



Missouri

End-of-Course Assessments

Missouri Department of Elementary and Secondary Education

Technical Report
Phase I Assessments
2010–2011

English II
Algebra I
Biology

Developed and published by The Riverside Publishing Company, a division of Houghton Mifflin Harcourt, 3800 Golf Road, Suite 200, Rolling Meadows, IL 60008-4015. Copyright © 2011 by the Missouri Department of Elementary and Secondary Education. All rights reserved. Only Missouri state educators and citizens may copy and/or download and print the document, located online at <http://dese.mo.gov/divimprove/assess/tech/index.html>.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	xi
List of Abbreviations	xii
Executive Summary	1
E.1 Background	1
E.2 Administration	1
E.3 Student Performance	2
E.4 Evidence Supporting the Validity of Inferences from the MO EOC Assessment Scores	2
Chapter 1: Introduction	7
1.1 History of Missouri’s End-of-Course Assessments	7
1.2 Brief Description of Missouri’s Current Assessment System	9
1.3 Summary of the MO EOC Assessments	9
1.4 Testing, Reporting, and Accountability	10
1.5 MO EOC Assessments Organizational Support	11
1.6 Purpose of the Technical Report	12
Chapter 2: Test Development	15
2.1 Introduction	15
2.2 Design of the MO EOC Assessments	15
2.3 Test Blueprints	19
2.4 Test Specifications	24
2.5 Development of Test Items	45
2.6 Test Form Assembly	50
2.7 Braille and Large Print Versions	52
2.8 Online Forms Construction	52
2.9 Quality Control for Test Construction	52
2.10 Summary	53
Chapter 3: Achievement-Level Setting	55
3.1 Introduction	55

3.2	Goal of the Standard Setting	55
3.3	Staff and Participants	55
3.4	Development of Achievement-Level Descriptors	56
3.5	Overview of Standard-Setting Activities	56
3.6	Specific Standard-Setting Activities	68
3.7	Session Results by Panel and Round	63
3.8	Results of Participant Evaluations	65
Chapter 4: Item Analysis		66
4.1	Introduction	66
4.2	Analysis of Forms for Each End-of-Course Assessment	66
4.3	Speededness	77
4.4	Item Bias Statistics	77
4.5	Summary	81
Chapter 5: Test Administration		83
5.1	Introduction	83
5.2	Students for Whom the EOC Assessments Are Appropriate	83
5.3	Students for Whom a School or District is Accountable	84
5.4	Dissemination of Testing Materials and Information	84
5.5	District and Test Examiner Training	84
5.6	Test Security	85
5.7	Test Administration	85
5.8	Accommodations and Modifications	90
5.9	Materials Handling and Return	91
5.10	Summary	92
Chapter 6: Scanning, Scoring, and Quality Control Procedures		94
6.1	Introduction	94
6.2	Quality Control Overview	94
6.3	Preparation and Materials Check-In	94
6.4	Materials Scanning	95
6.5	Quality Control in Report Production	96
6.6	Quality Assurance Product Review	97

Chapter 7: Scaling and Equating	98
7.1 Introduction	98
7.2 Item Response Theory	99
7.3 Scaling and Equating	99
Chapter 8: Reporting	124
8.1 Introduction	124
8.2 Individual Student Report	125
8.3 Student Score Label	127
8.4 Online Crystal Reports	127
Chapter 9: Summary Statistics	131
9.1 Introduction	131
9.2 Descriptive Statistics for Total Raw Score	131
9.3 Descriptive Statistics for Total Raw Score by Cluster	132
9.4 Descriptive Statistics for Scale Scores by Test Period and Subject	133
9.5 Descriptive Statistics by Demographic Group	135
Chapter 10: Reliability	174
10.1 Introduction	174
10.2 Reliability and Measurement Error	174
10.3 Evidence of Raw-Score Internal Consistency	175
10.4 Conditional Standard Error Estimates for Scale Scores	186
10.5 Reliability of Classifications	186
Chapter 11: Validity	191
11.1 Introduction	191
11.2 Federal Authority for School Accountability	191
11.3 Purpose and Intended Uses of Test Scores	191
11.4 MO EOC Assessment Scores	192
11.5 Content-Related Evidence of Validity	192
11.6 Validity Evidence Based on the Internal Structure of the MO EOC Assessments	194
11.7 Discriminant Validity Evidence for the MO EOC Assessments	198
11.8 Additional Validity Evidence for the MO EOC Assessments	199
11.9 Summary	200

References	202
Appendices	204
Appendix A: Demographic Characteristics of Standard-Setting Participants	204
Appendix B: Standard-Setting Session Agenda	206
Appendix C: Opening Session PowerPoint Presentation	208
Appendix D: Draft ALDs	220
Appendix E: Final ALDs	230
Appendix F: Qualifying Test	241
Appendix G: Participant Evaluation	243
Appendix H: Results for English II	249
Appendix I: Results for Algebra I	254
Appendix J: Results for Biology	259
Appendix K: Results of Participant Evaluation	264
Appendix L: Date Ranges for Hand Scoring of Performance Event/Writing Prompt Items	280

LIST OF TABLES

Table E.1	Percentage of Students at Each Achievement Level.....	2
Table 1.1	Main Activities for Groups Involved in MO EOC Organizational Support	11
Table 2.1	Linking Design for the MO EOC Assessments	16
Table 2.2	Test Construction Blueprint for English II with Writing Prompt.....	20
Table 2.3	Test Construction Blueprint for Algebra I with Performance Event	21
Table 2.4	Test Construction Blueprint for Biology with Performance Events	22
Table 2.5	Test Construction Blueprint for English II without Writing Prompt.....	23
Table 2.6	Test Construction Blueprint for Algebra I without Performance Event	23
Table 2.7	Test Construction Blueprint for Biology without Performance Events.....	24
Table 2.8	Target Point Distributions for the English II Operational Forms	26
Table 2.9	Actual Point Distributions for the 2010-2011 English II Operational Forms	29
Table 2.10	Target Point Distributions for the Algebra I Operational Forms.....	30
Table 2.11	Actual Point Distributions for the 2010-2011 Algebra I Operational Forms	33
Table 2.12	Target Point Distributions for the Biology I Operational Forms.....	34
Table 2.13	Actual Point Distributions for the 2010-2011 Biology Operational Forms	44
Table 2.14	2007 Content/Bias Item Review Acceptance Rates.....	49
Table 2.15	2008 Content/Bias Item Review Acceptance Rates.....	49
Table 2.16	Criteria for Flagged Items.....	51
Table 3.1	Median Recommended Cut Scores by Content Area and Round	64
Table 3.2	Projected Statewide Percentages of Students Scoring in the Various Performance Categories on the EOC Assessments, 2008.....	64
Table 4.1	Item Statistics for English II, Summer 2010 Operational Administration....	68
Table 4.2	Item Statistics for Algebra I, Summer 2010 Operational Administration	69
Table 4.3	Item Statistics for Biology, Summer 2010 Operational Administration.....	70
Table 4.4	Item Statistics for English II, Fall 2010 Operational Administration.....	71
Table 4.5	Item Statistics for Algebra I, Fall 2010 Operational Administration.....	72
Table 4.6	Item Statistics for Biology, Fall 2010 Operational Administration.....	73
Table 4.7	Item Statistics for English II, Spring 2011 Operational Administration	74

Table 4.8	Item Statistics for Algebra I, Spring 2011 Operational Administration	75
Table 4.9	Item Statistics for Biology, Spring 2011 Operational Administration	76
Table 4.10	Frequency Distribution of DIF Categories by Item Type for the Summer 2010, Fall 2010, and Spring 2011 Operational Assessments.....	79
Table 4.11	Frequency Distribution of DIF Categories by Item Type for the Spring 2008 Field Test Items	80
Table 4.12	Frequency Distribution of DIF Categories by Item Type for the Spring 2009 Field Test Items	80
Table 5.1	Frequency and Percentage of Students Receiving Each Type of Allowable Accommodation on the 2010-2011 EOC Assessments	91
Table 7.1	Linking Design for the MO EOC Assessments	99
Table 7.2	Mean Shift Values for the Spring 2009 Operational Test.....	100
Table 7.3	Comparison of 2010 Item Pool with 2010–2011 Operational Test Forms	103
Table 7.4	Summary Statistics for the Spring 2011 Operational Administration	103
Table 7.5	Results of the PCA for the Spring 2011 Operational Tests	104
Table 7.6	Summary Statistics for the Spring 2011 Operational Test Calibrations for English II.....	107
Table 7.7	Summary Statistics for the Spring 2011 Operational Test Calibrations for Algebra I.....	108
Table 7.8	Summary Statistics for the Spring 2011 Operational Test Calibrations for Biology	108
Table 7.9	Item Statistics for the Spring 2011 Operational Test Calibrations for English II.....	109
Table 7.10	Item Statistics for the Spring 2011 Operational Test Calibrations for Algebra I.....	110
Table 7.11	Item Statistics for the Spring 2011 Operational Test Calibrations for Biology.....	111
Table 7.12	Scale Score Ranges for the EOC Assessment Achievement Levels.....	112
Table 7.13	Summary of Slopes and Intercepts of Theta to Scale Score Transformation Functions by Content Area	113
Table 7.14	Raw Score to Scale Score Conversions for Summer 2010, English II	114
Table 7.15	Raw Score to Scale Score Conversions for Summer 2010, Algebra I.....	115
Table 7.16	Raw Score to Scale Score Conversions for Summer 2010, Biology	116
Table 7.17	Raw Score to Scale Score Conversions for Fall 2010, English II	117
Table 7.18	Raw Score to Scale Score Conversions for Fall 2010, Algebra I.....	118

Table 7.19	Raw Score to Scale Score Conversions for Fall 2010, Biology	119
Table 7.20	Raw Score to Scale Score Conversions for Spring 2011, English II.....	120
Table 7.21	Raw Score to Scale Score Conversions for Spring 2011, Algebra I.....	121
Table 7.22	Raw Score to Scale Score Conversions for Spring 2011, Biology	122
Table 9.1	Descriptive Statistics for Total Raw Score	131
Table 9.2	Descriptive Statistics for Total Raw Score by Test Administration, Content Area, and Cluster.....	132
Table 9.3	Scale Score Distributions for Each End-of-Course Assessment.....	133
Table 9.4	Achievement-Level Distributions for Each End-of-Course Assessment....	134
Table 9.5	Scale Score Distributions by Demographic Group—Gender	136
Table 9.6	Scale Score Distributions by Demographic Group—Ethnicity	137
Table 9.7	Scale Score Distributions by Demographic Group—Migrant Status	140
Table 9.8	Scale Score Distributions by Demographic Group—Free and Reduced Lunch	140
Table 9.9	Scale Score Distributions by Demographic Group—Limited English Proficient	141
Table 9.10	Scale Score Distributions by Demographic Group—Title I	141
Table 9.11	Scale Score Distributions by Demographic Group—Students with IEPs	142
Table 9.12	Scale Score Distributions by Demographic Group—Students with Accommodations	142
Table 9.13	Achievement-Level Distributions by Gender	143
Table 9.14	Achievement-Level Distribution by Ethnicity.....	146
Table 9.15	Achievement-Level Distribution—Migrant.....	155
Table 9.16	Achievement-Level Distribution—FRL	158
Table 9.17	Achievement-Level Distribution—LEP	161
Table 9.18	Achievement-Level Distribution—Title I.....	164
Table 9.19	Achievement-Level Distribution—IEP	167
Table 9.20	Achievement-Level Distribution—Accommodations	170
Table 10.1	Alpha Coefficients and Standard Errors of Measurement, English II, Summer 2010.....	177
Table 10.2	Alpha Coefficients and Standard Errors of Measurement, English II, Fall 2010	178
Table 10.3	Alpha Coefficients and Standard Errors of Measurement, English II, Spring 2011	179

Table 10.4	Alpha Coefficients and Standard Errors of Measurement, Algebra I, Summer 2010	180
Table 10.5	Alpha Coefficients and Standard Errors of Measurement, Algebra I, Fall 2010	181
Table 10.6	Alpha Coefficients and Standard Errors of Measurement, Algebra I, Spring 2011	182
Table 10.7	Alpha Coefficients and Standard Errors of Measurement, Biology, Summer 2010	183
Table 10.8	Alpha Coefficients and Standard Errors of Measurement, Biology, Fall 2010	184
Table 10.9	Alpha Coefficients and Standard Errors of Measurement, Biology, Spring 2011	185
Table 10.10	<i>CSEMs</i> at the Proficient Cut Score	186
Table 10.11	Classification Consistency Coefficients	187
Table 10.12	Raw Agreement Consistency Coefficients	188
Table 10.13	Classification Consistency Coefficients (Two Classification Categories)	188
Table 10.14	Raw Agreement Consistency Coefficients (Two Classification Categories)	189
Table 11.1	Pearson Correlation Coefficients Between Domains and Clusters for English II	194
Table 11.2	Pearson Correlation Coefficients Between Domains and Clusters for Algebra I	195
Table 11.3	Pearson Correlation Coefficients Between Domains and Clusters for Biology	196
Table 11.4	Pearson Correlation Between Phase I Assessments	198
Table 11.5	Pearson Correlation Between Phase I and Phase II Assessments	199
Appendix Table A.1	English II	204
Appendix Table A.2	Algebra I	204
Appendix Table A.3	Biology	205

LIST OF FIGURES

Figure 2.1	2008-2009 Field-Test and Operational Assessment Design, English II.....	15
Figure 2.2	2008-2009 Field-Test and Operational Assessment Design, Algebra I and Biology.....	16
Figure 5.1	Examples of Type Styles Used to Differentiate between Spoken and Unspoken Instructions in the <i>Test Examiner’s Manual</i>	88
Figure 5.2	Example Script from the <i>Test Examiner’s Manual</i> for the Algebra I EOC Assessment	88
Figure 5.3	Example Script from the <i>Online Test Examiner’s Manual</i> for the Online English II EOC Assessment.....	89
Figure 7.1	TCCs for Three Operational Forms for English II for the 2010-2011 Test Administration	101
Figure 7.2	TCCs for Three Operational Forms for Algebra I for the 2010-2011 Test Administration	102
Figure 7.3	TCCs for Three Operational Forms for Biology for the 2010-2011 Test Administration	102
Figure 7.4	Standardized Residual Correlations from the Spring 2011 Administration for English II.....	105
Figure 7.5	Standardized Residual Correlations from the Spring 2011 Administration for Algebra I	105
Figure 7.6	Standardized Residual Correlations from the Spring 2011 Administration for Biology.....	106
Figure 8.1	Individual Student Report.....	126
Figure 8.2	Student Score Label	127

LIST OF ABBREVIATIONS

Below is a list of abbreviations that appear frequently in this technical report.

ALD	Achievement-Level Descriptor
AMO	Additional Materials Order
ARC	Assessment Resource Center
AYP	Adequate Yearly Progress
CCSSO	Council of Chief State School Officers
CLE	Course-Level Expectation
CMS	Content Management System
<i>CSEM</i>	Conditional Standard Error of Measurement
CTT	Classical Test Theory
DESE	Department of Elementary and Secondary Education
DIF	Differential Item Functioning
DOK	Depth of Knowledge
EFT	Embedded Field Test
EOC	End-of-Course
FRL	Free and Reduced Lunch
FT	Field Test
GLE	Grade-Level Expectation
GRF	General Research File
IAP	Individualized Accommodation Program
IDEA	Individuals with Disabilities Education Act
IEP	Individualized Education Program
IPASS	Image-Based Performance Assessment Scoring System
IRT	Item Response Theory
ITS	Internet Testing Systems
LEP	Limited English Proficient
LOSS	Lowest Obtainable Scale Score
MAP	Missouri Assessment Program
MH	Mantel-Haenszel procedure
MOSIS	Missouri Student Information System
MS	Mean Square
NCLB	No Child Left Behind
PCA	Principal Components Analysis
PE	Performance Event
RPC	Riverside Publishing Company
RSS	Riverside Scoring Service [®]
<i>SE</i>	Standard Error
<i>SEM</i>	Standard Error of Measurement
SR	Selected Response
TAC	Technical Advisory Committee
TCC	Test Characteristic Curve
TDS	Test Development Specialist
WP	Writing Prompt

EXECUTIVE SUMMARY

This document provides a technical summary of the 2010–2011 administrations of the Missouri End-of-Course (MO EOC) Assessments in English II, Algebra I, and Biology. The criterion-referenced MO EOC Assessments are designed to assess students' knowledge of Missouri's Course-Level Expectations (CLEs) in these three content areas. The 2010–2011 school year marked the third operational administration of the assessments. For a technical summary of the first two operational administrations, see the *2008-2009 Missouri End-of-Course Assessments Technical Report* and the *2009-2010 Missouri End-of-Course Assessments Technical Report*.

E.1 Background

In 1993, the Missouri legislature passed the Outstanding Schools Act (Senate Bill 380), requiring the Missouri State Board of Education to adopt challenging academic performance standards that define the skills and competencies necessary for students to successfully advance through the public school system, prepare for post-secondary education and the workplace, and participate as citizens in a democratic society. The Missouri State Board of Education formally adopted the academic standards known as the Show-Me Standards in January 1996.

In addition to mandating the development of rigorous academic standards, the Outstanding Schools Act of 1993 required the development and implementation of a comprehensive assessment program to measure student proficiency in the knowledge, skills, and competencies identified in the standards. Upon adoption of the standards in 1996, Missouri began developing the Missouri Assessment Program (MAP).

In January 2007, the Missouri State Board of Education approved a plan to replace the MAP for high school students with MO EOC Assessments beginning with English II, Algebra I, and Biology in the 2008–2009 school year. The MO EOC assessments have been administered each summer, fall, and spring since the 2008-2009 school year.

E.2 Administration

The MO EOC Assessments are administered in summer, fall, and spring windows each year. Because the No Child Left Behind Act (NCLB) goal for every school in the state is Proficient, as defined by the Missouri State Board of Education, EOC testing is conducted as close as possible to the end of each course to allow school staff and students the greatest opportunity to achieve the goal of proficiency.

The scope of this technical report includes the Summer 2010, Fall 2010, and Spring 2011 administrations. Data analyses for the total assessed population, which include students who have not yet reached the secondary level, are based on a combination of assessment results as well as Missouri Department of Elementary and Secondary Education (DESE) demographic criteria as required under NCLB.

District- and state-level reports for the MO EOC Assessments are produced and distributed following each Spring administration. The score reports for the 2010–2011 assessment year contained information from the Summer 2010, Fall 2010, and Spring

2011 assessments. In future years, reports will continue to include information from the previous year’s summer, fall, and spring administrations.

E.3 Student Performance

A MO EOC assessment score describes the relationship of student performance to a defined level of achievement. Achievement-level descriptors (ALDs) associated with each level provide details about the content expectations that students at that level meet or exceed. Missouri uses four achievement levels for the EOC Assessments: Below Basic, Basic, Proficient, and Advanced.

Table E.1 displays the percentage of students at each achievement level for the Summer 2010, Fall 2010, and Spring 2011 MO EOC Assessments. The NCLB Act requires states to assess all students at least once in high school in mathematics, English/communication arts, and science. Students who take the MO EOC Assessment but are not yet in high school are not included in Missouri’s high school accountability data. Rather, their scores are “banked” until they actually reach high school, at which time they are rolled into the high school accountability data for that year. However, the data for *all* tested students are used each year for purposes of item analysis, scaling, and equating. For this reason, the numbers and/or percentages of tested students reported in Table E.1 and elsewhere in this technical report do not match the numbers of students reported by DESE for accountability purposes.

Table E.1 Percentage of Students at Each Achievement Level

Test Administration	Achievement Level	English II	Algebra I	Biology
Summer 2010	Below Basic	16.0	11.8	23.2
	Basic	44.7	49.9	41.7
	Proficient	35.3	30.2	30.5
	Advanced	4.0	8.1	4.7
Fall 2010	Below Basic	7.9	12.9	12.2
	Basic	29.3	36.2	34.8
	Proficient	42.6	26.3	31.8
	Advanced	20.2	24.6	21.2
Spring 2011	Below Basic	4.0	8.4	6.3
	Basic	21.1	29.6	31.0
	Proficient	48.2	41.5	46.8
	Advanced	26.6	20.5	15.9

E.4 Evidence Supporting the Validity of Inferences from the MO EOC Assessment Scores

The MO EOC Assessments are part of an integrated program of testing, accountability, and curricular and instructional support. This technical report provides extensive detail about the development and operation of EOC Assessments. While a section of this report is devoted specifically to the documentation of validity evidence for the MO EOC Assessment scores, all information contained in the report ultimately contributes to the argument for the validity of the scores for their intended purposes.

A summary of the information contained in this report follows.

Chapter 1: Introduction

Chapter 1 provides background information about MAP in general as well as some context for the MO EOC Assessments. Additionally, the chapter provides information about the organizational support provided by each contractor and subcontractor for the MO EOC Assessment program. The chapter ends with a statement of purpose for this technical report.

Chapter 2: Test Development

Chapter 2 contains thorough descriptions of each step in the development process for the MO EOC Assessments, including test design, test specifications and target point distributions, test blueprints, item writing, content and bias review procedures, test form assembly, and statistical item review. The evidence provided in this chapter is important to the content-related validity of the MO EOC Assessment scores. Additionally, the chapter covers principles of universal design and outlines the quality control processes employed throughout the test development process.

Chapter 3: Achievement-Level Setting

Chapter 3 details each step in the planning and execution of the 2009 standard-setting event that resulted in the cut scores for each of the MO EOC achievement levels. While this chapter was included in the *2009 Missouri End-of-Course Assessments Technical Report*, it is repeated here for the convenience of the reader because the results are relevant to the current test administrations. Chapter 3 covers selection of participants, development of ALDs, an overview of the methodology and considerations for the data that were available at the time of the standard-setting event, detailed information about each step in the process, and standard-setting results. Additionally, the chapter contains many appendices with examples of the materials that participants used during the standard-setting event.

Chapter 4: Item Analysis

Chapter 4 contains summary information, including item difficulty and discrimination indices, at the item level for each content area. The chapter also contains information on omit rates for the Summer 2010, Fall 2010, and Spring 2011 operational items, as well as differential item functioning (DIF) analyses performed on field-test item data.

Chapter 5: Test Administration

Chapter 5 contains information about the paper-and-pencil and online administration of the MO EOC Assessments, beginning with a description of students for whom the assessments are appropriate. Following this, the details of the administration are summarized. This summary includes a description of how the materials are distributed and how Test Examiners are trained, as well as information about the organization of the assessments, preparation of students to take the assessments, and directions for administration. Next, the chapter includes information about the accommodations allowed on the MO EOC Assessments. Finally, the chapter briefly describes how materials are submitted for processing and scoring.

Chapter 6: Scanning, Scoring, and Quality Control Procedures

Chapter 6 covers the processes involved with scanning, scoring, and controlling the quality of the resulting score information for the selected response items on the MO EOC Assessments. The chapter contains detailed information on how the Riverside Publishing Scoring Service® (RSS) prepared for processing the MO EOC selected response items, including a check of scanning procedures prior to receipt of materials, and how the materials were handled from the time they were received and processed at the RSS on through to report generation.

Chapter 7: Scaling and Equating

Chapter 7 begins with an introduction to the item response theory (IRT) model used for the scaling and equating of the MO EOC Assessments. Next, the steps are given for scaling and equating the 2008 standalone field-test items, the Spring 2009 operational forms (for establishing the base scale), and the Spring 2009 field-test items. Included next is a description of the process to reestablish the base scale in response to the removal of the PE items. Finally, information is presented for the Summer 2010, Fall 2010, and Spring 2011 operational forms, IRT model assumptions and evidence of data-to-model fit.

Chapter 8: Reporting

Chapter 8 contains information about the reports Riverside Publishing produced for the MO EOC Assessments, including the Individual Student Report and Student Score Label. A brief summary of state-produced reports is also included.

Chapter 9: Summary Statistics

Chapter 9 provides descriptive statistics for raw scores and scale scores for the MO EOC Assessments. Raw score statistics are summarized by test administration, content area, and cluster. Scale score statistics are summarized for each content area and are also broken down by gender, ethnicity, migrant status, free and reduced lunch (FRL), limited English proficiency, Title I, individualized education program, and accommodations.

Chapter 10: Reliability

Chapter 10 begins by defining reliability and providing an overview of reliability estimation techniques. Raw-score internal consistency reliability coefficients are presented for all students and for each demographic group. Conditional standard errors of measurement (*CSEMs*) are presented at each scale-score cut point. In previous technical manuals, inter-rater reliability statistics were presented for the performance events. However, the performance events were removed beginning with the Summer 2010 test administration, and therefore there is no inter-rater reliability data.

Chapter 11: Validity

Chapter 11 provides evidence supporting the validity of the MO EOC Assessments for their intended purposes. After a brief introduction to the validity evidence for the MO EOC Assessments, the chapter documents more specific evidence related to test content, the internal structure of the assessments, and other types of validity evidence proposed by the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999). The chapter summarizes and reiterates validity evidence presented in earlier chapters in addition to providing new information not presented elsewhere. It provides a thorough argument supporting the validity of the MO EOC Assessments for measuring Missouri students' mastery of the CLEs, for identifying students' strengths and weaknesses, for serving as a basis for evaluating accountability plans, and for program evaluation.

CHAPTER 1: INTRODUCTION

1.1 History of Missouri's End-of-Course Assessments

In 1993, the Missouri legislature passed the Outstanding Schools Act (Senate Bill 380), requiring the Missouri State Board of Education to adopt challenging academic performance standards defining the skills and competencies necessary for students to successfully advance through the public school system, prepare for post-secondary education and the workplace, and participate as citizens in a democratic society. The Missouri State Board of Education formally adopted the academic standards known as the Show-Me Standards in January 1996.

These 73 standards are organized around four broad goals that address application, communication, problem solving, and responsible decision-making. Thirty-three process standards emphasize the importance of engaging students of all ages in hands-on, active learning and integrating practical, challenging learning across all content areas. An additional 40 content standards define the academic skills and knowledge that provide the foundation for student learning in six content areas: Communication Arts, Mathematics, Science, Social Studies, Fine Arts, and Health/Physical Education. Content standards serve as the vehicle through which students demonstrate proficiency in the broader process standards. The Show-Me Standards are available for review on the Missouri Department of Elementary and Secondary Education's (DESE) website at <http://dese.mo.gov/standards/index.html>.

In 2001, DESE developed Grade-Level Expectations (GLEs) to assist districts in articulating the Show-Me Standards across grade levels and content areas. GLEs have been developed for Mathematics, Communication Arts, Science, Social Studies, Physical Education, Health, Music, Visual Arts, and Theater. GLEs are available for review on the DESE website at <http://dese.mo.gov/divimprove/curriculum/GLE/index.html>.

In addition to mandating the development of rigorous academic standards, the Outstanding Schools Act of 1993 also required the development and implementation of a comprehensive assessment program to measure student proficiency in the knowledge, skills, and competencies identified within the standards. Upon adoption of the standards in 1996, Missouri began developing the Missouri Assessment Program (MAP) in collaboration with contractor CTB/McGraw-Hill.

The Missouri State Board of Education adopted the purposes listed below to serve as guiding principles for development of the MAP:

- Improving students' acquisition of important knowledge, skills, and competencies
- Monitoring the performance of Missouri's educational system
- Empowering students and their families to improve their educational prospects
- Supporting the teaching and learning process

The first MAP assessments administered to students statewide were grade-span Mathematics assessments in grades 4, 8, and 10 in spring 1998. A voluntary grade-span Communication Arts Assessment for students in grades 3, 7, and 11 was also administered in spring 1998 and became mandatory in spring 1999. Required Science and Social Studies grade-span assessments (grades 3, 7, and 10, and grades 4, 8, and 11, respectively) were added to the program in subsequent years. A voluntary Health/Physical Education assessment was available in 2000 and required until spring 2002, and a Fine Arts assessment was field tested in 2001. Due to budget constraints, Fine Arts assessment development was suspended and Health/Physical Education Assessment was discontinued. Science and Social Studies grade span assessments returned to voluntary status in spring 2003. Social Studies assessments were discontinued in spring 2008 and required assessments in Science were implemented in grades 5, 8, and 11 to comply with NCLB requirements.

Through the spring 2005 administration, the MAP statewide assessment program included grade-span tests in the following grade levels/subject areas:

- Mathematics at grades 4, 8, and 10
- Communication Arts at grades 3, 7, and 11
- Science at grades 3, 7, and 10 (required spring 1998 through spring 2002; returned to voluntary status in spring 2003)
- Social Studies at grades 4, 8, and 11 (required spring 1999 through spring 2002; returned to voluntary status in spring 2003))

All MAP assessments included three types of items: selected response (SR), constructed response (CR), and performance events (PE). For all content areas, MAP assessments included selected response items from the TerraNova Survey Edition. Constructed response items and performance events were custom-developed with significant input from Missouri educators.

During the initial MAP development/implementation period, DESE developed two to four equivalent forms for each content area/grade level assessment, using the first form for a voluntary testing cycle and administering the next form(s) in subsequent years. Early in the development phase, DESE tried out new items using separate field tests that usually occurred in the fall of the school year. As the program continued, each test form contained embedded field test items. Small-scale pilots continued as well.

As each content area/grade level assessment was administered, DESE used the Bookmark approach to set achievement levels, defining student performance through spring 2005 as Advanced, Proficient, Nearing Proficiency, Progressing, or Step 1.

After nearly a decade of MAP administration, new federal and state legislation prompted change in the program. To comply with requirements of No Child Left Behind (NCLB) legislation, Missouri's assessment program needed to incorporate Mathematics and Communication Arts assessments at all elementary and middle school grade levels (grades 3 through 8) and at one high school grade level. As a result, new grade-level assessments were developed for both content areas. These assessments were administered for the first time in spring 2006.

Additional NCLB requirements necessitated the addition of a mandatory science assessment once in the elementary grade range, once in the middle school grade range, and once in the high school grade range, beginning in spring 2008. The voluntary Science assessment in grades 3, 7, and 10 became a requirement and was moved to grades 5, 8, and 11. The voluntary Social Studies MAP Assessment was eliminated following the spring 2007 administration. Missouri's assessment system changed further in 2008–2009, when high school content area MAP assessments were replaced by End-of-Course (EOC) Assessments.

1.2 Brief Description of Missouri's Current Assessment System

The current MAP system includes the following assessment components for elementary and middle school:

- Grades 3–8 Communication Arts
- Grades 3–8 Mathematics
- Grades 5 and 8 Science

The EOC Assessments administered in 2010–2011 included:

- English II
- Algebra I
- Biology

In addition, the statewide assessment program currently includes the Missouri Assessment Program–Alternate (MAP-A) for students with severe cognitive disabilities, WIDA ACCESS for ELLs, and a Personal Finance assessment for high school students who do not enroll in a personal finance course or who are receiving personal finance credit for embedded coursework.

1.3 Summary of the MO EOC Assessments

In response to feedback from Missouri districts regarding large-scale assessments for high school, the Missouri End-of-Course (MO EOC) Assessments were developed and first administered in 2008 for English II, Algebra I, and Biology. The MO EOC Assessments were created to assess the Course-Level Expectations while also meeting state and federal requirements. The Missouri State Board of Education identified the following purposes for the Missouri EOC Assessments:

- Measuring and reflecting students' mastery toward post-secondary readiness
- Identifying students' strengths and weaknesses
- Communicating expectations for all students
- Serving as the basis for state and national accountability plans
- Evaluating programs

Course-Level Expectations (CLEs) outline the ideas, concepts, and skills that form the foundation for an assessed EOC subject area, regardless of student grade level. During the 2008-2009 and 2009-2010 school years, each MO EOC Assessment included two types of test items: Selected Response (SR) items and Performance Events (PEs), which

include Writing Prompts (WPs). An SR item presents students with a question followed by four response options. The PE items are constructed response items that require students to perform more complicated tasks. A PE often allows more than one approach to arrive at a correct response. The advantage of this type of item is that it provides insight into a student's ability to apply knowledge and understanding in real-life situations. The WP, a special type of PE that appears in the English II EOC Assessment, is an open-ended item that requires students to demonstrate their writing proficiency. Beginning with the Summer 2010 administration, PE items and the WP were removed from the EOC Assessments.

The 2010-2011 MO EOC Assessments were offered in both paper-and-pencil and online administration modes.

1.4 Testing, Reporting, and Accountability

Evidence of students' progress in meeting the Show-Me Content Standards/CLEs is obtained from the MO EOC Assessments. These assessments provide the data that DESE uses to inform students, parents, the public, and the state legislature about students' performance, to help make informed decisions about educational issues, and to drive student services throughout the state.

The MO EOC Assessment reports provide useful information for determining the performance of students in a particular school and classroom. These reports help identify students who are below Proficient in a particular test area so that the school may determine a course of action that will meet the students' specific needs. Additionally, districts may use locally designed assessments, aligned to the Show-Me Content Standards/CLEs, to provide more detailed information for each student in specific test areas.

Testing for the MO EOC Assessments is conducted during three state-designated windows each year. Test windows are available for Summer, Fall, and Spring. Per contract requirements, however, paper reports for all administrations are provided only after the Spring testing window each year. (To aid in course grading, teachers may use an online interface to access student raw scores for the selected response items) Because the NCLB goal for every school in the state is Proficient, as defined by the Missouri State Board of Education, MO EOC testing is conducted as close as possible to the end of each course to allow school staff and students the greatest opportunity to achieve that goal.

Data for this technical report were collected during the Summer 2010, Fall 2010, and Spring 2011 operational administrations. Data analyses for the total assessed population, which includes students who have not yet reached the secondary level, are based on a combination of assessment results as well as DESE-provided demographic criteria required under NCLB.

1.5 MO EOC Assessments Organizational Support

DESE coordinates the development and implementation of the MO EOC Assessments. In addition to planning, scheduling, and directing all EOC activities, the staff is extensively involved in numerous test reviews, security, and quality assurance procedures. Riverside Publishing is the primary contractor working in partnership with Questar, the Assessment

Resource Center (ARC), Internet Testing Systems (ITS), Bookette, and others. The main activities for each of these groups are outlined in Table 1.1.

Table 1.1: Main Activities for Groups Involved in MO EOC Organizational Support

Group	Responsibilities
Riverside Publishing	<ul style="list-style-type: none"> • Provides program management, including primary contact with DESE; coordinates all meetings; handles all administrative costs/activities; generates all program management reports and status reports • Works with DESE to develop items with Missouri educators • Creates <i>Test Coordinator’s Manual</i>, <i>Test Examiner’s Manuals</i>, and other ancillary materials • Facilitates all review meetings with Missouri teachers and DESE • Conducts all psychometric analyses, reporting, linking/equating studies, and associated tasks, including participating in achievement-level setting • Provides all needed prepress work for program materials through camera-ready art • Produces all materials, including online, paper-and-pencil, Braille, and Large Print versions of test • Accounts for secure test books received after testing • Provides a direct customer service line, including technical support and general support to the program and customer interactions • Stores materials after testing • Participates in and presents at TAC meetings • Scores all selected response items • Produces and distributes all score reports and the <i>Guide for Interpreting Results</i> • Completes the technical report for DESE • Completes additional research studies
Questar	<ul style="list-style-type: none"> • Provides online enrollment and pre-ID system for use by Missouri districts • Packages and distributes all materials • Barcodes test books with security IDs • Leads facilitation and planning of achievement-level setting and provides members for the achievement-level-setting team • Contributes to the technical report • Participates in meetings with DESE, contributes to status reports, etc.

Table 1.1: Main Activities for Groups Involved in MO EOC Organizational Support (continued)

Group	Responsibilities
Assessment Resource Center (ARC)	<ul style="list-style-type: none"> • Provides facilities for item writing if contracted by DESE • Participates in meetings with DESE, contributes to status reports, etc.
Internet Testing Services (ITS)	<ul style="list-style-type: none"> • Sets up a Missouri DESE-branded website for access to the online testing system • Provides the online test delivery of one complete form for each administration • Provides system documentation for test administrators and the DESE website • Provides technical support from 8 A.M. to 6 P.M., Monday through Friday, for the Riverside Publishing help desk • Produces and hosts practice tests for the English II, Algebra I, and Biology content areas • Provides online tools for graphing and table creation/editing and provides an equation editor • Offers ruler and reference sheets in tests • Provides three administrations per contract year in fall, spring, and summer for all content areas • Supplies a data feed of results from ITS to Riverside Publishing •
Districts	<ul style="list-style-type: none"> • Distribute materials to the school buildings, track all secure materials, and promptly return all materials, including answer documents, for scoring • Assist in the timely resolution of scoring alerts • Act as liaison between Riverside Publishing and buildings
School Buildings	<ul style="list-style-type: none"> • Administer tests, track all secure materials, and promptly return materials to districts for scoring
RR Donnelly	<ul style="list-style-type: none"> • Prints all nonscannable testing materials
Techniforms	<ul style="list-style-type: none"> • Prints all scannable test books and answer documents
Region IV, 3X	<ul style="list-style-type: none"> • Prints Braille and Large Print versions, respectively

1.6 Purpose of the Technical Report

The purpose of this technical report is to provide information about the technical characteristics of the 2010–2011 operational administration of the MO EOC Assessments. Because this report is technical in nature and the intended audience is psychometric and educational research experts, it is best understood with a working knowledge of measurement concepts such as reliability and validity and statistical concepts such as correlation and central tendency. For some chapters, the reader is presumed to have basic familiarity with advanced topics in measurement and statistics such as item response theory (IRT).

This technical report provides extensive detail about the development and operation of the MO EOC Assessments. The empirical reliability of the assessments and validity of intended uses of the scores are reported explicitly in this document. While Chapter 10: Reliability is relatively straightforward, the steps in creating and operating the program are all aspects of validity, which is discussed in Chapter 11. The validity of score use and

interpretation for any assessment stems from the statement of the test’s purpose and the intended use of the scores; the steps taken in designing the test; and the processes of developing the content of the test, consulting with stakeholders, communicating about the test to users, scoring and reporting, and data analysis. The careful documentation of each of these steps is a necessary piece of a comprehensive, defensible validity argument for the intended uses of the assessment scores. In short, while there is a specific chapter devoted to validity, other parts of this document provide evidence necessary to assess the validity of the MO EOC Assessment scores for their intended purposes.

In reading this technical report, it is critical to remember that the testing program does not exist in a vacuum; it is not just a test. It is one part of a complex network intended to help schools focus their energies on improving student learning. The MO EOC Assessments are an integrated program of testing and accountability as well as curricular and instructional support. They can be evaluated properly only within their full context.

CHAPTER 2: TEST DEVELOPMENT

2.1 Introduction

The English II, Algebra I, and Biology End-of-Course (EOC) Assessments were first administered operationally during the 2008–2009 school year. This chapter provides an overview of the development of the Missouri End-of-Course (MO EOC) Assessments, including the test specifications, item development, item review, and test forms development. Forms development is described for the summer, fall, and spring administrations across the 2009-2010 and 2010-2011 school years. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) (hereafter referred to as the Standards), “Important validity evidence can be obtained from an analysis of the relationship between a test’s content and the construct it is intended to measure” (p. 11). Accordingly, the thorough descriptions of the test development procedures included in this chapter provide validity evidence of the MO EOC Assessments.

2.2 Design of the MO EOC Assessments

Figure 2.1 details the design of the Spring 2008 standalone field test, Fall 2008 operational administration, and Spring 2009 operational administration with embedded field test for English II. Additionally, Figure 2.1 displays the design of the Spring 2009 standalone Writing Prompt (WP) field test.

Figure 2.1: 2008-2009 Field-Test and Operational Assessment Design, English II

	Spring 2008 Stand-Alone Field Test	Fall 2008 Operational Administration	Spring 2009 Operational Administration	Spring 2009 Stand-Alone WP Field Test
Session I	Form A (18 items) Form B (18 items) Form C (18 items) Form D (18 items) Form E (18 items) Form F (18 items) Form G (18 items) Form H (18 items) Form I (18 items) Form J (18 items)	Selected Response (35 items) Filler Items (12 items)	Selected Response (35 items) EF 1 EF 2 EF 3 EF 4 EF 5 EF 6 EF 7 EF 8 ••• EF 32	
Session II	WP FT (Form 1) WP FT (Form 2) WP FT (Form 3) WP FT (Form 4) WP FT (Form 5) WP FT (Form 6) WP FT (Form 7) WP FT (Form 8) WP FT (Form 9) WP FT (Form 10)	Writing Prompt (1 item)	Writing Prompt (1 item)	WP FT (Form 1) WP FT (Form 2) WP FT (Form 3) WP FT (Form 4) WP FT (Form 5) WP FT (Form 6) WP FT (Form 7) WP FT (Form 8) ••• WP FT (Form 20)

Figure 2.2 details the design of the Spring 2008 standalone field test, Fall 2008 operational administration, and Spring 2009 operational administration with embedded field test for Algebra I and Biology.

Figure 2.2: 2008-2009 Field-Test and Operational Assessment Design, Algebra I and Biology

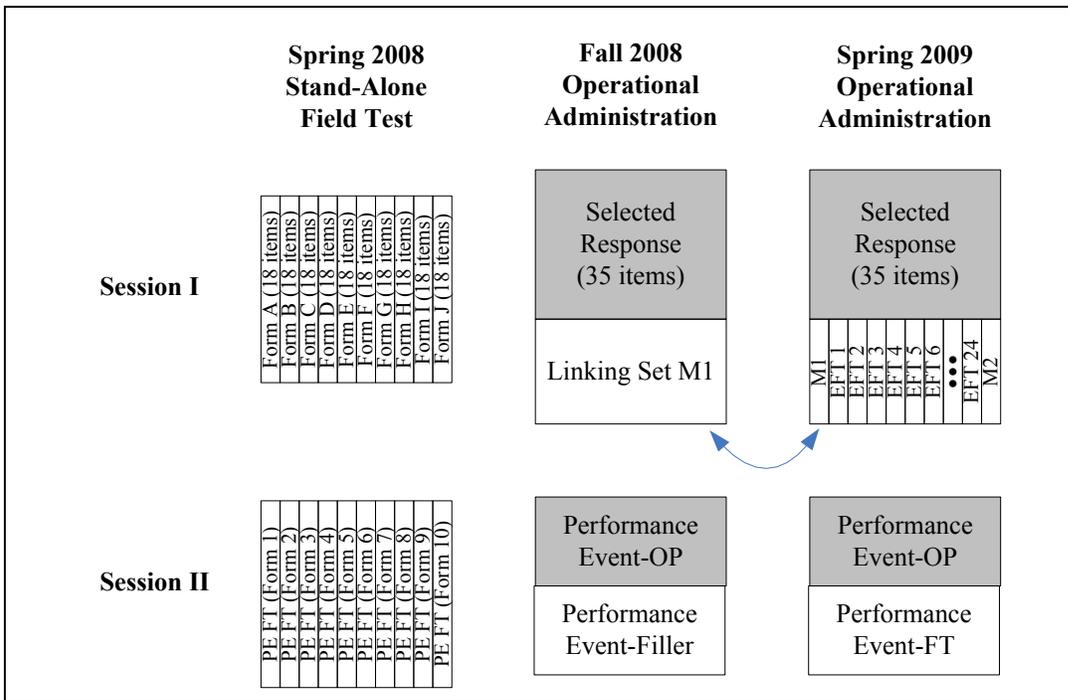


Table 2.1 lists the linking design for multiple operational administration years.

Table 2.1: Linking Design for the MO EOC Assessments

Administration	Assessment Form	Linking Set(s)
Fall 2008	A	M1
Spring 2009	B	M1, M2, Y1, Y2
Summer 2009	C	M2
Fall 2009	D	Y1
Spring 2010	A	Y1, Y2
Summer 2010	E	Y2
Fall 2010	B	M1, M2
Spring 2011	F	M1, M2

2.2.1 Spring 2008 Standalone Field Test

The Spring 2008 standalone field test provided item data to inform the 2008–2009 operational forms selection process. There were two sessions in the Spring 2008 field test. For each assessment, Session I included 10 unique forms of selected response (SR) items, with each form containing 19 items. Session II included 10 unique forms of Performance Event/Writing Prompt (PE/WP) items. For English II, the PE forms in Session II each consisted of one 4-point Writing Prompt. For Algebra I, each Session II form consisted of one 4-point PE. For Biology, each Session II form consisted of 10–12

constructed response (CR) items, for a total of 20 points on each form. Forms within each session were spiraled at the student level across the state.

2.2.2 Fall 2008 Operational Administration

The Fall 2008 administration consisted of three operational assessments. English II consisted of one 35-item SR form and one WP. Algebra I and Biology consisted of one 35-item SR form and one PE form each. The Algebra I PE consisted of one 4-point item. The Biology PE consisted of 10 CR items, ranging from 1 to 4 points each, for a total of 20 points.

In addition to the 35 scored items, each Fall 2008 Algebra I and Biology Session I test book contained a set of 12 linking items (designated as M1 in Figures 2.2 and 2.3). These linking items were used for the post-equating check of the pre-equating results following the Spring 2009 operational administration (indicated by an arrow in Figure 2.2). For the English II assessment, the 12 additional SR items in Session I were filler (nonscored) items.

2.2.3 Spring 2009 Operational Administration

The Spring 2009 assessments consisted of three operational assessments. For all three content areas, Session I consisted of 35 operational SR items. For English II, there were 32 unique sets of 12 embedded field test items (labeled as EFT 1 through EFT 32 in Figure 2.1). For Algebra I and Biology, there were 24 unique sets of 12 embedded field test items (labeled as EFT 1 through EFT 24 in Figure 2.2). Additionally, the sets of items used to link the Spring 2009 form to the Fall 2008 (M1) and Summer 2009 (M2) assessments for the post-equating check occupied two of the embedded field test slots on the Algebra I and Biology assessments.

Session II of the Spring 2009 English II assessment contained one 4-point WP item. Session II of the Algebra I assessment contained one 4-point PE. Finally, Session II of the Biology assessment contained 11 CR items, ranging from 1 to 4 points each, for a total of 20 points. Session II of the Algebra I and Biology assessments also contained an embedded field test PE.

2.2.4 Spring 2009 Standalone WP Field Test

There was a separate standalone WP field test in which 20 WPs were administered statewide, each on its own form. In Figures 2.1 and 2.2, the shaded portions designate the operational (scored) items.

2.2.5 Released Forms

In addition to the operational forms that were constructed for 2008–2009, the Missouri Department of Elementary and Secondary Education (DESE) and Riverside Publishing also worked together to construct “released” forms for each operational assessment. These forms were posted on the DESE website in August 2008. They were constructed to mirror the test content of the actual operational forms (minus the EFT items) to allow Missouri teachers, parents, and students the opportunity to review the new format and representative content of the EOC Assessments. Although these forms were constructed

to parallel the operational forms, the items in these released forms were never used on an operational EOC Assessment.

2.2.6 Fall 2009 Operational Administration

The Fall 2009 assessments consisted of three operational assessments. For all three assessments, Session I consisted of 35 operational SR items. Each form also contained 12 linking items used to equate the Fall 2009 forms to the Spring 2009 forms as shown in Figure 2.3.

Session II of the Fall 2009 English II assessment contained one 4-point WP item. Session II of the Algebra I assessment contained one 4-point PE. Finally, Session II of the Biology assessment contained 12 CR items, ranging from 1 to 4 points each, for a total of 20 points.

2.2.7 Spring 2010 Operational Administration

The Fall 2008 operational form was reused for the Spring 2010 operational SR assessment. There were two forms each of the English II, Algebra I, and Biology assessments, which contained 12 linking items. As shown in Figure 2.3, there were two unique sets of linking items per course. Set Y1 linked the Spring 2010 Form A to the Fall 2009 and the Spring 2008 forms. Set Y2 linked the Spring 2010 Form B to the Spring 2008 and the Summer 2009 forms.

Session II of the Spring 2010 English II assessment contained one 4-point WP item. Session II of the Algebra I assessment contained one 4-point PE item. Finally, Session II of the Biology assessment contained 12 CR items, ranging from 1 to 4 points each, for a total of 20 points. Session II of the Algebra I and Biology assessments also contained an embedded field test PE.

2.2.8 Summer 2010 Operational Administration

The Summer 2010 assessments consisted of three newly created operational assessments. For all three assessments, Session I consisted of 35 operational SR items. Each form also contained 12 linking items that were also administered on the Summer 2009 test form and the Spring 2010 test form as shown in Figure 2.3.

Session II of the Fall 2009 English II assessment contained one 4-point WP item. Session II of the Algebra I assessment contained one 4-point PE item. Finally, Session II of the Biology assessment contained 12 CR items, ranging from 1 to 4 points each, for a total of 20 points. Session II of the Algebra I and Biology assessments also contained an embedded field test PE. After the forms were created and printed, a decision was made to discontinue the use of PE and WP items on the MO EOC Assessments. Therefore, the student responses from this administration were not scored.

2.2.9 Fall 2010 Operational Administration

The Fall 2010 operational form was reused for the Spring 2009 operational SR assessment. There were two forms each of the English II, Algebra I, and Biology assessments, which contained 12 linking items. As shown in Figure 2.3, there were two unique sets of linking items per course. Set M1 linked the Fall 2010 form to the Fall 2008

and the Spring 2009 forms. Set M2 linked the Fall 2010 form to the Summer 2008 and the Spring 2009 forms.

2.2.10 Spring 2011 Operational Administration

The Spring 2011 assessments consisted of three newly created operational assessments. For all three assessments, Session I consisted of 35 operational SR items. Each form also contained 12 linking items (sets M1 and M2) as shown in Figure 2.3. In addition to the linking items, additional items were field tested in the embedded test positions. Each of the three content areas field tested an additional 12 forms' worth of new items.

2.3 Test Blueprints

The test blueprint specifies the relative percentages of items in each high-level content strand. This document helps ensure that each strand is represented by the minimum number of points (8) for student score reports.

Riverside Publishing content experts worked with DESE to develop blueprints for each course before item writing began in fall 2007. Blueprint development was guided by the Missouri Show-Me Standards. In the summer of 2010, a decision was made to remove the WP and PE items from the MO EOC Assessments. The test blueprints were revised to reflect this change.

Tables 2.2 through 2.4 outline the test construction blueprints with WP and PE items for English II, Algebra I, and Biology. Tables 2.5 through 2.7 show the blueprints used for the assessments without WP and PE items.

Table 2.2: Test Construction Blueprint for English II with Writing Prompt

Big Idea	Target # of Points	Point Range*	Target % Total	Minimum Emphasis	Maximum Emphasis
1. Develop and apply skills and strategies to the reading process	12	10–14	31%	26%	36%
2. Develop and apply skills and strategies to comprehend, analyze, and evaluate fiction, poetry, and drama	9	8–11	23%	23%	28%
3. Develop and apply skills and strategies to comprehend, analyze, and evaluate nonfiction	9	8–11	23%	23%	28%
Writing					
1. Apply a writing process in composing text					
2. Compose well-developed text					
3. Write effectively in various forms and types of writing	9	8–9	23%	23%	23%
Total	39		100%		

Note: Total score points for each content strand may vary depending on which passages are selected for a particular administration. The percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

*The minimum number of points in each strand will be 8.

This blueprint was built under the following assumptions:

1. The operational test will be composed of two sessions. Session I will have thirty-five (35) 1-point selected response items, and Session II will have one (1) 4-point WP.
2. The reading passages will generally be balanced between nonfiction and fiction. A slight imbalance may occur if an odd number of passages appears on the operational test.
3. Content strand 1 has a larger percentage of total points because it can be assessed using both fiction and nonfiction passages.
4. The writing form/type will vary depending on the WP selected for a particular administration. Writing prompts will be aligned to a primary CLE; however, multiple writing CLEs may be assessed to reflect the holistic rubric.

Table 2.3: Test Construction Blueprint for Algebra I with Performance Event

Content Strand	Target # of Points	10% Tolerance	Point Range	Target % Total	Minimum Emphasis	Maximum Emphasis
Number and Operations	8	0.8	7–9	21%	19%	23%
Algebraic Relationships	23	2.3	21–25	58%	53%	63%
Measurement*	0	0	0	0%	0%	0%
Data and Probability	8	0.8	7–9	21%	19%	23%
Total	39			100%		

Note: Total score points for the operational tests may vary depending on which PE is selected for a particular administration. Regardless of the total score points on a particular operational test, the percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

*These strands are not included on the EOC Assessment but are assessed locally.

This blueprint was built under the following assumptions:

1. The operational test will be composed of two sessions. Session I will have thirty-five (35) 1-point selected response items, and Session II will have one (1) 4-point PE item.
2. Each PE will be aligned to one CLE from the Algebraic Relationships strand.

Table 2.4: Test Construction Blueprint for Biology with Performance Events

Content Strand	Target # of Points	10% Tolerance	Point Range	Target % Total	Minimum Emphasis	Maximum Emphasis
Strand 1: Properties and Principles of Matter and Energy*	0	0	0	0%	0%	0%
Strand 2: Properties and Principles of Force and Motion*	0	0	0	0%	0%	0%
Strand 3: Characteristic and Interactions of Living Organisms	22	2.2	20–24	40%	36%	44%
Strand 4: Changes in Ecosystems and Interactions of Organisms with Their Environments	13	1.3	12–14	24%	22%	27%
Strand 5: Processes and Interactions of the Earth’s Systems (Geosphere, Atmosphere, and Hydrosphere)*	0	0	0	0%	0%	0%
Strand 6: Composition and Structure of the Universe and the Motion of the Objects Within It*	0	0	0	0%	0%	0%
Strand 7: Scientific Inquiry	20	(2.0)	20	36%	36%	36%
Strand 8: Impact of Science, Technology, and Human Activity*	0	0	0	0%	0%	0%
Total:	55			100%		

Note: Total score points for the operational tests may vary depending on which PE prompts are selected for a particular administration. Regardless of the total score points on a particular operational test, the percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

*These strands are not included on the EOC Assessment but are assessed locally.

This blueprint was built under the following assumptions:

1. The operational test will be composed of two sessions. Session I will have thirty-five (35) 1-point selected response items, and Session II will have one (1) 20-point performance task that is made up of a main context and several prompts.
2. Prompts within PEs will be aligned to CLEs from strand 7 only.

Table 2.5: Test Construction Blueprint for English II without Writing Prompt

Big Idea	Target # of Points	Point Range	Target % Total	Minimum Emphasis	Maximum Emphasis
1. Develop and apply skills and strategies to the reading process	12	10–14	34%	29%	40%
2. Develop and apply skills and strategies to comprehend, analyze, and evaluate fiction, poetry, and drama	9	8–11	26%	23%	31%
3. Develop and apply skills and strategies to comprehend, analyze, and evaluate nonfiction	9	8–11	26%	23%	31%
Writing					
1. Apply a writing process in composing text					
2. Compose well-developed text					
3. Write effectively in various forms and types of writing	5	5	14%	14%	14%
Total	35		100%		

Note: Total score points for each content strand may vary depending on which passages are selected for a particular administration. The percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

This blueprint was built under the following assumptions:

1. The operational test will be composed of thirty-five (35) 1-point selected response items.
2. The reading passages will generally be balanced between nonfiction and fiction. A slight imbalance may occur if an odd number of passages appears on the operational test.
3. Content strand 1 has a larger percentage of total points because it can be assessed using both fiction and nonfiction passages.

Table 2.6: Test Construction Blueprint for Algebra I without Performance Event

Content Strand	Target # of Points	10% Tolerance	Point Range	Target % Total Points	Minimum Emphasis	Maximum Emphasis
Number and Operations	8	0.8	7–9	23%	20%	26%
Algebraic Relationships	19	1.9	17–21	54%	49%	60%
Geometric and Spatial Relationships*	0	0	0	0%	0%	0%
Measurement*	0	0	0	0%	0%	0%
Data and Probability	8	0.8	7–9	23%	20%	26%
Total	35			100%		

*These strands are not included on the EOC Assessment but are assessed locally.

The percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

The operational test will be composed of thirty-five (35) 1-point selected response items.

Table 2.7: Test Construction Blueprint for Biology without Performance Events

Content Strand	Target # of Points	10% Tolerance	Point Range	Target % Total	Minimum Emphasis	Maximum Emphasis
Strand 1: Properties and Principles of Matter and Energy*	0	0	0	0%	0%	0%
Strand 2: Properties and Principles of Force and Motion*	0	0	0	0%	0%	0%
Strand 3: Characteristic and Interactions of Living Organisms	22	2.2	20–24	63%	57%	69%
Strand 4: Changes in Ecosystems and Interactions of Organisms with Their Environments	13	1.3	12–14	37%	34%	40%
Strand 5: Processes and Interactions of the Earth’s Systems (Geosphere, Atmosphere, and Hydrosphere)*	0	0	0	0%	0%	0%
Strand 6: Composition and Structure of the Universe and the Motion of the Objects Within It*	0	0	0	0%	0%	0%
Strand 7: Scientific Inquiry	0	0	0	0%	0%	0%
Strand 8: Impact of Science, Technology, and Human Activity	0	0	0	0%	0%	0%
Total:	35			100%		

*These strands are not included on the EOC Assessment but are assessed locally.

The percentage of total score points from each content strand (emphasis) will fall within the blueprint range described above.

The actual 2008–2011 English II, Algebra I, and Biology point distributions for each form fell within the blueprint targets.

2.4 Test Specifications

Standard 1.6¹ specifically addresses the appropriateness of test content and its relationship to a solid validity argument. Additionally, Standard 3.3² defines “test specifications” and provides examples of the type of information that should be included in a specification document. The test specifications describe the content and format of the test and delineate the ideal number of items and points assessed for each Course-Level Expectation (CLE). This section details the development and use of the test specification documents for the MO EOC Assessments.

¹ **Standard 1.6:** When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified (p. 18).

² **Standard 3.3:** The test specifications should be documented, along with their rationale and the process by which they were developed. The test specifications should define the content of the test, the proposed number of items, the item formats, the desired psychometric properties of the items, and the item and section arrangement. They should also specify the amount of time for testing, directions to the test takers, procedures to be used for test administration and scoring, and other relevant information (p. 43).

In 2007, Riverside Publishing content experts developed draft test specifications for each course. These draft test specifications were subsequently reviewed and approved by DESE. The specifications were finalized in fall 2007, before the development of items for field-test forms.

The test specification document served as the foundation for all test item development. The material in the test specifications was designed for use by Riverside Publishing content experts and DESE to construct tests containing the following items:

- Aligned to Missouri CLEs
- Aligned to Norman Webb’s depth of knowledge (DOK) cognitive levels
- SR and PE/WPs
- Standalone and passage-based

Detailed descriptions of the test content measured in English II, Algebra I, and Biology are presented in the following sections.

2.4.1 English II

The English II MO EOC Assessment measures students’ achievement in reading and writing. Session I of the test contains commissioned passages that comprise both fiction and nonfiction and cover a wide range of genres, including poems, short stories, newspaper articles, historical fiction, functional texts, and webpages. The questions associated with each passage are in SR format. There are 35 SR items on the English II Assessment. The English II EOC Assessment also contains standalone SR items that assess grammar and language usage. In previous years, Session II of the English II EOC Assessment contained a WP. However, Session II was either not scored or not administered during the 2010-2011 administrations.

Table 2.8 contains targets for the CLE point distribution on the English II operational forms for Session I only. Some of the CLE point targets may not be met because the use of a passage or scenario is not conducive to items written to the CLE. Some Big Ideas are not represented in this chart because they are not assessed at this course level.

Table 2.9 contains actual point distributions for the Summer 2010, Fall 2010, and Spring 2011 operational forms.

Table 2.8: Target Point Distributions for the English II Operational Forms

READING STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Develop and apply skills and strategies to the reading process	C. Phonics	*Apply decoding strategies to “problem-solve” unknown words when reading when needed		Assessed locally
	D. Fluency	*Read grade-level instructional text a. with fluency: accuracy, comprehension and appropriate expression b. adjusting reading rate to difficulty and type of text		Assessed locally
	E. Vocabulary	Develop vocabulary through text, using a. roots and affixes b. context clues c. glossary, dictionary, and thesaurus	2	3–4
	F. Prereading	*Apply prereading strategies to aid comprehension a. access prior knowledge b. preview c. predict with text support or rationale d. set a purpose and rate for reading		Assessed locally
	G. During Reading	*During reading, utilize strategies to a. determine meaning of unknown words b. self-monitor comprehension c. question the text d. infer e. visualize f. paraphrase g. summarize	2	Assessed locally
	H. Post Reading	Apply post-reading skills to comprehend, interpret, analyze, and evaluate text: *a. question to clarify *b. reflect c. draw conclusions d. paraphrase e. summarize	3	5–6
	I. Making Connections	Compare, contrast, analyze and evaluate connections: a. text to text (information and relationships in various fiction and nonfiction works) *b. text to self (text ideas and own experiences) *c. text to world (text ideas and the world by analyzing and evaluating the relationship between literature and its historical period and culture)	3	1–2

Table 2.8: Target Point Distributions for the English II Operational Forms (continued)

READING STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
2. Develop and apply skills and strategies to comprehend, analyze, and evaluate fiction, poetry and drama from a variety of cultures and times	A. Text Features	Analyze and evaluate the text features in grade-level text	3	2–3
	B. Literary Techniques	Identify and explain literary techniques, emphasizing a. understatement b. parallelism c. allusion d. analogy e. analyze and evaluate literary techniques previously introduced	3	2–3
	C. Literary Elements	Use details from text(s) to a. demonstrate comprehension skills previously introduced b. analyze character, plot, setting, point of view c. analyze the development of a theme across genres d. identify and analyze tone	3	4–5
3. Develop and apply skills and strategies to comprehend, analyze, and evaluate nonfiction (such as biographies, newspapers, technical manuals) from a variety of cultures and times	A. Text Features	Explain, analyze, and evaluate the author’s use of text features to clarify meaning	3	1–2
	B. Literary Techniques	Identify, explain, and analyze literary techniques in nonfiction, emphasizing a. understatement b. parallelism c. allusion d. analogy e. figurative language and sound devices previously introduced	3	2–3
	C. Text Structures	Use details from informational and persuasive text(s) to a. analyze and evaluate the organizational patterns b. identify and analyze faulty reasoning and unfounded inferences c. evaluate proposed solutions d. evaluate for accuracy and adequacy of evidence e. evaluate effect of tone on the overall meaning of work f. analyze and evaluate point of view g. analyze and evaluate author’s viewpoint/perspective h. demonstrate comprehension skills previously introduced	3	3–4

Table 2.8: Target Point Distributions for the English II Operational Forms (continued)

READING STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
	D. Understanding Directions	*Read and apply multistep directions to perform complex procedures and/or tasks	2	Assessed locally
WRITING STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Apply a writing process in composing text	A. Writing Process	*Apply a writing process to write effectively in various forms and types of writing (W3A)		Assessed locally
	A. Audience and Purpose	*Compose text: a. showing awareness of audience b. choosing a form and point of view appropriate to purpose and audience		Assessed locally
2. Compose well-developed text	B. Ideas and Content	*Compose text with a. strong controlling idea b. relevant specific details c. complex ideas d. freshness of thought		Assessed locally
	C. Organization and Sentence Structure	*Compose text with a. effective beginning, middle, and end b. a logical order c. effective paragraphing d. cohesive devices e. varied sentence structure f. clarity of expression g. active voice		Assessed locally
	D. Word Choice	*Compose text using a. precise and vivid language b. writing techniques such as imagery, humor, voice, and figurative language		Assessed locally
	E. Conventions	In written text, apply a. conventions of capitalization b. conventions of punctuation c. standard usage	1	5

Table 2.8: Target Point Distributions for the English II Operational Forms (continued)

WRITING STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
3. Write effectively in various forms and types of writing	A. Forms/Types/Modes of Writing	Compose a variety of texts a. using narrative, descriptive, expository, and/or persuasive features b. in various formats, including workplace communication c. including summary d. including literary analysis e. including reflective writing		Assessed locally

* These CLEs are locally assessed.

Table 2.9: Actual Point Distributions for the 2010-2011 English II Operational Forms

Reporting Categories	Summer 2010		Fall 2010		Spring 2011	
	Target	Actual	Target	Actual	Target	Actual
Reading Process	12	10	12	11	12	10
Reading (fiction)	9	11	9	10	9	9
Reading (nonfiction)	9	9	9	9	9	11
Writing	5	5	5	5	5	5
Total Items/Points	35	35	35	35	35	35

2.4.2 Algebra I

The Algebra I EOC Assessment measures students' ability to solve problems by applying mathematical concepts. There are three strands assessed on the Algebra I Assessment:

- Numbers and Operations
- Algebraic Relationships
- Data and Probability

The 35 selected response questions in Session I are aligned to the strands listed above. In previous years, Session II of the Algebra I EOC Assessment contained a PE. However, Session II was either not scored or not administered during the 2010-2011 administrations.

Table 2.10 contains targets for the CLE point distribution on the Algebra I operational forms. Some Big Ideas are not represented in this table because they are not assessed at this course level. Table 2.11 contain actual point distributions for the Summer 2010, Fall 2010, and Spring 2011 Algebra I operational forms.

Table 2.10: Target Point Distributions for the Algebra I Operational Forms

NUMBERS AND OPERATIONS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Understand numbers, ways of representing numbers, relationships among numbers, and number systems	A. Read, write, and compare numbers	Compare and order rational and irrational numbers, including finding their approximate locations on a number line	1	3–5
	B. Represent and use real numbers	Use real numbers and various models, drawing, etc. to solve problems	3	3–5
	C. Compose and decompose numbers	*Use a variety of representations to demonstrate an understanding of very large and very small numbers	2	Assessed locally
2. Understand meanings of operations and how they relate to one another	B. Describe effects of operations	*Describe the effects of operations, such as multiplication, division, and computing powers and roots on the magnitude of quantities	2	Assessed locally
	D. Apply operations on real and complex numbers	*Apply operations to real numbers, using mental computation or paper-and-pencil calculations for simple cases and technology for more complicated cases	2	Assessed locally
3. Compute fluently and make reasonable estimates	D. Estimate and justify solutions	*Judge the reasonableness of numerical computations and their results	3	Assessed locally
	E. Use proportional reasoning	*Solve problems involving proportions	2	Assessed locally

Table 2.10: Target Point Distributions for the Algebra I Operational Forms (continued)

ALGEBRAIC RELATIONSHIPS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Understand patterns, relations, and functions	B. Create and analyze patterns	Generalize patterns using <u>explicitly</u> or <u>recursively</u> defined functions	2	1–2
	C. Classify objects and representations	Compare and contrast various forms of <u>representations</u> of patterns	3	1–2
	D. Identify and compare functions	Understand and compare the properties of <u>linear</u> and <u>nonlinear functions</u>	2	2–3
	E. Describe the effects of parameter changes	Describe the effects of <u>parameter changes</u> on <u>linear</u> , <u>exponential growth/decay</u> , and <u>quadratic functions</u> including intercepts	2	2–3
2. Represent and analyze mathematical situations and structures using algebraic symbols	A. Represent mathematical situations	Use <u>symbolic algebra</u> to represent and solve problems that involve linear and quadratic relationships, including equations and inequalities	3	2–3
	B. Describe and use mathematical manipulation	Describe and use algebraic manipulations, including factoring and rules of integer exponents, and apply <u>properties of exponents</u> , including order of operations, to simplify expressions	2	2–3
	C. Use equivalent forms	Use and solve equivalent forms of equations (linear, absolute value, and quadratic)	2	1–2
	D. Use systems	Use and solve systems of linear equations or inequalities with 2 variables	2	1–2
3. Use mathematical models to represent and understand quantitative relationships	A. Use mathematical models	Identify quantitative relationships and determine the type(s) of functions that might model the situation to solve the problem	2	3–4
4. Analyze change in various contexts	A. Analyze change	Analyze linear and quadratic functions by investigating rates of change, intercepts, and zeros	3	3–4

Table 2.10: Target Point Distributions for the Algebra I Operational Forms (continued)

GEOMETRIC AND SPATIAL RELATIONSHIPS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Analyze characteristics and properties of two- and three-dimensional geometric shapes and develop mathematical arguments about geometric relationships	B. Apply geometric relationships	*Apply geometric properties such as similarity and angle relationship to solve multistep problems in two dimensions	2	Assessed locally
4. Use visualization, spatial reasoning, and geometric modeling to solve problems	B. Draw and use visual models	*Draw or use <u>visual models</u> to represent and solve problems	3	Assessed locally
MEASUREMENT STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
2. Apply appropriate techniques, tools, and formulas to determine measurements	D. Analyze precision	*Describe the effects of operations, such as multiplication, division, and computing powers and roots, on magnitudes of quantities and the effects of computation on <u>precision</u> , which include the judging of reasonable numerical computations <u>and their results</u>	2	Assessed locally
	E. Use relationships within a measurement system	*Use <u>unit analysis</u> to solve problems	2	Assessed locally
DATA AND PROBABILITY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them	A. Formulate questions	Formulate questions and collect data about a characteristic, which include <u>sample spaces</u> and <u>distributions</u>	3	1–2
	C. Represent and interpret data	Select and use appropriate graphical representation of data and, given <u>one-variable quantitative data</u> , display the distribution and describe its shape	3	2–3

Table 2.10: Target Point Distributions for the Algebra I Operational Forms (continued)

DATA AND PROBABILITY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
2. Select and use appropriate statistical methods to analyze data	A. Describe and analyze data	Apply statistical measures of center to solve problems	2	2–3
	C. Represent data algebraically	Given a scatterplot, determine an equation for a <u>line of best fit</u>	2	1–2
3. Develop and evaluate inferences and predictions that are based on data	A. Develop and evaluate inferences	Make <u>conjectures</u> about possible relationships between 2 characteristics of a sample on the basis of scatterplots of the data	3	2–3

* These CLEs are locally assessed.

Table 2.11: Actual Point Distributions for the 2010-2011 Algebra I Operational Forms

Reporting Categories	Summer 2010		Fall 2010		Spring 2011	
	Target	Actual	Target	Actual	Target	Actual
Number and Operation	8	9	8	8	8	8
Algebraic Relationships	19	17	19	19	19	19
Data and Probability	8	9	8	8	8	8
Total Items/Points	35	35	35	35	35	35

2.4.3 Biology

The Biology EOC Assessment measures students' achievement in the following content and process strands:

- Characteristics and Interactions of Living Organisms
- Changes in Ecosystems and Interactions of Organisms with Their Environments

The 35 SR questions in Session I are aligned to the two strands listed above. In previous years, Session II of the Biology EOC Assessment contained a PE. However, Session II was either not scored or not administered during the 2010-2011 administrations.

Table 2.12 is used as a target for the CLE point distribution for the Biology operational forms. Some Big Ideas are not represented in this table because they are not assessed at this course level. Table 2.13 contains actual point distributions for the Summer 2010, Fall 2010, and Spring 2011 operational forms.

Table 2.12: Target Point Distributions for the Biology I Operational Forms

CHARACTERISTICS AND INTERACTIONS OF LIVING ORGANISMS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. There is a fundamental unity underlying the diversity of all living organisms	B. Organisms progress through life cycles unique to different types of organisms	a. Recognize cells both increase in number and differentiate, becoming specialized in structure and function, during and after embryonic development	1	1–2
	C. Cells are the fundamental units of structure and function of all living things	b. Describe the structure of cell parts (e.g., cell wall, cell membrane, cytoplasm, nucleus, chloroplast, mitochondrion, ribosome, vacuole) found in different types of cells (e.g., bacterial, plant, skin, nerve, blood, muscle) and the functions they perform (e.g., structural support, transport of materials, storage of genetic information, photosynthesis and respiration, synthesis of new molecules, waste disposal) that are necessary to the survival of the cell and organism	2	1–2
2. Living organisms carry out life processes in order to survive	A. The cell contains a set of structures called organelles that interact to carry out life processes through physical and chemical means	c. Explain physical and chemical interactions that occur between organelles (e.g., nucleus, cell membrane, chloroplast, mitochondrion, ribosome) as they carry out life processes	2	1–2
	B. Photosynthesis and cellular respiration are complementary processes necessary to the survival of most organisms on Earth	a. Explain the interrelationship between the processes of photosynthesis and cellular respiration (e.g., recycling of oxygen and carbon dioxide), comparing and contrasting photosynthesis and cellular respiration reactions (Do NOT assess intermediate reactions.)	2	1–2
	B. Photosynthesis and cellular respiration are complementary processes necessary to the survival of most organisms on Earth	b. Determine what factors affect the processes of photosynthesis and cellular respiration (i.e., light intensity, availability of reactants, temperature)	2	1–2

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

CHARACTERISTICS AND INTERACTIONS OF LIVING ORGANISMS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
2. Living organisms carry out life processes in order to survive	D. Cells carry out chemical transformations that use energy for the synthesis or breakdown of organic compounds	*a. Summarize how energy transfer occurs during photosynthesis and cellular respiration as energy is stored in and released from the bonds of chemical compounds (i.e., ATP)	2	Assessed locally
	D. Cells carry out chemical transformations that use energy for the synthesis or breakdown of organic compounds	*b. Relate the structure of organic compounds (e.g., proteins, nucleic acids, lipids, carbohydrates) to their role in living systems		Assessed locally
	D. Cells carry out chemical transformations that use energy for the synthesis or breakdown of organic compounds	*d. Explain how protein enzymes affect chemical reactions (e.g., the breakdown of food molecules, growth and repair, regulation)	1	Assessed locally
	F. Cellular activities and responses can maintain stability internally while external conditions are changing (homeostasis)	a. Explain the significance of the selectively permeable membrane to the transport of molecules	2	1–2
	F. Cellular activities and responses can maintain stability internally while external conditions are changing (homeostasis)	b. Predict the movement of molecules across a selectively permeable membrane (i.e., diffusion, osmosis, active transport) needed for a cell to maintain homeostasis given concentration gradients and different sizes of molecules	2	1–2
	F. Cellular activities and responses can maintain stability internally while external conditions are changing (homeostasis)	c. Explain how water is important to cells (e.g., is a buffer for body temperature, provides a soluble environment for chemical reactions, serves as a reactant in chemical reactions, provides hydration that maintains cell turgidity, maintains protein shape)	2	1–2

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

CHARACTERISTICS AND INTERACTIONS OF LIVING ORGANISMS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
3. There is a genetic basis for the transfer of biological characteristics from one generation to the next through reproductive processes	B. All living organisms have genetic material (DNA) that carries hereditary information	a. Describe the chemical and structural properties of DNA (e.g., DNA is a large polymer formed from linked subunits of four kinds of nitrogen bases; genetic information is encoded in genes based on the sequence of subunits; each DNA molecule in a cell forms a single chromosome) (Assess the concepts; do NOT memorize the nitrogen base pairs.)	1	1–2
	B. All living organisms have genetic material (DNA) that carries hereditary information	b. Recognize the DNA codes for proteins, which are expressed as the heritable characteristics of an organism.	1	1–2
	B. All living organisms have genetic material (DNA) that carries hereditary information	a. Identify possible external causes (e.g., heat, radiation, certain chemicals) and effects of DNA mutations (e.g., altered proteins which may affect chemical reactions and structural development)	2	1–2
	C. Chromosomes are components of cells that occur in pairs and carry hereditary information from one cell to daughter cells and from parent to offspring during reproduction	a. Recognize the chromosomes of daughter cells, formed through the processes of asexual reproduction and mitosis, the formation of somatic (body) cells in multicellular organisms, are identical to the chromosomes of the parent cell	1	1–2
	C. Chromosomes are components of cells that occur in pairs and carry hereditary information from one cell to daughter cells and from parent to offspring during reproduction	b. Recognize that during meiosis, the formation of sex cells, chromosomes are reduced to half the number present in the parent cell	1	1–2

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

CHARACTERISTICS AND INTERACTIONS OF LIVING ORGANISMS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
3. There is a genetic basis for the transfer of biological characteristics from one generation to the next through reproductive processes	C. Chromosomes are components of cells that occur in pairs and carry hereditary information from one cell to daughter cells and from parent to offspring during reproduction	c. Explain how fertilization restores the diploid number of chromosomes	2	1–2
	D. There is heritable variation within every species of organism	a. Describe the advantages and disadvantages of asexual and sexual reproduction with regard to variation within a population	2	1–2
	D. There is heritable variation within every species of organism	*c. Recognize that new heritable characteristics can only result from new combinations of existing genes or from mutations of genes in an organism’s sex cells	1	Assessed locally
	E. The pattern of inheritance for many traits can be predicted by using the principles of Mendelian genetics	b. Predict the probability of the occurrence of specific traits, including sex-linked traits, in an offspring by using a monohybrid cross	2	1–2
CHANGES IN ECOSYSTEMS AND INTERACTIONS OF ORGANISMS WITH THEIR ENVIRONMENTS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Organisms are interdependent with one another and with their environment	A. All populations living together within a community interact with one another and with their environment in order to survive and maintain a balanced ecosystem	a. Explain the nature of interactions between organisms in predator/prey relationships and different symbiotic relationships (i.e., mutualism, commensalism, parasitism)	1	1–3

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

CHANGES IN ECOSYSTEMS AND INTERACTIONS OF ORGANISMS WITH THEIR ENVIRONMENTS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Organisms are interdependent with one another and with their environment	A. All populations living together within a community interact with one another and with their environment in order to survive and maintain a balanced ecosystem	b. Explain how cooperative (e.g., symbiotic) and competitive (e.g., predator/prey) relationships help maintain balance within an ecosystem	2	1–2
	B. Living organisms have the capacity to produce populations of infinite size, but environments and resources are finite	a. Identify and explain the limiting factors (biotic and abiotic) that may affect the carrying capacity of a population within an ecosystem	2	1–3
	D. The diversity of species within an ecosystem is affected by changes in the environment, which can be caused by other organisms or outside processes	a. Predict the impact (beneficial or harmful) a natural environmental event (e.g., forest fire, flood, volcanic eruption, avalanche) or human caused change (e.g., acid rain, global warming, pollution, deforestation, introduction of an exotic species) may have on the diversity of different species in an ecosystem	2	1–2
	D. The diversity of species within an ecosystem is affected by changes in the environment, which can be caused by other organisms or outside processes	*b. Predict the impact (beneficial or harmful) a natural or human caused environmental event (e.g., forest fire, flood, volcanic eruption, avalanche, acid rain, global warming, pollution, deforestation, introduction of an exotic species) may have on the biodiversity of a community	2	Assessed locally
2. Matter and energy flow through the ecosystem	A. As energy flows through the ecosystem, all organisms capture a portion of that energy and transform it to a form they can use	c. Predict how the use and flow of energy will be altered due to changes in a food web	2	1–2

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

CHANGES IN ECOSYSTEMS AND INTERACTIONS OF ORGANISMS WITH THEIR ENVIRONMENTS STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
3. Genetic variation sorted by the natural selection process explains evidence of biological evolution	A. Evidence for the nature and rates of evolution can be found in anatomical and molecular characteristics of organisms and in the fossil record	*b. Evaluate the evidence that supports the theory of biological evolution (e.g., fossil records, similarities between DNA and protein structures, similarities between developmental stages of organisms, homologous and vestigial structures)	2	Assessed locally
	B. Reproduction is essential to the continuation of every species	*a. Define a species in terms of the ability to mate and produce fertile offspring	1	Assessed locally
	B. Reproduction is essential to the continuation of every species	b. Explain the importance of reproduction to the survival of a species (i.e., the failure of a species to reproduce will lead to extinction of that species)	1	1–2
	C. Natural selection is the process of sorting individuals based on their ability to survive and reproduce within their ecosystem	a. Identify examples of adaptations that may have resulted from variations favored by natural selection (e.g., long-necked giraffes, long-eared jack rabbits) and describe how that variation may have provided populations an advantage for survival	2	1–2
	C. Natural selection is the process of sorting individuals based on their ability to survive and reproduce within their ecosystem	c. Explain how environmental factors (e.g., habitat loss, climate change, pollution, introduction of non-native species) can be agents of natural selection	2	1–2

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

SCIENTIFIC INQUIRY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Science understanding is developed through the use of science process skills, scientific knowledge, scientific investigation, reasoning, and critical thinking	A. Scientific inquiry includes the ability of students to formulate a testable question and explanation, and to select appropriate investigative methods in order to obtain evidence relevant to the explanation	a. Formulate testable questions and hypotheses	3	Assessed locally
	A. Scientific inquiry includes the ability of students to formulate a testable question and explanation, and to select appropriate investigative methods in order to obtain evidence relevant to the explanation	b. Analyzing an experiment, identify the components (i.e., independent variable, dependent variables, control of constants, multiple trials) and explain their importance to the design of a valid experiment	3	Assessed locally
	A. Scientific inquiry includes the ability of students to formulate a testable question and explanation, and to select appropriate investigative methods in order to obtain evidence relevant to the explanation	c. Design and conduct a valid experiment	4	Assessed locally
	A. Scientific inquiry includes the ability of students to formulate a testable question and explanation, and to select appropriate investigative methods in order to obtain evidence relevant to the explanation	d. Recognize it is not always possible, for practical or ethical reasons, to control some conditions (e.g., when sampling or testing humans, when observing animal behaviors in nature)	2	Assessed locally

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

SCIENTIFIC INQUIRY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Science understanding is developed through the use of science process skills, scientific knowledge, scientific investigation, reasoning, and critical thinking	A. Scientific inquiry includes the ability of students to formulate a testable question and explanation, and to select appropriate investigative methods in order to obtain evidence relevant to the explanation	g. Evaluate the design of an experiment and make suggestions for reasonable improvements	3	Assessed locally
	B. Scientific inquiry relies upon gathering evidence from qualitative and quantitative observations	b. Measure length to the nearest millimeter, mass to the nearest gram, volume to the nearest milliliter, force (weight) to the nearest Newton, temperature to the nearest degree Celsius, time to the nearest second	1	Assessed locally
	B. Scientific inquiry relies upon gathering evidence from qualitative and quantitative observations	c. Determine the appropriate tools and techniques to collect, analyze, and interpret data	2	Assessed locally
	B. Scientific inquiry relies upon gathering evidence from qualitative and quantitative observations	d. Judge whether measurements and computation of quantities are reasonable	2	Assessed locally
	B. Scientific inquiry relies upon gathering evidence from qualitative and quantitative observations	e. Calculate the range, average/mean, percent, and ratios for sets of data	1	Assessed locally

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

SCIENTIFIC INQUIRY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Science understanding is developed through the use of science process skills, scientific knowledge, scientific investigation, reasoning, and critical thinking	C. Scientific inquiry includes evaluation of explanations (laws/principles, theories/models) in light of evidence (data) and scientific principles (understandings)	a. Use quantitative and qualitative data as support for reasonable explanations (conclusions)	3	Assessed locally
	C. Scientific inquiry includes evaluation of explanations (laws/principles, theories/models) in light of evidence (data) and scientific principles (understandings)	b. Analyze experimental data to determine patterns, relationships, perspectives, and credibility of explanations (e.g., predict/extrapolate data, explain the relationship between the independent and dependent variable)	3	Assessed locally
	C. Scientific inquiry includes evaluation of explanations (laws/principles, theories/models) in light of evidence (data) and scientific principles (understandings)	c. Identify the possible effects of errors in observations, measurements, and calculations, on the validity and reliability of data and resultant explanations (conclusions)	3	Assessed locally
	C. Scientific inquiry includes evaluation of explanations (laws/principles, theories/models) in light of evidence (data) and scientific principles (understandings)	d. Analyze whether evidence (data) and scientific principles support proposed explanations (laws/principles, theories/models)	3	Assessed locally

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

SCIENTIFIC INQUIRY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
1. Science understanding is developed through the use of science process skills, scientific knowledge, scientific investigation, reasoning, and critical thinking	D. The nature of science relies upon communication of results and justification of explanations	<p>A. Communicate the procedures and results of investigations and explanations through:</p> <ul style="list-style-type: none"> • oral presentations • drawings and maps • data tables (allowing for the recording and analysis of data relevant to the experiment such as independent and dependent variables, multiple trials, beginning and ending times or temperatures, derived quantities) • graphs (bar, single, and multiple line) • equations and writings 	3	Assessed locally
	D. The nature of science relies upon communication of results and justification of explanations	c. Explain the importance of the public presentation of scientific work and supporting evidence to the scientific community (e.g., work and evidence must be critiqued, reviewed, and validated by peers; needed for subsequent investigations by peers; results can influence the decisions regarding future scientific work)	2	Assessed locally

Table 2.12: Target Point Distributions for the Biology I Operational Forms (continued)

IMPACT OF SCIENCE, TECHNOLOGY AND HUMAN ACTIVITY STRAND				
Big Idea	Concept	CLE	DOK Limit	Range of Points per CLE on the Operational Test
2. Historical and cultural perspectives of scientific explanations help to improve understanding of the nature of science and how science knowledge and technology evolve over time	B. Scientific theories are developed based on the body of knowledge that exists at any particular time and must be rigorously questioned and tested for validity	*a. Identify and describe how explanations (laws/principles, theories/models) of scientific phenomena have changed over time as a result of new evidence (e.g., cell theory, theories of spontaneous generation and biogenesis, theories of extinction, evolution theory, structure of the cell membrane, genetic theory of inheritance)	2	Assessed locally
	B. Social, political, economic, ethical and environmental factors strongly influence, and are influenced by, the direction of progress of science and technology	*e. Analyze and evaluate the drawbacks (e.g., design constraints, unintended consequences, risks), benefits, and factors (i.e., social, political, economic, ethical, and environmental) affecting progress toward meeting major scientific and technological challenges (e.g., limitations placed on stem-cell research or genetic engineering, introduction of alien species, deforestation, bioterrorism, nuclear energy, genetic counseling, use of alternative energies for carbon fuels, use of pesticides)	3	Assessed locally

* These CLEs are locally assessed.

Table 2.13: Actual Point Distributions for the 2010-2011 Biology Operational Forms

Reporting Categories	Summer 2010		Fall 2010		Spring 2011	
	Target	Actual	Target	Actual	Target	Actual
Characteristics and Interactions of Living Organisms Total	22	22	22	22	22	22
Changes in Ecosystems and Interactions of Organisms with Their Environments Total	13	13	13	13	13	13
Total Items/Points	35	35	35	35	35	35

2.5 Development of Test Items

Content-related evidence of validity supporting test interpretation is presented in terms of how the 2008–2011 MO EOC Assessments were assembled for English II, Algebra I, and Biology. Detailed information regarding both item-development procedures and content coverage is included in this section.

The forms for the Fall 2008 through Summer 2011 administrations were constructed using items that were field tested in spring 2008 or spring 2009. During the process of building the forms for the 2008–2011 operational test administrations, statistical characteristics (i.e., *p*-values and point-biserial correlations) were monitored to ensure that the statistical properties of the forms were similar within each content area and across operational test forms for fall, spring, and summer.

Riverside Publishing Test Development Specialists (TDSs) created a detailed item and passage development plan based on the blueprints for each content area. The plans included the number of items necessary for each assessable CLE, as well as an outline of the review process for developed items and passages. This process included internal Riverside Publishing reviews, a DESE review on a percentage of the items, and a content and bias review by Missouri educators.

2.5.1 Item Writing

Missouri educators, DESE staff members, Regional Instructional Facilitators, and Riverside Publishing TDSs created all the test items, including the PEs. English II passages and WPs were developed by item writers trained by Riverside Publishing, Riverside Publishing TDSs, and DESE staff. These passages were developed and refined prior to the item-writing workshops. Requirements to be an item writer included experience in classroom teaching and expert content knowledge.

In September 2007 and June 2008, Riverside Publishing conducted item-writing workshops to develop SR items for English II, Algebra I, and Biology as well as PEs for Algebra I and Biology. These workshops were conducted at the Assessment Resource Center (ARC) in Columbia, Missouri. Participants in the workshops included Missouri educators, DESE staff, Regional Instructional Facilitators, and Riverside Publishing TDSs. The workshops were held over a five-day period and were conducted with 15–20 teacher participants per content area. Teacher participants were selected by DESE to represent school districts throughout Missouri. The content developed at the workshops was based on the Missouri Show-Me Standards and CLEs.

The English II participants wrote SR items associated with the passages that had been developed prior to the item-writing workshops. The Algebra I and Biology participants wrote SR items and PEs along with rubrics. Biology PEs consist of a science investigation scenario and several associated constructed response items. The Biology PE items were written based on an existing science PE development template that specified the types of tasks and numbers of items that compose a PE.

During the item-writing workshops, Riverside Publishing TDSs conducted training sessions with the item writers and provided instructions on avoiding bias and stereotyping of groups and individuals on the basis of gender, race, ethnicity, religion, age, language, socioeconomic group, and disability. Riverside Publishing TDSs also trained item writers

to write items that adhere to the principles of universal design, making the items accessible to the widest range of students. For example, items and passages were written using clear and concise language, and all art, graphs, and tables were labeled and were not overly crowded with extraneous information. Instruction was also provided on developing items at particular cognitive levels based on Norman Webb’s Depth of Knowledge (DOK) levels.

Riverside Publishing TDSs trained item writers to enter content into the company’s electronic Content Management System (CMS). During training, each item writer wrote several items and received feedback on them. Participants also received feedback through CMS, as Riverside Publishing TDSs responded to teachers’ items as they were submitted. As items were produced, they were continuously reviewed, revised, edited, and evaluated by Riverside Publishing TDSs and DESE staff. Item writers who generated high-quality work on or ahead of schedule were given additional assignments.

As items were written, they were tracked according to the item development plan. Riverside Publishing kept careful records to maintain a workflow that generated items in assessment strands and CLEs as required by the test blueprint. All items and passages went through several rounds of internal reviews, including content and editorial reviews. Riverside Publishing TDSs reviewed each item with respect to alignment, clarity, and correspondence with item specifications.

2.5.2 Universal Design

Riverside Publishing TDSs are experienced in employing the principles of universal design in item development so that all students have equal access to the assessments. Riverside Publishing included these principles when training Missouri teachers to write the test items.

According to the *NCEO Synthesis Report 44* (Thompson, Johnstone, and Thurlow, 2002), there are seven elements of universally designed assessments:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, nonbiased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

All items for the MO EOC Assessments were developed with these elements in mind. Riverside Publishing ensured the development of MO EOC items in accordance with these principles in the following manner:

- Items were developed to include a wide array of contexts and cultures. These item types may make students feel more included, increase motivation, and avoid bias.
- The test and item specifications served as a model for precisely defining the constructs that the tests would measure. These specifications indicated to the item

- writer, content reviewer, and TDS exactly what was to be measured. The item could assess a particular part of a standard or a combination of elements within a standard. The reviews served as a method for eliminating items that included assessment of knowledge outside the standard. For example, a mathematics item should have nonmathematical vocabulary below grade level; otherwise, the item might also be assessing reading ability, introducing construct-irrelevant variance.
- The review of items, which included Missouri teachers from diverse ethnic and geographic backgrounds, served to ensure that all items were accessible to as many students as possible.
 - Riverside Publishing staff members trained Missouri teachers to create clear and simple instructions so that students would have a clear understanding of the task needed to answer an item. Teacher review committees had an opportunity to review the instructions to ensure that they were appropriate for the grade levels and subject areas. To ensure the appropriateness of the level of the vocabulary, *Children’s Writer’s Word Book* and *EDL Core Vocabulary* were employed by test developers and item review committees.
 - Finally, items with text, art, tables, maps, and diagrams were constructed with maximum legibility.

2.5.3 Content and Bias Review Process

Standard 3.6³ specifically addresses the importance of item review by both an examination of the item statistics and the use of expert panels of judges. This section details the steps that were taken to ensure that the items chosen for the operational forms of the MO EOC Assessments were of high technical quality and were free from bias. Content and bias reviews were conducted in November 2007 and July 2008 in Columbia, Missouri. The content review committees included DESE staff, Missouri educators from around the state, Regional Instructional Facilitators, and Riverside Publishing staff.

The content and bias review committees reviewed selected response items, PEs, and WPs using the following criteria:

- Overall quality and syntactical clarity
- Content coverage and content appropriateness
- Alignment to the specified CLE
- Appropriate contexts
- One clearly correct answer and plausible distractors for selected response items
- Freedom from bias or any racial, socioeconomic, gender, or other sensitivity issues

³ **Standard 3.6:** The type of items, response formats, scoring procedures, and test administration procedures should be selected based on the purposes of the test, the domain to be measured, and the intended test takers. To the extent possible, test content should be chosen to ensure that intended inferences from test scores are equally valid for members of different groups of test takers. The test review process should include empirical analyses and, when appropriate, the use of expert judges to review items and response formats. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented (p. 44).

Before reviewing the items, a group training session was held with all committee members. Riverside Publishing presented a PowerPoint that described the MO EOC program, the test development process, and the content and bias review procedures. After the large-group session, the committee members went to their respective break-out rooms to discuss the week's activities in more detail. The committee members were provided with copies of the CLEs and Item Specifications for the courses for the items they were to review. Each Riverside Publishing content facilitator reviewed these documents with the committee and answered any questions. The committee members were given the following checklists that could be referenced throughout the review process:

For SR items:

- Does the item assess the assigned CLE?
- Is the item clear, concise, and complete?
- Does the item contain accurate and sufficient content information?
- Is the item grade-level appropriate, and are the vocabulary and syntax appropriate for the students at the intended grade? (Reference the *EDL Core Vocabularies*.)
- Is the item fair to all students and free of bias and sensitivity issues?
- Does the item have correct punctuation, and is it grammatically correct?
- Is the item free from spelling and typographical errors?
- Is clueing avoided within an item stem and options, as well as among items?
- Does the item stand alone? (The answer to one item should not be dependent on the content of another item.)
- Are the equations, tables, charts, graphs, and other art clear, accurate, and necessary?
- Does the item have only one correct answer?
- Does the item have unique, plausible distractors containing common errors students would make?
- Are all the options parallel in form and arranged in logical order?
- Do all distractors contain clear rationale statements? (Math and Science only)
- Is the item free from absolutes (“none of the above,” “all of the above”) as options and free from the use of negatives (“not,” “none,” “except”) in the stem?
- Does the item avoid repeating words from the stem in the options?
- Does the item pose a single problem (although the solution may require more than one step)?

For PE/WP items:

- Does the item assess the assigned CLE?
- Does the item clearly specify how the student should respond?

- Does the item allow for a variety of acceptable responses for the student to get full credit?
- Is the item grade-level appropriate, and are the vocabulary and syntax appropriate for the students at the intended grade? (Reference the *EDL Core Vocabularies*.)
- Is the item rich enough to elicit an appropriate range of responses covering all possible score points?
- Is the item fair to all students and free of bias and sensitivity issues?
- Does the rubric clearly define an acceptable answer or answers at each score point level?

Twenty Missouri educators participated in the review process for each content area. The committee members read and reviewed each item. Discussions were held about whether the items met the criteria listed above. The committees then rejected or revised any items they deemed unsatisfactory. If there was disagreement about how to proceed with an item, the Riverside Publishing facilitator polled the group and followed the direction of the majority. Approximately 95% of the items were accepted (as-is or with edits) by the content and bias committees. Table 2.25 shows the number of items that were reviewed in 2007. The accepted items were placed in a pool of items from which the 2008 standalone field-test forms were built. Table 2.26 shows the number of items that were reviewed in 2008 and placed on embedded field test forms in the Spring and Summer 2009 operational administrations.

To further preserve validity, all item review sessions were held in secure meeting rooms, and all materials were confidential. Committee members were required to sign confidentiality agreements so that the integrity of the test content was not compromised. Although educators were encouraged to share information with their colleagues about the process of the item review, they were made fully aware of the expectation that any information about specific items and passages was to remain secure and confidential.

Table 2.14: 2007 Content/Bias Item Review Acceptance Rates

	Total Number of Items Presented for Review	Number of Items Accepted (As Is or With Edits)	Acceptance Rate (Items Accepted As Is or With Edits)
English II	404	398	99%
Algebra I	239	233	97%
Biology	402	365	91%

Table 2.15: 2008 Content/Bias Item Review Acceptance Rates

	Total Number of Items Presented for Review	Number of Items Accepted (As Is or With Edits)	Acceptance Rate (Items Accepted As Is or With Edits)
English II	298	298	100%
Algebra I	288	288	100%
Biology	164	161	98%

2.6 Test Form Assembly

2.6.1 Field-Test Selection and Administration

The items accepted at the content/bias review were used to build the standalone field-test forms that were administered in spring 2008. Field-test items were selected so that each form met the established operational blueprint requirements for content coverage as closely as possible. For any standalone field-test form that deviated slightly from the blueprint, another field-test form made up for that difference so that the entire pool of field-tested items met the blueprint requirements.

The MO EOC Spring 2008 field test consisted of 10 SR forms per course, 10 English II WPs, 10 Algebra I PE forms, and 10 Biology PE forms. All field-test forms were reviewed and approved by DESE. They were administered to a sample of Missouri students in April 2008.

2.6.2 Statistical Item Review

After the 2008 field-test item scoring was completed, Riverside Publishing TDSs and psychometricians reviewed the statistical characteristics of the items. Riverside Publishing used classical item statistics, including n -counts, p -values, percentage choosing each response option, point-biserial correlations, and differential item functioning (DIF) analysis for the SR items. Additionally, the Rasch model was used for distractor analysis for the SR items and for DIF analysis for the PE/WP items.

During the data review, which was held June 6–8, 2008, Riverside Publishing Research and Test Development staff and DESE staff reviewed students' performance on the Spring 2008 field-test items. Items were carefully reviewed with respect to their statistical characteristics. Item reviewers from DESE and Riverside Publishing were provided with the following information:

- Form
- Position
- Item as it appeared in the printed books
- Item alignment to the Missouri Show-Me Standards
- The p -value of the correct answer and percentage of students who selected each distractor (for SR items only)
- Mean and SD of item score (for PE/WP items only)
- Point-biserial correlation of correct response and point-biserial for each distractor (for SR items only)
- Total number of students who attempted to answer each question
- DIF using the Mantel-Haenszel (MH) (1959) procedure and Educational Testing Service (ETS) classification (for SR items only)

Riverside Publishing and DESE staff reviewed items that were flagged because of statistics that fell outside the parameters determined by the Riverside Publishing Research staff. Table 2.27 contains the guidelines that were used for data review.

Table 2.16: Criteria for Flagged Items

Item Flagging Criteria	Indicates
If p -value of keyed response < 0.35	Difficult item
If p -value of keyed response > 0.95	Easy item
If p -value of keyed response $< p$ -value of distractor	Possible miskey
If p -value of distractor > 0.35	Possible second correct option
If point-biserial of keyed response < 0.20	Poorly discriminating item
If point-biserial of a distractor is > 0.00	Possible second correct option
If ETS classification is B or C (from DIF analysis)	Possible bias in item

Each flagged item was reviewed, and then Riverside Publishing and DESE made a decision about whether the item should be accepted or rejected. The review included items flagged with moderate to severe DIF (an ETS classification of B or C). A flagged item was accepted if the review team determined that the item was strong and tested students on content that they were expected to know. Accepted items were then made available in the pool of items that could be used to create the operational forms. Items that the review team felt were biased or inappropriate for the Missouri EOC Assessments were rejected. Rejected items were removed from the item pool, making them invalid for the MO EOC Assessments. Of the 690 total items reviewed, 91% were accepted.

2.6.3 Operational Test Selection and Administration

In June and July 2008, Riverside Publishing TDSs selected operational items for test forms for use in 2008 and 2009. Items for the 2009–2010 administration cycle were selected in June and July of 2009. Using item response theory (IRT) item difficulty values, four equivalent operational forms were selected for each content area. These four forms are the operational component of the Fall, Spring, and Summer EFT forms, as well as the released form. Three forms were selected for the 2009–2010 administrations. The Fall forms were administered in November 2008 and 2009, the Spring forms in April 2009 and 2010, and the Summer forms in June 2009 and 2010.⁴

The operational form construction process was based on content requirements and statistical criteria. The steps associated with assembling the test forms included the following:

1. *Determine form design.* Each form consists of operational items and embedded field-test items.
2. *Select items that meet content specifications.* Each form was constructed based on the test specifications for that content area. The test specifications delineate the item distribution across assessment strands. They also outline the test length, type of items, and number of points to be assessed at each CLE.
3. *Evaluate statistical specifications and select items to meet these specifications.* Spreadsheets (form matrices) are used to ensure that the test forms meet statistical specifications. These matrices contain the following statistics: average p -values,

⁴ The Summer 2009 administration is part of the 2009–2010 assessment year, and its results are not included in this technical report.

point-biserial correlations, and DIF statistics. Riverside Publishing psychometricians conducted a review of the test forms to ensure equivalence of test difficulty across forms.⁵

4. *Review and approve test forms.* Once the content and statistical specifications were met for each content area, the forms were reviewed and approved by MO DESE. The forms were then released for production and additional content and editorial reviews.

2.7 Braille and Large Print Versions

Beyond employing the principles of universal design, all operational assessments were offered in Braille and Large Print versions for visually impaired students taking the MO EOC Assessments. The Fall 2008 operational paper-and-pencil version was converted into Braille and Large Print to accommodate these students.

Once the Braille and Large Print forms were created, two separate reviews were held with educators from Missouri who had specialized training in working with visually impaired students.

The Large Print form review was held in Jefferson City, Missouri, at the DESE offices on September 29, 2008. The Braille review was held in St. Louis, Missouri, at the Missouri School for the Blind on October 10, 2008.

The teachers consulted the *Large Print and Braille Style Guide*, which was also used during form composition, and relied on their own expertise to determine whether changes to directions, passages, or items were needed, or whether items should be omitted. Riverside Publishing Braille vendor (Region IV) also reviewed the forms and made recommendations based on how items, passages, and directions would be transcribed to Braille.

Riverside Publishing and DESE reviewed the recommendations from all of these sources. It was determined that no items had to be omitted to accommodate Large Print students. For the Braille version of the form, one item from English II was removed because the content of the item prohibited transcription to Braille. Students taking the Braille form were given credit for this item. The embedded field test items were eliminated from both versions of these forms due to the irregular testing conditions and the small sample sizes for these groups. These versions of the Large Print and Braille forms were used through the Summer 2010 test administration.

2.8 Online Forms Construction

All items were written so that they could be presented in the online delivery system without any alterations.

2.9 Quality Control for Test Construction

Checklists and quality control procedures accompany each stage of form development. Following is a list of some quality control procedures used during the assembly of the MO EOC Assessment forms:

⁵ Rasch values were not available for all items when the 2008–2009 operational forms were built.

- Construct forms based on all content requirements noted in the test blueprint and test specifications.
- Verify correct number of items per standard or reporting category based on test blueprint.
- Review items to ensure a wide sampling of the knowledge and skills being measured.
- Ensure that all items have been through the appropriate review procedures and are approved for use by DESE.
- Check for a variety of item topics, equal distribution of males and females, ethnicities, etc.
- Verify appropriate portions of items with and without artwork.
- Check for clueing across all items on each form.
- Verify equal or nearly equal distribution of answer choices for SR items.
- Ensure that the test meets the required statistical specifications (i.e., that as many items as possible have p -values between .35 and .90 and as many items as possible have point-biserial correlations above .20).
- Consider any statistical flags or problems.
- Check statistics to ensure that the collection of items on a given form yields an overall difficulty that falls within the specified range.
- Verify that items have not been released to the public.
- Verify correct answer key for each item.
- Perform content review of form (senior staff).
- Perform statistical review of form (psychometrician/statistician).
- Send form to DESE for review and approval.

2.10 Summary

The MO EOC Assessments in English II, Algebra I, and Biology provide an indication of student progress toward achieving the knowledge and skills identified in the Missouri Show-Me Standards. Just as the Show-Me Standards guided the item development and selection process, the consideration of content played an equally important role in form development. Form development required a balance of both content coverage and item difficulty. As items were selected for inclusion on particular forms, every effort was made to balance the content coverage to ensure the items aligned to the Missouri Show-Me Standards/CLEs being assessed while simultaneously considering the overall difficulty of the forms.

CHAPTER 3: ACHIEVEMENT-LEVEL SETTING

3.1 Introduction

To provide full documentation for the Missouri End-of-Course (MO EOC) Assessments, important information from the 2009 standard-setting workshop is provided as the cut scores established at the 2009 standard-setting sessions were applied to 2010 test scores. Reports of past standard-setting activities provide continuity and will help readers more fully understand the MO EOC program and its impact on student learning.

One purpose of assessment is to establish clear guidelines for educational decision making. By assigning meaning to test scores, standard setting allows policymakers, administrators, teachers, parents, and students to make statements about the level of proficiency of individual students and groups of students. The purpose of this chapter is to provide documentation of the achievement-level-setting (or standard-setting) event conducted for the MO EOC Assessments on November 3, 4, and 5, 2008. These activities were undertaken for three MO EOC Assessments: English II, Algebra I, and Biology. These three assessments were administered operationally for the first time during the 2008–2009 school year.

3.2 Goal of the Standard Setting

The main goal of the standard-setting event was to establish three cut scores for each test in the MO EOC Assessments:

1. The cut score that differentiates Below Basic performance from Basic performance
2. The cut score that differentiates Basic performance from Proficient performance
3. The cut score that differentiates Proficient performance from Advanced performance

The determination of three cut scores yields four performance categories for each assessment.

3.3 Staff and Participants

Staff from Questar Assessment, Inc., a subcontractor to Riverside Publishing, planned and facilitated the standard-setting workshops. Questar's most experienced facilitators—Michael Beck, Sheila Potter, and Martha Caswell—served as facilitators for the workshops. Each of these individuals has facilitated standard-setting sessions for multiple clients for both elementary level and high school level assessments.

In addition to the staff from Questar, two psychometricians from Riverside Publishing attended the workshops. Their function was to enter panelist data, produce tables and reports, and oversee data quality control. A Riverside Publishing program manager was present for the entire workshop to assist Missouri Department of Elementary and Secondary Education (DESE) staff and the panelists with logistical issues. Content area specialists from Riverside Publishing's Content Development group were present in the three panel rooms to serve as

resources for content-related questions. Finally, curriculum staff from DESE attended the standard-setting workshops to serve as content resources to the appropriate panels.

A total of 46 panelists participated in the standard-setting workshop: 14 in English II, 15 in Algebra I, and 17 in Biology. One to three members of each panel had participated in an earlier achievement-level-setting workshop for other Missouri assessments. The significant majority of panelists had not been members of any of the assortment of committees for MO EOC development activities. More than half of each panel was made up of active classroom teachers in the relevant content area; several other panel members were other professional educators, such as administrators and curriculum coordinators. One or two members of each panel were business professionals with expertise in the relevant field. Lists of the standard-setting participants appear in Appendix A.

3.4 Development of Achievement-Level Descriptors

The MO EOC Assessments utilize the same achievement-level labels used for previous high school Missouri Assessment Program (MAP) assessments: Advanced, Proficient, Basic, and Below Basic. For each of these levels, the achievement-level descriptor (ALD) describes the specific knowledge and skills that a student at that level is able to demonstrate. As suggested by Missouri's Technical Advisory Committee (TAC), DESE conducted sessions devoted to developing draft ALDs prior to the standard-setting workshop.

Riverside Publishing staff used the ALDs from the Missouri Assessment Program (MAP) and the End-of-Course Course-Level Expectations (CLEs) to create the initial draft ALDs for English II. This document was then reviewed and approved by DESE. This content was used as a template for the Algebra I and Biology ALDs. This template was provided to committees comprised of people representing higher education in Missouri. These committees reviewed the CLEs and added skills associated with each level. These ALDs were reviewed and approved by DESE.

At the standard-setting workshop, participants devoted a significant portion of time to fine-tuning the draft ALDs for each assessment. The facilitators provided the panelists with draft copies of the appropriate ALDs, copies of the MO EOC Assessment blueprint, and the appropriate CLEs. Using these materials as references and drawing on the expertise of the panelists, the Questar facilitators led each panel in an extended discussion and exercise to refine and elaborate each of the ALDs. Once this activity was complete, the panels relied on the resulting ALDs as a reference during the actual standard-setting activities. In addition, the panelists were allowed to make appropriate, though generally minor, revisions and refinements to the ALDs during and after the standard-setting activities.

3.5 Overview of Standard-Setting Activities

3.5.1 Methodology and Data Considerations

The specific methodology used for the standard-setting activities was a modified Angoff procedure, as recommended by the state's TAC. The Angoff procedure and its modifications are well-recognized and heavily researched methods for establishing student performance standards for tests such as the EOC. Prior Missouri standard-setting

workshops utilized an item-mapping procedure commonly known as Bookmark standard setting; however, that method requires placing the items in a difficulty-ordered item book, which necessitates that the item difficulty parameters be known. In the case of the MO EOC Assessments, because the Spring 2009 operational assessment had not yet been administered at the time of the standard-setting workshop, parameter estimates for the operational test forms were not available. The modified Angoff method does not require placing the items in difficulty order; it was, therefore, a suitable choice of methods for this event.

Consistent with the methods used for prior MAP standard-setting events, the modified Angoff method allows three distinct rounds of panelist judgments. Between the first and second rounds, Riverside Publishing provided the panelists with item-difficulty data for their consideration. Because operational data were not available in November, the item data were derived from the 2008 field test. Panelists were appropriately cautioned about the limitations of such data. (For more information about the similarity of the data from the Spring 2008 standalone field test and the Spring 2009 operational assessment, the reader is referred to Section 7.3.3 of this technical report.)

Before the last round of judgments, Riverside Publishing staff provided the panelists with statewide impact data for the assessment. These data were intended to serve as an anchor for the panelists' recommendations. Again, because actual performance data were not available, the data were based on projected statewide score distributions generated from the field test. It is likely that a standalone field test would produce lower-than-expected results due to decreased student effort; therefore, Riverside Publishing psychometricians would consider the field test data "lower-bound" estimates of actual student performance in an operational event. As with the item-level data estimates, the facilitators cautioned the panelists about relying too much on these impact data.

Despite the limitations of the field test data for the standard-setting activities, Riverside Publishing psychometricians believed that providing panelists with even tentative data was desirable, both to mirror procedures used for establishing standards for previous Missouri assessments and to provide panelists with an "external reality check" on their evolving recommendations. Missouri TAC discussions confirmed the use of these statewide impact data.

In addition to the caveats about item level and impact data, panel facilitators clearly communicated to the panelists that the results of their standard-setting activities would be purely advisory to DESE. DESE would consider the recommendations and select the final cut scores for each assessment.

3.5.2 Description of the Test Forms and Considerations

DESE chose to use the MO EOC Spring 2009 operational forms for the standard-setting event. These forms were selected from the several available operational forms because they would be the most widely used in the 2008–2009 test administration year.

The MO EOC Assessments are comprised of selected response (multiple choice) items and a Performance Event/Writing Prompt (PE/WP). Each English II and Algebra I form includes a single PE/WP worth 4 points. On the Biology assessment, the Performance

Event consists of 11 open-ended items, each worth between 0 and 4 points (for a total of 20 points).

3.6 Specific Standard-Setting Activities

The Standard-Setting Session Agenda provided a general guide regarding the time devoted to each activity. This agenda is included as Appendix B. Questar facilitators held closely to the times contained in the agenda. They used identical processes, including presentation slides and scripts, across all sessions to minimize any intersession differences related to facilitator or session variance.

The following sections provide details about the processes that Questar and Riverside Publishing followed during the course of the standard-setting workshop.

3.6.1 General Process Overview

The first 90 minutes of the three-day session served as an introduction and overview to the general standard-setting processes. First, Michael Muenks, Coordinator of Curriculum and Assessment for DESE, oriented the panelists to the MO EOC program and briefly outlined the session purpose and intended outcomes.

Next, Michael Beck of Questar led a brief general overview, “What Is Standard Setting?” Its purpose was to ensure a common understanding of the fundamental elements of the process. Mr. Beck included a brief overview of the general process of establishing student performance standards, ground rules for panelist activities, and some key elements for the panelists to focus on when attempting to set standards. Mr. Beck also advised the panelists that their work was advisory to DESE. This introduction was a high-level overview of the standard-setting process; individual facilitators provided more detail about each step in the process after the panels broke into content-specific groups. The PowerPoint slides presented during the opening session are included as Appendix C of this report.

Finally, Dr. Sheila Potter of Questar provided a general overview of ALDs and their importance to the standard-setting process. Since the panels would be reviewing, editing, and expanding on draft versions of the ALDs provided by the state, it was important for panelists to understand the critical role of ALDs in the standard-setting process. Following this activity, panelists divided into the three content-specific panel break-out rooms, where all remaining work for the sessions took place.

3.6.2 Panelists Take the Operational Assessments

After reconvening in the content-area panels, panelists first introduced themselves and signed DESE-provided confidentiality forms. Facilitators introduced themselves and reiterated the high-level standard-setting processes that Mr. Beck had discussed during the opening session. Facilitators then allowed the panelists time to take and score the Spring 2009 form of the operational assessment. For this activity, panelists had access to the test administration procedures, the actual test content, and all relevant scoring materials. Field test items that were included in these forms were removed from the test books seen by the panelists. Because these were “live” materials, facilitators stressed the confidentiality of all of the items.

The primary purpose of this activity was to familiarize panelists with the actual, complete assessment content prior to beginning the standard-setting judgments. Following this review of the tests, each panel spent a short time reacting to the assessment content: difficulty, sources of challenge, scoring issues, and general and specific reactions. This exercise provided the panelists, especially those not familiar with the MO EOC Assessments, with a context concerning the definition of *Proficient* as conveyed by the assessments.

3.6.3 Panelists Discuss and Fine-Tune the ALDs

All three panels began this activity with a review of the draft ALDs for the particular content area. Separate panels of Missouri educators had developed these draft ALDs during DESE-led sessions several weeks earlier. The ALD review activity was highly interactive, with panelists suggesting changes and other refinements—both substantive and editorial—to the draft ALDs. The ultimate task was to operationalize specific behaviors indicating performance at the Advanced, Proficient, Basic, and Below Basic levels in the content area. Panel suggestions were discussed until consensus was reached and were then recorded on the draft ALDs, a copy of which was given to each panelist or placed on chart paper displayed around the room. Panelists were later able to refer to these pages, along with the original drafts, during the actual judgment activities. The thoroughness of the ALD refinement activities and the extent to which the panelists, individually and as a group, internalized the ALDs significantly impacted the soundness of the subsequent standard-setting activities. For this reason, approximately two hours was devoted to this activity in the session agenda.

At the conclusion of the standard-setting sessions, DESE collected the panelist recommendations for ALD revisions for consideration in the wording of the final ALDs. Appendix D contains a copy of the draft ALDs that were distributed to the panelists at the outset of the standard-setting workshop. Appendix E contains a copy of the final ALDs.

3.6.4 Orientation to the Modified Angoff Procedures

After the ALD activity was complete, facilitators oriented the panels to the specific tasks involved with the modified Angoff standard-setting process. The modified Angoff process requires panelists to read and make judgments about each successive item in the test book using the following procedures. When reading an item, panelists were to consider the item's importance in the context of the underlying CLE, the task(s) required of the student, and the item's difficulty. They were to decide what percentage of minimally Proficient students should be able to answer the item correctly. Panelists were then to decide what percentage of minimally Advanced students would answer the item correctly. Finally, they were to decide what percentage of minimally Basic students would answer the item correctly. (While the MO EOC Assessments contain four levels of student performance, cuts are made at only three locations on the score distribution.) The panelists were instructed to consider their judgments in this order—Proficient, Advanced, and Basic—as it anchors the item judgments on the most important cut, Proficient. In addition, once panelists made their judgment for the Proficient students, they had a clearer, more defined range of values to consider for the other two cuts.

For the constructed response item(s), panelists were to consider the average item score of minimally Proficient, minimally Advanced, and minimally Basic students. In other words, judgments for the constructed response items were made as whole-point values (i.e., 1, 2, 3, etc.) rather than as percentages of students answering correctly.

The facilitators included the following important points in their presentations:

- Panelists should focus on the *threshold* of performance in each category.
- Panelists should review and recall what each performance descriptor means.
- Panelists should focus on MO EOC students statewide, not just in the school or district in which they work.

Finally, the facilitators explained that the panelists' judgments should be made independently and anonymously and that security of the testing materials should be maintained at all times.

The steps outlined in Sections 3.6.1 through 3.6.4 composed the activities of the first day of the workshop. The second day began with an overview of the previous day's activities and outcomes, after which panelists took a five-item multiple choice qualifying test concerning the standard-setting procedures they were about to use. A copy of this instrument is provided as Appendix F. This instrument was used to ensure that all panelists understood the importance of the ALDs and selected elements of the modified Angoff procedure before beginning the process of making item judgments.

Next, the facilitators led their panels in a practice exercise using the modified Angoff rating procedures. The practice test contained five multiple choice items and one performance item (with score points comparable to the performance items on the operational test). The practice items were released Grade 4 NAEP Mathematics items selected to distance the practice exercise from the content area and grade levels of the EOC Assessments. During this exercise, panelists had the opportunity to practice the mechanical aspects of the modified Angoff judgment process and procedures for recording their recommendations before beginning work on the real test. Additionally, the practice test allowed the facilitators to check the panelists' understanding of the mechanics of the technique and corresponding recording of judgments. Facilitators allowed the panelists about 20 minutes to complete the practice activity. The practice judgments were reviewed on a group basis by discussing the range of judgments made about each item.

Following completion of the practice exercise, facilitators asked the panelists to complete and sign a form indicating that they understood the information they had received and discussed and that they felt prepared to make their Round 1 judgments. All panelists so indicated.

3.6.5 Round 1 Judgments

At this point, panelists were ready to make their Round 1 judgments. This work was completed anonymously (via judge numbers known only to Riverside Publishing staff) and independently. Panelists indicated their judgments on specially designed scannable rating sheets developed for each content area. These rating sheets contained three fields for each test item: one for Basic, one for Proficient, and one for Advanced. For the

selected response items, each field contained a set of bubbles corresponding to the percentage of students expected to choose the correct answer. As panelists made their judgments for each item, facilitators instructed them to “bubble in” one value for each achievement level. In other words, for Item 1, the panelist entered a number corresponding to the percentage of students expected to choose a correct answer at the minimally Basic level, a number for the minimally Proficient level, and a number for the minimally Advanced level. Panelists were constrained to choosing multiples of 5 (i.e., 5%, 10%, 15%, etc.), as they appeared on the Rating Forms. Panelists then followed this same procedure for all the remaining selected response items. For the PE/WP items, each field contained bubbles corresponding to the various point values possible for the item. Facilitators instructed the panelists to make a judgment about how many points a borderline student at each achievement level would score on that item (in other words, how many points a Basic student would score, how many points a Proficient student would score, and how many points an Advanced student would score).

Most panelists completed their first round of judgments within 60 minutes; however, there was no time limit for this activity, and some panelists required 90 minutes to complete their judgments. This is not unusual for the first round of judgments in a modified Angoff workshop; often some panelists are still struggling to understand the task at this point, thus requiring more time to make their judgments. After panelists completed their judgments, they turned in their rating sheets and were excused for a lunch break.

3.6.6 Feedback and Discussion of Round 1 Judgments

While the panelists were at lunch, the Riverside Publishing psychometricians prepared reports of the Round 1 judgment results. The afternoon session began with an overview of these reports. The first report was a table displaying all three raw score cuts as determined individually by each panelist’s judgments. This table also contained the entire panel’s average, median, highest, and lowest raw score cuts, as well as the standard deviation of all the panelists’ judgments for each of the three raw score cuts. The second report contained a frequency display of all three cut scores (Basic, Proficient, and Advanced) recommended by each panelist. This bar graph displayed all the panelists’ judgments on a single graph so that areas of dispersion or overlap in the raw cut scores would be apparent. These reports were anonymous; ID numbers, rather than names, were used to identify individual panelists.

Facilitators spent time reviewing these reports with the participants to ensure that everyone understood how to interpret the information contained in them. Using the Round 1 results, facilitators then led an extended discussion of the Round 1 judgments. This discussion focused primarily on the panelists’ judgments of individual items. Facilitators actively engaged all the panelists in the discussion to gauge whether they had indicated the item percentage values that they intended, that the reasoning processes they followed in making their judgments were consistent with good practice, and that the panelists clearly understood the mechanics of making item judgments. Throughout these discussions, facilitators focused on the key elements of the standard-setting process: establishing the *threshold* of each cut, projecting the cuts for a *statewide* population of these students, and focusing on the particular *course* and *performance level* of the target populations.

Much like a jury deliberation, this discussion also allowed the panelists to hear their peers' comments and rationales for their judgments. This phase took around two hours, depending on the session; facilitators permitted discussion to continue until they perceived that all panelists were prepared to make their second round of judgments.

Next, facilitators distributed statewide item difficulty data derived from the 2008 field test. For the selected response items, the derived item difficulties were item *p*-values. For the PE/WP, the item difficulties were average item scores. Facilitators advised the panelists that caution should be taken in interpreting the item difficulty data, since the data were collected during a standalone field test (and student motivation may not have been the same as it would be on an operational assessment). Facilitators also explained that these data were relevant, but not critical, to the process of setting standards.

Before making Round 2 judgments, panelists again signed a short form indicating that they understood the procedures and were prepared to make Round 2 recommendations.

3.6.7 Round 2 Judgments

During Round 2, panelists again worked independently to make judgments about the percentage of students at the threshold of each achievement level who would answer each item correctly. Facilitators explained to the panelists that they were free to maintain their Round 1 judgments or to revise them as they deemed appropriate. Before beginning this round of judgments, panelists were once more reminded of the key elements of the process and were asked to focus specifically on the ALDs for their assessment. Again, there was no time limit, although this round required significantly less time than did Round 1 because the panelists more clearly understood the judgment process. In addition, they were increasingly familiar with the specific items for which they were making the judgments. Further, many panelists had begun to formulate some or all of their Round 2 item judgments during the discussion of the Round 1 results.

After panelists completed their Round 2 judgments and recorded their recommendations on their rating sheets, they submitted the forms and were excused for the evening. After all rating sheets were collected, Riverside Publishing psychometricians prepared the reports of the Round 2 judgments.

3.6.8 Feedback and Discussion of Round 2 Judgments

When the panels convened on the third day of the standard-setting workshop, facilitators presented the results of the Round 2 judgments. The reports showing the Round 2 results were used to guide another discussion of specific items. The presentation and discussion at this stage were similar to, although more focused than, those following Round 1.

Following this discussion, facilitators provided panelists with estimated statewide impact data—that is, the percentages of students statewide whose performance would likely be labeled Below Basic, Basic, Proficient, or Advanced if the panels' Round 2 judgments were adopted. The panels' median Round 2 judgments were used to determine cut scores for this report. Again, facilitators advised the panelists that the impact data were relevant to, but not essential for, setting performance standards. (This cautionary information was especially important in the case of MO EOC, as the data were not grounded in an operational administration of the assessments.)

When facilitators were comfortable that all panelists were prepared to make their final recommendations, they proceeded to Round 3 of judgments.

3.6.9 Round 3 of Judgments, Meeting Evaluation, and Final Inspection of ALDs

For Round 3, the panelists' judgments consisted of one recommended cut score for each achievement level; panelists were not required to make item-level judgments. Panelists were given unlimited time to complete their Round 3 (final) recommendations, although most completed their judgments within 20 minutes. All panelists clearly understood that only the Round 3 judgments counted as their recommendations and that the three rounds were not combined in any way to form the proposed cuts.

After completing their final round of judgments, individual panelists were excused for lunch. Following the lunch break, facilitators provided the panels with the results of their Round 3 judgments.

Immediately following the final presentation of cut scores, panelists completed a written evaluation of the process. This evaluation covered the panelists' opinions of the adequacy of the training provided and their comfort with and confidence in their judgments on a round-by-round basis. The form also contained spaces for the panelists to write other comments concerning the workshop. A copy of this evaluation is included as Appendix G of this report.

After facilitators collected the panelist evaluations, they allowed the panels 30 to 40 minutes for a final review of the ALDs. During this time, panelists were allowed to discuss and, if necessary, fine-tune or revise the ALDs. Finally, panelists were thanked for their participation and dismissed.

3.7 Session Results by Panel and Round

Appendices H, I, and J contain the feedback reports by round for English II, Algebra I, and Biology, respectively. Selected data from these graphs and tables are summarized below for ease of cross-round and cross-content-area comparison.

The standard-setting literature typically considers the *median* recommendation to be the best indicator of a panel's judgment, as the median would not be impacted by the judgments of a few outlying panelists. In the case of this standard-setting event, as a review of Appendices H, I, and J indicates, all median and mean cut scores are within a single rounded raw-score point for all of the content areas. Therefore, the choice of a measure of central tendency for these particular panels would not markedly impact the resulting cut scores.

Table 3.1 contains the median recommended cut scores for all rounds and content areas. As data in Table 3.1 indicate, the panels did not markedly change their typical recommended cut scores across the three rounds of judgments. This is not to say that individual panelists made the same recommendations across rounds. In fact, across the nine sets of judgments between rounds (three content areas with three cut scores each), the mean change in median raw cut scores was -0.5 between Rounds 1 and 2, -1.0 between Rounds 2 and 3, and -1.5 between Rounds 1 and 3. (The median raw-score change between any pair of rounds was 0.) Though the mean changes were minimal from round to round, individual panelists changed their round-to-round recommendations by as

much as 17 raw-score points. Across all panels, the mean absolute value of raw cut score changes made was 1.5 between Rounds 1 and 2, 1.9 between Rounds 2 and 3, and 2.6 between Rounds 1 and 3.

Table 3.1: Median Recommended Cut Scores by Content Area and Round

Cut*	Content Area								
	English II			Algebra I			Biology		
	BB-B	B-P	P-A	BB-B	B-P	P-A	BB-B	B-P	P-A
Round 1	16.5	24.5	32	13	23	32	20	35	48
Round 2	16.5	25.5	33	13	23	31	19	34	46
Round 3/Final	15.5	24	33	13	22	31	18	32	45
No. Points Possible	39			39			55		

* BB = Below Basic; B = Basic; P = Proficient; A = Advanced

As is typically the case with standard-setting activities conducted over multiple rounds, the standard deviations of panelists' recommendations got smaller across rounds, indicating both an increasing level of panelist understanding of the process and increasing interpanel agreement based on group discussions between rounds of judgments. This is illustrated graphically through an examination of the frequency bar charts in Appendices H, I, and J, as well as statistically in the tabled results. The colored bars representing Basic, Proficient, and Advanced cut scores clearly become taller and more compact over each round of judgments. While panelists came closer to their peers in judging the most appropriate cut scores, even in Round 3—not unexpectedly—there was still a fair amount of spread in the recommended scores. That variability is especially notable in the Biology assessment; however, this assessment is significantly longer than the others, which may partially account for the larger Round 3 variability.

Standard errors of the median judgments were computed for all cut scores across all panels. In no case did the Round 3 standard error reach a whole raw-score unit. Most were lower than half of a raw-score point. This indicates that the final median judgments are highly stable. These standard errors are, of course, at least partially a function of the relatively small standard deviations and large panel sizes.

Table 3.2 summarizes the projected statewide percentages of students whose EOC scores will fall in each of the four performance categories. These data are based on the 2008 field test results and may be viewed as “lower-bound” estimates of the likely statewide results that will be obtained at the end of the 2008–2009 school year.

Table 3.2: Projected Statewide Percentages of Students Scoring in the Various Performance Categories on the EOC Assessments, 2008

Assessment	Performance Category*			
	Below Basic	Basic	Proficient	Advanced
English II	15%	31%	39%	16%
Algebra I	18%	38%	33%	11%
Biology	12%	39%	39%	10%

*Percentages may not sum to 100% due to rounding.

After the standard-setting event, the Missouri State Board of Education adopted the panels' median cut scores without revision.

3.8 Results of Participant Evaluations

Appendix K contains the data collected from panelists on the evaluation form. For the questions pertaining to the organization and adequacy of information provided in the opening session, the panelists generally provided ratings of 4 or 5 (on a scale of 1 to 5, with 5 being the highest). For the evaluation questions pertaining to the discussions of the achievement-level descriptors and the panelists' understanding of each of the ALDs after the discussions, in all cases at least 70% of the panelists provided ratings of 4 or 5. The questions pertaining to the panelists understanding of the judgment process and feedback on the results of each round received similar scores. Overall, these data indicate that the panelists generally understood what was expected of them, were comfortable with the process, and were comfortable with the resulting cut scores.

CHAPTER 4: ITEM ANALYSIS

4.1 Introduction

Item analyses were conducted for Missouri End-of-Course (MO EOC) Assessments in English II, Algebra I, and Biology. In this chapter, the summary information, which includes mean item score and discrimination indices, is presented at the item level for each content area. The item summary statistics presented in this section (p -values, point-biserial correlations, and omit rates) are based on the operational administrations that included responses from 1986 students for Summer 2010, 6965 students for Fall 2010, and 189,747 students for Spring 2011 across the three content areas. The differential item functioning (DIF) analyses are based upon data obtained when the item was field tested.

For selected response (SR) items, the mean score is simply the proportion of students who gave correct responses to the item (usually referred to as item difficulty, or p -value), and the discrimination index is the point-biserial correlation between the item score and the total score based on the remaining items.

4.2 Analysis of Forms for Each End-of-Course Assessment

Tables 4.1 through 4.9 summarize item difficulty, discrimination, and omit rates for the SR items that composed each assessment for the Summer 2010, Fall 2010, and Spring 2011 operational administrations. For SR items, the p -value is the proportion of students who answered the item correctly. The item discrimination, or corrected point-biserial correlation, is the correlation between students' item scores and their total scores on the remaining test items. Both item difficulty and item discrimination are expressed in the raw score metric.

Table 4.1: Item Statistics for English II, Summer 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.70	0.21	0.00
2	0.43	0.26	0.00
3	0.70	0.32	0.00
4	0.63	0.17	0.00
5	0.45	0.14	0.00
6	0.82	0.29	0.00
7	0.74	0.34	0.00
8	0.56	0.39	0.00
9	0.53	0.15	0.00
10	0.45	0.46	0.00
11	0.38	0.31	0.01
12	0.51	0.44	0.00
24	0.70	0.49	0.00
25	0.40	0.31	0.00
26	0.47	0.44	0.00
27	0.31	0.10	0.00
28	0.55	0.33	0.00
29	0.53	0.39	0.00
30	0.28	0.19	0.00
31	0.34	0.31	0.01
32	0.63	0.27	0.00
33	0.43	0.37	0.00
34	0.48	0.42	0.00
35	0.34	0.25	0.00
36	0.51	0.36	0.00
37	0.42	0.27	0.00
38	0.45	0.36	0.00
39	0.53	0.33	0.00
40	0.58	0.52	0.00
41	0.51	0.55	0.00
43	0.71	0.11	0.00
44	0.50	0.20	0.00
45	0.64	0.30	0.00
46	0.66	0.24	0.00
47	0.70	0.27	0.00

Table 4.2: Item Statistics for Algebra I, Summer 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.46	0.08	0.00
2	0.78	0.31	0.00
3	0.68	0.29	0.00
4	0.68	0.41	0.00
5	0.38	0.23	0.00
10	0.71	0.33	0.00
11	0.46	0.26	0.00
12	0.79	0.26	0.00
13	0.41	0.12	0.00
14	0.49	0.29	0.00
15	0.81	0.39	0.00
16	0.61	0.35	0.00
17	0.26	0.36	0.00
18	0.33	0.33	0.00
19	0.37	0.21	0.00
20	0.66	0.33	0.00
21	0.34	0.14	0.00
26	0.60	0.42	0.00
27	0.31	0.26	0.00
28	0.42	0.31	0.00
29	0.34	0.17	0.00
30	0.46	0.39	0.00
31	0.57	0.36	0.00
32	0.32	0.08	0.00
33	0.43	0.27	0.00
34	0.29	0.39	0.00
35	0.29	0.17	0.00
36	0.38	0.35	0.00
37	0.13	0.11	0.00
38	0.55	0.35	0.00
43	0.57	0.42	0.00
44	0.68	0.35	0.00
45	0.62	0.36	0.00
46	0.45	0.33	0.00
47	0.61	0.41	0.00

Table 4.3: Item Statistics for Biology, Summer 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.44	0.19	0.01
2	0.56	0.21	0.00
3	0.84	0.40	0.00
4	0.71	0.28	0.00
5	0.49	0.31	0.00
10	0.68	0.27	0.00
11	0.43	0.29	0.00
12	0.31	0.25	0.00
13	0.81	0.37	0.00
14	0.53	0.20	0.00
15	0.47	0.40	0.00
16	0.71	0.42	0.00
17	0.37	0.26	0.00
18	0.58	0.21	0.00
19	0.52	0.39	0.00
20	0.79	0.25	0.00
21	0.66	0.37	0.00
26	0.76	0.41	0.00
27	0.87	0.38	0.00
28	0.34	0.20	0.00
29	0.51	0.36	0.00
30	0.27	0.23	0.00
31	0.61	0.31	0.00
32	0.68	0.49	0.00
33	0.40	0.23	0.01
34	0.32	0.38	0.00
35	0.38	0.33	0.00
36	0.74	0.42	0.00
37	0.30	0.03	0.00
38	0.54	0.43	0.00
43	0.45	0.23	0.00
44	0.55	0.35	0.00
45	0.22	0.13	0.00
46	0.49	0.35	0.01
47	0.46	0.42	0.00

Table 4.4: Item Statistics for English II, Fall 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.66	0.44	0.00
2	0.69	0.32	0.00
3	0.84	0.31	0.00
4	0.73	0.43	0.00
5	0.82	0.34	0.00
6	0.33	0.10	0.00
7	0.82	0.39	0.00
8	0.65	0.49	0.00
9	0.46	0.25	0.00
10	0.65	0.36	0.00
11	0.73	0.52	0.00
12	0.53	0.30	0.00
24	0.55	0.35	0.00
25	0.48	0.30	0.00
26	0.80	0.45	0.00
27	0.46	0.46	0.00
28	0.61	0.33	0.00
29	0.46	0.15	0.00
30	0.39	0.37	0.00
31	0.71	0.45	0.00
32	0.64	0.50	0.00
33	0.40	0.33	0.00
34	0.59	0.34	0.00
35	0.80	0.44	0.00
36	0.84	0.46	0.00
37	0.57	0.34	0.00
38	0.50	0.37	0.00
39	0.76	0.44	0.00
40	0.58	0.39	0.00
41	0.63	0.50	0.00
43	0.81	0.41	0.00
44	0.50	0.33	0.00
45	0.75	0.30	0.00
46	0.77	0.21	0.00
47	0.67	0.37	0.00

Table 4.5: Item Statistics for Algebra I, Fall 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.89	0.30	0.00
2	0.73	0.22	0.00
3	0.71	0.50	0.00
4	0.70	0.35	0.00
5	0.68	0.56	0.00
10	0.79	0.36	0.00
11	0.67	0.40	0.00
12	0.72	0.46	0.00
13	0.59	0.48	0.00
14	0.44	0.41	0.00
15	0.65	0.50	0.00
16	0.67	0.51	0.00
17	0.80	0.41	0.00
18	0.71	0.54	0.00
19	0.60	0.65	0.00
20	0.59	0.38	0.00
21	0.54	0.45	0.00
26	0.57	0.53	0.00
27	0.50	0.49	0.00
28	0.61	0.43	0.00
29	0.55	0.50	0.00
30	0.55	0.32	0.00
31	0.56	0.50	0.00
32	0.48	0.47	0.00
33	0.45	0.50	0.00
34	0.50	0.52	0.00
35	0.63	0.43	0.00
36	0.41	0.48	0.00
37	0.57	0.54	0.00
38	0.39	0.41	0.00
43	0.32	0.40	0.00
44	0.59	0.37	0.00
45	0.46	0.22	0.00
46	0.16	0.26	0.00
47	0.35	0.45	0.00

Table 4.6: Item Statistics for Biology, Fall 2010 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.80	0.39	0.00
2	0.77	0.41	0.00
3	0.92	0.20	0.00
4	0.59	0.37	0.00
5	0.79	0.36	0.00
10	0.64	0.39	0.00
11	0.70	0.40	0.00
12	0.56	0.35	0.00
13	0.67	0.46	0.00
14	0.52	0.40	0.00
15	0.60	0.30	0.00
16	0.77	0.45	0.00
17	0.90	0.39	0.00
18	0.68	0.49	0.00
19	0.50	0.51	0.00
20	0.30	0.16	0.00
21	0.54	0.20	0.00
26	0.72	0.52	0.00
27	0.67	0.42	0.00
28	0.46	0.60	0.00
29	0.44	0.37	0.00
30	0.42	0.26	0.00
31	0.41	0.35	0.00
32	0.59	0.47	0.00
33	0.58	0.43	0.00
34	0.57	0.46	0.00
35	0.70	0.32	0.00
36	0.61	0.48	0.00
37	0.45	0.32	0.00
38	0.73	0.46	0.00
43	0.62	0.27	0.00
44	0.63	0.36	0.00
45	0.48	0.41	0.00
46	0.60	0.45	0.00
47	0.83	0.44	0.00

Table 4.7: Item Statistics for English II, Spring 2011 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.76	0.37	0.00
2	0.90	0.37	0.00
3	0.70	0.35	0.00
4	0.84	0.31	0.00
5	0.84	0.32	0.00
6	0.66	0.32	0.00
7	0.76	0.42	0.00
8	0.69	0.31	0.00
9	0.71	0.30	0.00
10	0.66	0.47	0.00
11	0.79	0.31	0.00
12	0.84	0.49	0.00
24	0.84	0.41	0.00
25	0.71	0.36	0.00
26	0.60	0.19	0.00
27	0.76	0.50	0.00
28	0.58	0.36	0.00
29	0.43	0.30	0.00
30	0.67	0.35	0.00
31	0.67	0.47	0.00
32	0.66	0.39	0.00
33	0.65	0.38	0.00
34	0.74	0.23	0.00
35	0.69	0.41	0.00
36	0.65	0.20	0.00
37	0.64	0.48	0.00
38	0.68	0.42	0.00
39	0.78	0.35	0.00
40	0.60	0.38	0.00
41	0.60	0.40	0.00
43	0.91	0.29	0.00
44	0.60	0.34	0.00
45	0.67	0.20	0.00
46	0.53	0.31	0.00
47	0.85	0.42	0.00

Table 4.8: Item Statistics for Algebra I, Spring 2011 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.46	0.23	0.00
2	0.73	0.40	0.00
3	0.54	0.49	0.00
4	0.73	0.40	0.00
5	0.45	0.47	0.00
10	0.61	0.37	0.00
11	0.69	0.52	0.00
12	0.84	0.35	0.00
13	0.69	0.41	0.00
14	0.68	0.43	0.00
15	0.54	0.10	0.00
16	0.16	0.22	0.00
17	0.37	0.26	0.00
18	0.82	0.45	0.00
19	0.68	0.40	0.00
20	0.84	0.24	0.00
21	0.64	0.36	0.00
26	0.59	0.41	0.00
27	0.54	0.38	0.00
28	0.66	0.38	0.00
29	0.29	0.28	0.00
30	0.63	0.37	0.00
31	0.66	0.47	0.00
32	0.29	0.31	0.00
33	0.88	0.32	0.00
34	0.80	0.38	0.00
35	0.52	0.45	0.00
36	0.72	0.52	0.00
37	0.83	0.30	0.00
38	0.44	0.39	0.00
43	0.67	0.32	0.00
44	0.89	0.18	0.00
45	0.31	0.29	0.00
46	0.68	0.37	0.00
47	0.60	0.31	0.00

Table 4.9: Item Statistics for Biology, Spring 2011 Operational Administration

Item	<i>p</i>-Value/Mean	Corrected Point-Biserial Correlation	Omit Rate %
1	0.83	0.42	0.00
2	0.65	0.31	0.00
3	0.84	0.27	0.00
4	0.81	0.32	0.00
5	0.72	0.28	0.00
10	0.71	0.40	0.00
11	0.46	0.35	0.00
12	0.67	0.21	0.00
13	0.71	0.38	0.00
14	0.39	0.28	0.00
15	0.60	0.44	0.00
16	0.81	0.33	0.00
17	0.52	0.41	0.00
18	0.90	0.39	0.00
19	0.64	0.28	0.00
20	0.64	0.35	0.00
21	0.24	0.27	0.00
26	0.75	0.49	0.00
27	0.88	0.45	0.00
28	0.68	0.45	0.00
29	0.59	0.38	0.00
30	0.84	0.36	0.00
31	0.75	0.24	0.00
32	0.57	0.19	0.00
33	0.52	0.22	0.00
34	0.49	0.18	0.00
35	0.51	0.33	0.00
36	0.68	0.40	0.00
37	0.86	0.39	0.00
38	0.72	0.27	0.00
43	0.69	0.38	0.00
44	0.54	0.41	0.00
45	0.71	0.32	0.00
46	0.53	0.33	0.00
47	0.60	0.26	0.00

4.3 Speededness

The consequence of time limits on examinees' scores is called speededness. A test is speeded if examinees taking it score lower than they would have had the test not been timed. Most speededness statistics are based on the number of items that were not attempted by students. For the purpose of this analysis, if a student did not attempt the last item on any of the separately timed subsections of the test, it was assumed that the student might not have reached the item because he or she ran out of time.

The MO EOC Assessments were not designed to be speeded tests. Rather, they were intended to be "power tests"; that is, all students were expected to have ample time to finish all items and prompts.

The last column in Tables 4.1 through 4.9 shows the percentage of students who omitted each SR item for each MO EOC Assessment. It is clear from the tables that the omit rates are negligible or zero for the majority of items.

4.4 Item Bias Statistics

Differential Item Functioning (DIF) occurs when an item has difficulty measures that vary across contexts for similarly able subgroups of examinees. Field test data from Spring 2008, 2009, and 2010 was examined for DIF with the Mantel-Haenszel (MH) (1959) procedure.

The Mantel-Haenszel method is a nonparametric approach to DIF. In the MH procedure, total raw scores are held constant while the odds ratio is estimated. In practice, the odds ratio is generally converted to the delta metric, and the Educational Testing Service (ETS) categorization is applied to flag the significance of DIF effects (Dorans and Holland, 1993).

With the groups matched on raw score, the comparable examinees can be placed in 2×2 tables of group by item response, where j equals the number of levels of the matching variable. For these analyses, j equals each observed score category of the k -item tests, with $j = 0, 1, 2, \dots, k$, then one 2×2 table for a given item with score category j can be represented as

	Correct	Incorrect	Total
Reference	y_j	x_j	m_j
Focal	y'_j	x'_j	m'_j
Total	n_j	n'_j	N_j

The Delta MH test statistic and variance have the following form:

$$DeltaMH = 2.35 \ln \frac{\left[\sum_{j=0}^K \frac{(y_j x'_j - y'_j x_j)}{N_j} \right]}{\sum_{j=0}^K \frac{y'_j x_j}{N_j}},$$

Where y_j , x_j , y'_j , and x'_j are the frequency counts of cells of the 2×2 tables, and N_j is the total n for the cells.

The critical values of the ETS categorizations are 1.00 and 1.50 on the delta scale for categories A, B, and C. Specifically, if the absolute value of delta is smaller than 1.00, the item is categorized as A. If the absolute value of delta is larger than or equal to 1.50, the item is classified as C. Otherwise, items are categorized as B. In both the A and C categories, statistical significance is set at the 5% level for a single item.

Results of the DIF analyses for the items contained in the Summer 2010, Fall 2010, and Spring 2011 operational administrations are summarized in Table 4.10. Tables 4.11 and 4.12 contain DIF statistics for the entire pool of MO EOC Assessment items.

Table 4.10: Frequency Distribution of DIF Categories by Item Type for the Summer 2010, Fall 2010, and Spring 2011 Operational Assessments

Test	Group**	Selected Response Items					
		A*	A-*	B*	B-*	C*	C-*
Summer 2010							
English II	M/F	33	0	1	1	0	0
	W/B	33	0	0	2	0	0
	W/H	35	0	0	0	0	0
Algebra I	M/F	34	0	0	1	0	0
	W/B	31	0	2	2	0	0
	W/H	32	0	1	2	0	0
Biology	M/F	35	0	0	0	0	0
	W/B	33	0	0	2	0	0
	W/H	35	0	0	0	0	0
Fall 2010							
English II	M/F	35	0	0	0	0	0
	W/B	34	0	1	0	0	0
	W/H	33	0	0	2	0	0
Algebra I	M/F	34	0	0	1	0	0
	W/B	34	0	0	1	0	0
	W/H	31	0	3	1	0	0
Biology	M/F	32	0	2	1	0	0
	W/B	34	0	0	1	0	0
	W/H	34	0	0	1	0	0
Spring 2011							
English II	M/F	34	0	1	0	0	0
	W/B	34	0	0	1	0	0
	W/H	32	0	1	2	0	0
Algebra I	M/F	33	0	0	2	0	0
	W/B	29	0	3	3	0	0
	W/H	32	0	0	2	1	0
Biology	M/F	33	0	1	1	0	0
	W/B	33	0	0	2	0	0
	W/H	32	0	2	1	0	0

Note: Classifications with a negative sign (“-”) favor the reference group, while classifications with no sign favor the focal group.

* DIF categories: A, negligible; B, slight to moderate; and C, moderate to severe.

* DIF contrast groups: M/F, male versus female; W/B, white versus black; and W/H, white versus Hispanic.

Table 4.11: Frequency Distribution of DIF Categories by Item Type for the Spring 2008 Field Test Items

Test	Group ^{***}	Selected Response Items [*]						PE/WP Items [*]					
		A ^{**}	A- ^{**}	B ^{**}	B- ^{**}	C ^{**}	C- ^{**}	A ^{**}	A- ^{**}	B ^{**}	B- ^{**}	C ^{**}	C- ^{**}
English II	M/F	281	0	8	7	1	3	7	0	3	0	0	0
	W/B	285	0	8	6	0	1	10	0	0	0	0	0
	W/H	285	0	5	10	0	0	10	0	0	0	0	0
Algebra I	M/F	178	0	0	2	0	0	10	0	0	0	0	0
	W/B	161	0	7	10	1	1	8	0	0	2	0	0
	W/H	167	0	9	4	0	0	10	0	0	0	0	0
Biology	M/F	173	0	3	4	0	0	105	0	0	0	0	0
	W/B	169	0	3	6	0	2	97	0	2	6	0	0
	W/H	169	0	5	6	0	0	104	0	0	1	0	0

Note: Classifications with a negative sign (“-”) favor the reference group, while classifications with no sign favor the focal group.

* The Mantel-Haenszel procedure is applied for the SR items and WINSTEPS for the PE/WP items.

** DIF categories: A, negligible; B, slight to moderate; and C, moderate to severe.

*** DIF contrast groups: M/F, male versus female; W/B, white versus black; and W/H, white versus Hispanic.

Table 4.12: Frequency Distribution of DIF Categories by Item Type for the Spring 2009 Field Test Items

Test	Group ^{***}	Selected Response Items [*]						PE/WP Items [*]					
		A ^{**}	A- ^{**}	B ^{**}	B- ^{**}	C ^{**}	C- ^{**}	A ^{**}	A- ^{**}	B ^{**}	B- ^{**}	C ^{**}	C- ^{**}
English II	M/F	365	0	5	12	0	2	1	0	7	0	2	0
	W/B	344	0	11	24	0	5	10	0	0	0	0	0
	W/H	365	0	7	9	1	2	10	0	0	0	0	0
Algebra I	M/F	277	0	2	9	0	0	24	0	0	0	0	0
	W/B	257	0	14	15	0	2	24	0	0	0	0	0
	W/H	271	0	9	7	1	0	23	0	0	1	0	0
Biology	M/F	279	0	4	4	0	1	172	0	2	3	0	2
	W/B	263	0	14	10	0	1	150	0	10	12	2	5
	W/H	274	0	7	6	0	1	172	0	4	2	1	0

Note: Classifications with a negative sign (“-”) favor the reference group, while classifications with no sign favor the focal group.

* The Mantel-Haenszel procedure is applied for the SR items and WINSTEPS for the PE/WP items.

** DIF categories: A, negligible; B, slight to moderate; and C, moderate to severe.

*** DIF contrast groups: M/F, male versus female; W/B, white versus black; and W/H, white versus Hispanic.

4.5 Summary

The item analyses provided in this chapter show that the MO EOC Assessments have sound psychometrics properties. For example, p -values show that MO EOC Assessment items measure achievement across a broad range of difficulty. Also, item discrimination values show that most items are appropriately correlated with the total test score and thus contribute to distinguishing between lower-performing and higher-performing students. In addition, very few students omitted items during testing. The low percentage of students omitting selected response items provides evidence that the test is a power test of the students' skills and not a speeded test. Finally, item bias statistics based on data from the 2008 standalone field test and the 2009 embedded field test administrations show the items to be generally free from statistical bias.

CHAPTER 5: TEST ADMINISTRATION

5.1 Introduction

This chapter contains information about the Missouri Department of Elementary and Secondary Education (DESE) and Riverside Publishing processes that ensure the standardized administration of the Missouri End-of-Course (MO EOC) Assessments. The *Standards* (AERA, APA, and NCME, 1999) state that, “For tests designed to assess the examinee’s knowledge, skills, or abilities, standardization helps to ensure that all examinees have the same opportunity to demonstrate their competencies” (p. 61). In other words, careful attention to the details of information dissemination, Test Examiner training, accommodations and modifications, and test security help ensure that students taking the EOC Assessments in different locations have equal opportunities for success.

The *EOC Test Coordinator’s Manual* and *Test Examiner’s Manual* contain detailed information about the testing guidelines, materials handling, and standardized administration instructions for the EOC Assessments. While those manuals are not included here, much of the information contained in this chapter can be found in them.

For the MO EOC Assessments, districts can choose either a paper-and-pencil or online delivery format. The *Test Coordinator’s Manual* and the *Online Test Examiner’s Manual* contain information specific to the registration for and administration of the online version of the MO EOC Assessments. Relevant information related to the online delivery, where it differs from the paper-and-pencil format, is included in this chapter.

5.2 Students for Whom the EOC Assessments Are Appropriate

The responsibility and authority for testing students in the Missouri EOC Assessments at the appropriate time in the course of instruction belongs to the local district. The EOC Assessments are based on Course-Level Expectations (CLEs) rather than on Grade-Level Expectations (GLEs). Therefore, when the content of the CLEs is covered in the local school district’s curriculum, the test may be administered regardless of student grade level or course name.

5.2.1 Students with Individualized Education Programs

A student with disabilities, as classified under the Individuals with Disabilities Education Act (IDEA), has an Individualized Education Program (IEP) that, in part, governs whether a particular assessment is appropriate for the student. In the case of the EOC Assessments, decisions about whether a student with a disability will participate in the assessments are made by the student’s IEP team and are documented in the IEP. All students must take the three Phase I EOC Assessments (English II, Algebra I, and Biology), plus the Government EOC Assessment from Phase II. If, however, a student’s disability qualifies him or her to take the MAP-Alternate Assessment (MAP-A), that student will not be required to participate in the EOC Assessment.

5.2.2 Students with Individual Accommodation Programs

Students with Individual Accommodation Programs (IAPs) are considered disabled under Section 504 of the 1973 Rehabilitation Act. These students are not served under IDEA

and are not documented with a particular designation for the EOC Assessment. However, professionals who are knowledgeable about a student’s disability and educational needs should make accommodation decisions for the student as they would for a student with an IEP.

5.2.3 English Language Learner Students

Students who have been in the United States for 12 consecutive months or less at the time of test administration may be exempted by the local school district from taking EOC Communication Arts assessments.

5.3 Students for Whom a School or District is Accountable

For accountability purposes, Missouri must include the results for any student who is eligible to take the EOC Assessments and has been enrolled at least one full academic year in a school (for school accountability) or district (for district accountability) without transferring out of the building or district for a significant period of time and re-enrolling. A full academic year is defined as the last Wednesday in September through the EOC Assessment administration. A significant period of time is considered “one more than half of the eligible days between the last Wednesday in September and the test administration.” DESE obtains enrollment information from the Missouri Student Information System (MOSIS) data that are reported by school districts. This rule applies to the building and district summary levels independently. For example, a student who is coded as “In building less than a year,” but was in the district a full academic year, is excluded from the building totals but is included in the district totals.

5.4 Dissemination of Testing Materials and Information

Riverside Publishing works with Questar Assessment, a subcontractor for the EOC Assessment program, to gather all enrollment counts and distribute all paper-and-pencil testing materials. Riverside Publishing distributes all password information for the online system. Before the start of the test window, districts enter their enrollment counts and scheduled testing window into ServicePoint, an online enrollment and materials ordering system. From those enrollment counts, Questar generates each district’s order. All paper-and-pencil materials are shipped one week before the district’s designated testing window. Districts that administer the assessments online receive an e-mail message with password information one week prior to test administration. The District Test Coordinator (DTC) is responsible for inventorying all paper-and-pencil materials, as well as for distributing the online test information to the test administrators. If additional materials are needed, the Test Coordinator is responsible for placing an Additional Materials Order (AMO) through ServicePoint.

5.5 District and Test Examiner Training

DESE is responsible for training the Test Coordinators on EOC test administration. The Regional Instructional Facilitators (RIFs) are first trained by the Assistant Director of Assessment on all information covered in the *Test Coordinator’s Manual* and the *Test Examiner’s Manual*. The RIFs then conduct training sessions for the districts within their region. The RIFs also provide assistance with test administration and serve as a liaison

between DESE and the districts. Both DESE and Riverside Publishing are available to answer any questions the districts may have about the EOC Assessment administration.

Riverside Publishing provides training to districts that administer the EOC Assessments online. The hour-long training session is conducted via WebEx and gives an overview of both the administrative and student sides of the online system.

5.6 Test Security

The EOC Assessment test books and online assessment are secure. Test Coordinators are instructed to keep the materials in a locked room or cabinet at all times when not in use. No testing materials may be photocopied, duplicated, scanned, or made accessible to personnel who are not responsible for testing. Additionally, written or oral discussion of specific EOC Assessment items breaches the security and integrity of the test. In accordance with the *Standards*, the *Test Coordinator's Manual* and *Test Examiner's Manual* contain explicit instructions about test security for Test Coordinators and Test Examiners.⁶ When the tests are delivered online, Test Examiners do not have access to the student screens for the online assessment, only to the test administrator features. In addition, a secure browser must be installed on each student computer prior to administration of the online assessments. Test items, as well as student responses, are encrypted during transmission to and from student computers.

5.7 Test Administration

5.7.1 Test Organization

Students take the EOC Assessments in one session. The EOC Assessments for the Summer 2010, Fall 2010, and Spring 2011 were revised to contain only selected response (SR) items. Each item consists of a stem followed by four response options. The tests are not timed. All items are contained in a single test book and answers are marked on a separate answer sheet.

The online assessment is also completed in one session. As in the case of the paper-and-pencil administration, the tests are not timed. Students are required to complete the practice tests on the DESE website prior to testing. These practice tests include instructions on how to use the tools in the system and practice questions for the students. For items on the online assessment, the student clicks an answer choice.

5.7.2 Test and Ancillary Materials

District Test Coordinators or School Test Coordinators are responsible for distributing all EOC Assessment materials to Test Examiners. The materials provided by Riverside Publishing and/or DESE include the following:

- *Test Examiner's Manual* (online and paper-and-pencil)
- *Test Coordinator's Manual*
- Building Identification Sheets
- Group Identification Sheets

⁶ **Standard 5.7:** Test users have the responsibility of protecting the security of test materials at all times (p. 64).

- Student barcode labels
- Session I test books
- Session I answer sheets
- Session II test books
- Math reference sheets (if applicable)
- Return kit materials

Students need the following additional materials for the paper-and-pencil assessment; these materials are not provided by Riverside Publishing or DESE:

- No. 2 pencils
- Scratch paper
- Metric ruler (for Biology)
- Dictionary, thesaurus, grammar book (for English II)

For the online assessment, each student needs a computer with a monitor, a mouse, and a keyboard. Adequate space should be left between workstations. Students can use scratch, grid, or draft paper and a writing utensil while taking the online assessment. The Test Examiner needs the following:

- A computer for logging on to the proctor interface
- A writing board and utensil

Additionally, students taking either the paper-and-pencil or online version may use a four-function calculator for the Algebra I assessment. (This is not required.)

5.7.3 Preparing the Classroom and the Students

The *Test Examiner's Manual* contains specific instructions for teachers and other test administrators regarding how the classroom should be prepared for testing. These include the following:

- Planning for the distribution and collection of materials
- Planning the seating arrangement to prevent students from seeing other students' responses
- Eliminating distractions such as bells and telephones
- Using a "Do Not Disturb" sign on the door
- Removing from students' view any classroom maps, charts, or other materials that relate to the test content
- Making arrangements for students who may not finish testing in the allotted time

Before students begin the assessment using the online system, a representative of the district or school must do the following:

- Read the entire *Online Test Examiner's Manual*
- Run a system check on each workstation used for testing
- Ensure that the MO EOC browser is downloaded to each workstation for test delivery

- Read the frequently asked questions from the link on the Test Examiner’s login page
- Input identification information for students who were not included in the MOSIS precode file
- Contact Riverside Publishing if any changes need to be made to the student roster
- Create a test session immediately before testing

Additionally, while students await proctor approval, the Test Examiner must set and verify class information and set students’ testing status codes and/or accommodations information in the online system.

The *Test Examiner’s Manual* and *Online Test Examiner’s Manual* explain some ways teachers may prepare their students for testing, including the following:

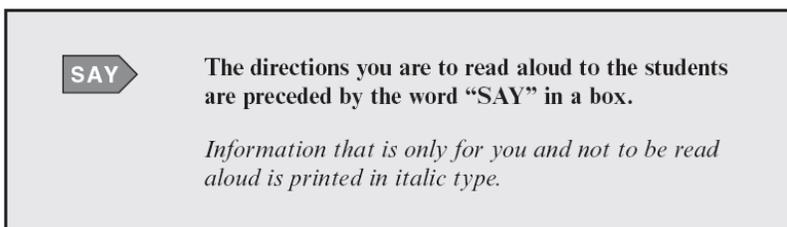
- Helping students approach the testing with a relaxed, positive attitude
- Encouraging and motivating students to do their best work
- Explaining test strategies, such as skipping harder items and coming back to them later
- Reassuring students that they will be given ample time to do their best work

Students are not allowed to use electronic devices, such as cellular phones, digital cameras, gaming devices, or scanners during the testing session. However, students may use four-function calculators during the Algebra I test session.

5.7.4 Directions for Administration

In accordance with Standard 5.1,⁷ specific standardized directions for administration are printed in the *Test Examiner’s Manual*. Directions that are to be read aloud to the students are printed in **bold** type and have a callout arrow in the margin for clarity. Information for the teacher that should not be read aloud is in *italic* type. Figure 5.1 provides an example of the type styles used in the *Test Examiner’s Manual* to differentiate between spoken and unspoken instructions. Figure 5.2 provides an example of a script from the English II EOC Assessment. Figure 5.3 provides an example of a script from the online English II EOC Assessment.

Figure 5.1: Examples of Type Styles Used to Differentiate between Spoken and Unspoken Instructions in the *Test Examiner’s Manual*



⁷ **Standard 5.1:** Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer, unless the situation or a test taker’s disability dictates that an exception should be made (p. 63).

Figure 5.2: Example Script from the *Test Examiner’s Manual* for the Algebra I EOC Assessment

Directions for Administering the Algebra I Assessment

Teacher Directions:

Distribute the test books, answer sheets, reference sheets, scratch paper, and graph or grid paper. Make sure each student has the correct answer sheet if student barcode labels were affixed before testing. Have the students write their name on the line provided on the front cover of the test book.

Before administering the test, take a moment to have the students look through the test book. Be sure that students understand what this picture means.



means that a student may want to use the Algebra I reference sheet.

Ensure that all students use a nonmechanical No. 2 pencil. If you have decided the students should use calculators, make sure all students have a working calculator. Remember, the use of a calculator is not required.

For students testing in Spring 2011 only, please fill in the FORM letter that corresponds to the student’s test book. The FORM letter must be included or scoring the assessment will be negatively impacted.

SAY

For the questions in this test, you will select an answer from a list of given choices. Use scratch paper and graph or grid paper to work the problems. Please record your final answers on the answer sheet. Remember to fill in the circle on the answer sheet that goes with the answer you chose. Your score on these questions will depend on how well you follow directions and show your understanding of Algebra I. The reference sheet and calculator can be used.

SAY

Open your test book to page 1 and read the directions. When you have finished reading the directions, turn to page 2.

Check to see that all students are on the correct page in their test books.

SAY

When you come to the word “STOP,” you have finished. You may go back over the test and check your answers. When you have finished checking your answers, close your test book and sit quietly until everyone has finished. Do you have any questions?

When you are sure that all students understand the directions, continue.

SAY

You may begin.

If a student does not understand a word, you may pronounce the word for the student, but do not define, explain, or paraphrase it. If a student has not finished in the allotted time and is making adequate progress, the student should be allowed to finish. Tests may not be returned to students to complete or to correct incomplete or inaccurate answers. When all students have finished the test,

SAY

Stop. You have finished the test.

Collect all testing materials and secure them after accounting for all test books and answer sheets.

Figure 5.3: Example Script from the *Online Test Examiner’s Manual* for the Online English II EOC Assessment

Directions for Administering the English II Assessment

Teacher Directions:

Distribute the test books and answer sheets. Make sure each student has the correct answer sheet if student barcode labels were affixed before testing. Have the students write their name on the line provided on the front cover of the test book. Before administering the test, take a moment to have the students look through the test book.

Students may NOT use a dictionary, grammar handbook, thesaurus, or bilingual dictionary during the English II Assessment. Ensure that all students use a nonmechanical No. 2 pencil.

For students testing in Spring 2011 only, please fill in the FORM letter that corresponds to the student’s test book. The FORM letter must be included or scoring the assessment will be negatively impacted.

SAY

For the questions in this test, you will select an answer from a list of given choices. Please record your final answers on the answer sheet. Remember to fill in the circle on the answer sheet that goes with the answer you chose. Your score on these questions will depend on how well you follow directions and show your understanding of what you read. You may choose to look over the questions before reading the passage. Read the passage and then answer the questions.

SAY

Open your test book to page 1 and read the directions. When you have finished reading the directions, turn to page 2.

Check to see that all students are on the correct page in their test books.

SAY

When you come to the word “STOP,” you have finished. You may go back over the test and check your answers. When you have finished checking your answers, close your test book and sit quietly until everyone has finished. Do you have any questions?

When you are sure that all students understand the directions, continue.

SAY

You may begin.

If a student does not understand a word, you may pronounce the word for the student, but do not define, explain, or paraphrase it. You may pronounce only one word per sentence. If a student has not finished in the allotted time and is making adequate progress, the student should be allowed to finish. Tests may not be returned to students to complete or to correct incomplete or inaccurate answers. When all students have finished the test,

SAY

Stop. You have finished the test.

Collect all testing materials and secure them after accounting for all test books and answer sheets.

5.8 Accommodations and Modifications

A student's IEP team has the responsibility and authority to determine individual accommodations to support and ensure his or her participation in the EOC Assessments. Allowable accommodations are intended to assist the student by reducing the effects of his or her disability without reducing performance expectations. Allowable accommodations for the EOC Assessments include the following:

- A student may receive a Braille or Large Print edition.
- A teacher may present the test content to a student in a nonstandard way, such as by reading it aloud in English or in the student's native language, paraphrasing it, or using sign language. For the English II Assessment, this will result in the lowest obtainable scale score (LOSS).
- A student may be allowed additional time to complete one or more sessions of the assessment.
- A student may use an assistive communicative device.
- A student may be tested individually or in a small group.
- A student may be allowed to use a computer, another word-processing device, or a teacher scribe to record his or her responses.
- A student may use other assistive materials such a calculator (on the English II or Biology Assessment) or a bilingual dictionary.

Modifications are alterations in the test that change construct-related requirements. The resulting information may not be equal to the information that might be obtained without modifications. While modifications invalidate the use of student scores for No Child Left Behind (NCLB) accountability determinations, the following modifications for the EOC Assessments can be provided:

- Oral reading of the and English II Assessment, including paraphrasing questions
- Oral reading in a student's native language
- Use of a bilingual dictionary for any part of the English I or English II, Session I, Assessment

As noted above, the modifications listed will result in the lowest obtainable scale score (LOSS) on the EOC Assessments. For more information on accommodations and modifications and their effects on the interpretation of the EOC Assessment scores, see the Appendix to the *Test Examiner's Manual*.

In accordance with Standard 5.2,⁸ Test Examiners indicate an accommodation, when allowed by a student's IEP and used for the EOC Assessment, by filling in the bubble corresponding to the accommodation on page 1 of the Session I answer sheet.

Table 5.1 contains information about the percentage of students who received each type of allowable accommodation for each EOC Assessment. The most prevalent type of accommodation across all three EOC Assessments was testing in a small group (provided to between 1.81% and 3.88% of students across assessments).

⁸ **Standard 5.2:** Modifications or disruptions of standardized test administration procedures or scoring should be documented (p. 63).

Table 5.1: Frequency and Percentage of Students Receiving Each Type of Allowable Accommodation on the 2010-2011 EOC Assessments

Accommodation	English II		Algebra I		Biology	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Braille	5	0.01	12	0.02	7	0.01
Large Print	22	0.03	18	0.03	18	0.03
Oral Reading	62	0.09	1702	2.41	2180	3.22
Oral Reading— Blind/Partial Sight	4	0.01	31	0.04	27	0.04
Signing of Assessment	1	0.00	12	0.02	14	0.02
Paraphrasing	6	0.01	4	0.01	3	0.00
Other Administrations	0	0.00	4	0.01	2	0.00
Oral Reading in Native Language	5	0.01	27	0.04	32	0.05
Extended Time Administered Using More Than Allotted Periods	1603	2.41	1587	2.25	1534	2.27
Other Timing	503	0.76	525	0.74	489	0.72
Other Timing	28	0.04	32	0.05	35	0.05
Use of Scribe	71	0.11	62	0.09	73	0.11
Use of Calculator, Math Tables, etc.	101	0.15	1160	1.64	670	0.99
Using Bilingual Dictionary	0	0.00	13	0.02	18	0.03
Other Response	9	0.01	5	0.01	10	0.01
Testing Individually	270	0.41	293	0.42	304	0.45
Testing in Small Group	3279	4.92	3354	4.75	3444	5.09
Other Setting	147	0.22	93	0.13	148	0.22

5.9 Materials Handling and Return

The *Test Coordinator's Manual* and *Test Examiner's Manual* contain detailed instructions for how schools and districts should collect and package the paper-and-pencil testing materials at the end of the test administration. For Test Examiners, these activities include, but are not limited to, the following:

- Collecting test books and answer sheets from the students
- Counting the test books and answer sheets and comparing the number to the totals from pretesting
- Returning all used and unused test books and answer sheets to the School Test Coordinator
- Collecting all scratch paper used during testing
- Properly handling all contaminated test books (i.e., books having contact with bodily fluids such as blood or with any potentially hazardous material)
- Verifying that the barcode labels are affixed properly to the answer sheets and Session II test books

- Verifying that the information contained on the Student Information Sheet (SIS) is accurate and complete

For School Test Coordinators, these activities include, but are not limited to, the following:

- Collecting testing materials from the Test Examiners
- Counting all test books and verifying against the pretesting total
- Completing Group Identification Sheet for each class
- Verifying that the Building Identification Sheets are correct or completing new Building Identification Sheets if incorrect
- Returning all answer sheets and test books (scorable and nonscorable) to the District Test Coordinator
- Destroying all unused answer sheets and other nonsecure testing materials

After receiving the answer sheets and scorable and nonscorable test books from the School Test Coordinators, District Test Coordinators complete the following steps:

- Verify 100% return of test books
- Complete the Test Book Accountability Form and fax it to Riverside Publishing
- Verify that each group of scorable materials is accompanied by a Group Identification Sheet
- Verify that Group Identification Sheets are used consistently for Session I and Session II scorables

For the online system, the student needs to click the End button once he or she has finished testing to submit the test for scoring. No additional information is needed from the Test Examiner after the student has completed the test. All demographic information is edited or added by the test administrator before the student starts the assessment.

5.10 Summary

The distribution, administration, and collection of the EOC Assessments is carefully communicated and executed in the detailed *Test Examiner's Manual* and *Test Coordinator's Manual*. All standards related to test security, administration, and accommodations are adhered to throughout the process. The most important steps and procedures have been covered in this chapter. Readers interested in further detail should consult the *Test Examiner's Manual* and *Test Coordinator's Manual* for the EOC Assessments.

CHAPTER 6: SCANNING, SCORING, AND QUALITY CONTROL PROCEDURES

6.1 Introduction

This chapter describes the processes used to scan, score, and provide quality control for the Missouri End-of-Course (MO EOC) Assessments. Beginning in Summer 2010, the EOC Assessments contained selected response (SR) items only and were scored by Riverside Publishing. Sections 6.2 through 6.6 of this chapter pertain to Riverside Publishing's scanning, scoring, and quality control processes for the (SR) items.

6.2 Quality Control Overview

Riverside Publishing adheres to the guidelines listed in the SCASS/TILSA *Quality Control Checklist for Processing, Scoring, and Reporting* provided by the Council of Chief State School Officers (2003). Quality assurance in processing, scoring, and reporting is the highest consideration in all stages of score report delivery. Additionally, Standard 5.8⁹ of the *Standards* (AERA, APA, and NCME 1999) specifically addresses the issue of quality control in the scoring process. To comply with this standard, Riverside Publishing employed a set of checks at each stage in the process of scoring and reporting the SR items to ensure a zero error rate for the MO EOC Assessments. Riverside Publishing documented the various quality control procedures through a variety of reports and checklists during both the preproduction and post-production phases. Documentation took the form of issues logs and quality audit reports.

6.3 Preparation and Materials Check-In

6.3.1 Preparation for Processing

Before any MO EOC Assessment answer documents were processed for operational testing, Riverside Publishing programming staff conducted a complete check of scanning programs using the program specifications and a transfer file. A test set of documents was gridded to include all response ranges, ID ranges, blanks, double grids, all correct responses, all incorrect responses, and other scenarios, depending on the specified scoring rules. These mock data were then processed through the scanning program, the editing programs, and the scoring system. The resulting file was thoroughly hand checked to ensure that the machine was scanning correctly, that the pre-edit program was picking up the proper errors, that the post-edit program was accepting corrections properly, and that the scoring system was applying the answer keys correctly. If any errors were found in the programs, the programmer was notified to make the corrections, and quality control checks were run again.

Once the programs were found to be functioning correctly, a batch of live data was processed. This pilot run involved test results from one Missouri district. The resulting data file was put through the same quality control procedures described above, and documents were hand checked against the transfer file created. Riverside Publishing

⁹ **Standard 5.8:** Test scoring services should document the procedures that were followed to assure accuracy of scoring. The frequency of scoring errors should be monitored and reported to users of the service on reasonable request. Any systematic source of scoring errors should be corrected (p. 64).

quality control staff checked and verified the live data. All quality control checks were completed successfully before the rest of the live documents were released for processing. This procedure ensured that the scanning programs were accurate and reliable.

When the MO EOC Assessment documents were first checked in at the Riverside Scoring Service[®] (RSS), they were issued a barcode number and a color-coded sheet that included vital information about the school. The documents were put into barcoded containers that were scanned at each stage of processing to constantly track the location of a client's documents in the Scoring Center.

Next, RSS staff checked that document counts matched the Return Packing Form and that all submitted materials were complete and included fully completed header sheets. If not, the documents were tagged to alert the Scoring Project Manager and Riverside Publishing Customer Service that resolution was necessary.

6.3.2 Materials Check-In

When a shipment of MO EOC Assessment documents was delivered to the Riverside Publishing Scoring Center, the arrival date, time, carrier type, and number of boxes delivered was immediately recorded in the RSS database, thus starting the clock for processing and delivering score reports. As an additional quality step, one of Riverside Publishing's trained receiving clerks hand counted the boxes and entered the number into the RSS tracking system. Any discrepancies were entered into the alert system for resolution.

Box contents (answer sheets) were verified against the Order for Scoring Services forms, and any discrepancies were entered into the RSS alert resolution system. Each order was issued a unique barcoded number that enabled the order to be tracked as it was processed through the RSS.

6.4 Materials Scanning

All documents were scanned using Scan Optics 9000M scanners, which use four mounted cameras (two on top and two on the bottom) to capture both the grayscale and the bitonal images of each page. As each document was scanned, a Print After Scan (PAS) number was printed on the edge of the document. The first six digits in the PAS were identical to the numbers identifying the container in which the documents moved through the Scoring Center. The last digits represented the order of the document in the stack. The PAS number was used by RSS staff to identify the location of an answer document in the processing system. The scanner read the skunk codes at the top of the page to determine which document code should be used for editing and scoring. Image scanners captured the entire test page as if it were a photocopy.

6.4.1 Handling of Unscannable Documents

The scanner is programmed to detect anchor points and zones to capture the image. Occasionally, a page cannot be scanned and is automatically sent by the scanner to the rejection bin. When this occurs, the scanner stops. The scanning operator follows procedures to either scan the document correctly or insert an Unscannable Document Header along with the document or page that is unscannable. Some reasons that a

document might be unscannable include manipulation during the test administration or pages missing or removed from the answer document before it was submitted for scoring. Photocopied documents are also unscannable.

6.4.2 Resolution of n-Count Discrepancies

Throughout the scanning of the MO EOC Assessment documents, the scanning station was monitored to ensure that images were gathered for all answer documents submitted with each school's or district's materials. A Scan Integrity Report compared the scanned *n*-count with the expected *n*-count on each Group/Class Header Sheet. Any discrepancies were logged into the system and resolved through a physical check of the documents before the container passed to the next station. If a resolution could not be reached, the order was entered into the alert system.

6.4.3 Application of Editing Rules

Riverside Publishing has numerous quality control procedures in place to ensure the accuracy of the scanning of the MO EOC Assessment answer documents. The scoring process applied editing rules to each document as it completed the scanning stage. The editing rules identified conflicts caused either by the student or by the scanner. Examples of these conflicts are double marks, excessive omits, or light marks. Based on these rules, documents were placed in the editing queue for an editor to resolve the conflicts. To ensure that the scanners and the editing rules were working properly, a small percentage of documents from each batch were randomly selected to go to editing, even without any mistakes or errors. If an issue could not be resolved in the editing process, an alert was sent, and a Riverside Publishing alerts specialist contacted the MO EOC Assessments program manager, who worked with the particular school or district to resolve the issue as soon as possible.

Documents that could not be read by the scanner (for instance, because the images were too light, pages were bent, etc.) were manually entered. In these instances, the first editor manually key-entered the student responses. A different editor then manually keyed the student responses a second time. The second editor was not able to see the work of the first editor. Upon completion of the two separate key entries, the system notified the second editor if there were differences in the two entries. If discrepancies were identified, the document was reviewed to determine the correct response.

6.5 Quality Control in Report Production

Riverside Publishing uses OCE V7400, OCE PS372, OCE PS88, and IBM 1245 printers. A trained Riverside Publishing print operator inspects all reports for print quality according to defined tolerances and then reprints any documents that fail this inspection.

A product assembler collated and sorted the reports for each order into folders. As the reports were put into folders, the assembler conducted a final quality control check, specifically looking at print quality, data integrity, and stray or extra sheets. Using the customer packing list, the assembler tracked the foldering process. The compiled folders were organized according to customer specifications, and the collated order was returned to Quality Control for a final check.

6.6 Quality Assurance Product Review

Riverside Publishing worked with the Missouri Department of Elementary and Secondary Education (DESE) to determine which districts were to be used in the review of the first live order. The Process and Quality Engineering department in the RSS reviewed each score report deliverable. The techniques and procedures followed in the quality assurance plan are defined below.

6.6.1 Techniques

Score reports and data were reviewed for accuracy and completeness in the following ways:

- To verify the accuracy of the data, RSS staff hand scored a sampling of student responses to ensure that the scoring system was functioning according to specification.
- To validate the completeness of the data, RSS staff verified that all records were accounted for in the district General Research File (GRF) and score reports (based on the reporting requirements).
- RSS staff reviewed the score reports to ensure that they met the reporting requirements defined for the MO EOC Assessment program.

6.6.2 Procedures

The quality review was documented in the form of a Quality Audit Report, which outlined the data elements of each score report deliverable that was audited. Nonconformance issues were documented in an issues log and were communicated to the project team.

CHAPTER 7: SCALING AND EQUATING

7.1 Introduction

This chapter details the scaling and equating procedures implemented by Riverside Publishing for the 2010-2011 Missouri End-of-Course (EOC) Assessments. Due to the removal of PE items beginning in Summer 2010, additional steps were taken to equate this and subsequent test forms. These steps were completed under guidance from the Missouri Technical Advisory Committee and DESE. This chapter will describe the scaling and equating procedures taken for each year of the operational test administrations. The intent of the processes implemented is to provide continuity across test administrations.

A pre-equating model (Kolen and Brennan, 2004) was used to produce equated forms for each EOC Assessment. The chapter begins with an overview of the equating design. Next, the item response theory (IRT) models used for equating are described, and the model assumptions are examined. This is followed by a description of the steps used to carry out the scaling and equating for the operational assessments.

7.1.1 Equating Design

The removal of PE items beginning in Summer 2010 did not change the assessment design from the standpoint of linking items. The use of horizontal links continued in the same format that has been in place since the beginning of the program. At the May 2008 meeting, Missouri's Technical Advisory Committee (TAC) recommended that post-equating be conducted as a check on the pre-equating results. To accomplish this, a common-item test design was developed in which each form was equated to a base form through a set of linking, or anchor, items. Figure 7.1 shows the post-equating design for the EOC Assessments.

Each Phase I assessment contains a set of operational items as well as 12 additional item slots. New test forms for the Summer 2010 operational administration were built from the operational item bank. Items in the additional item slots on the Summer 2010 forms were repeated from the Spring 2010 operational administration. The Fall 2010 operational tests were administered previously during the Spring 2009 administration and the items in the additional item slots were used on the Fall 2008 test administration in the additional item slots. New test forms for the Spring 2011 operational administration were built from the operational item bank. Items in the additional item slots on the Spring 2011 forms were repeated across several operational test administrations with the most recent administration being Fall 2010. The items in the additional item slots were then used for a post-equating check of the test forms.

Table 7.1: Linking Design for the MO EOC Assessments

Administration	Assessment Form	Linking Set(s)
Fall 2008	A	M1
Spring 2009	B	M1, M2, Y1, Y2
Summer 2009	C	M2
Fall 2009	D	Y1
Spring 2010	A	Y1, Y2
Summer 2010	E	Y2
Fall 2010	B	M1, M2
Spring 2011	F	M1, M2

7.2 Item Response Theory

WINSTEPS software (Linacre, 2006b) was used to accomplish the scaling and equating for the Missouri EOC Assessments during the administrations with and without PE items. WINSTEPS is designed to produce a single scale by jointly analyzing data from students' responses to both selected response (SR) and open-ended items. SR items were calibrated using the Rasch model (Rasch, 1960; Wright and Stone, 1979), while the partial credit model (Masters 1982) was used to calibrate the Performance Event/Writing Prompt (PE/WP) items in prior test administrations.

Rasch scaling is “a method for obtaining objective, fundamental, linear measures from stochastic observations of ordered category responses” (Linacre 2006a, p. 10). One feature of the Rasch model that distinguishes it from classical test theory is the placement of estimates of a person's ability and item difficulty on the same scale. The Rasch model expresses the probability of a correct response to an item as a function of the ability of the person and the difficulty of the item. In the Rasch model, the probability of a correct response to item i , given θ , is

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}}$$

where θ = latent trait, or ability, level and b_i = the difficulty parameter for item i .

Masters (1982) developed the partial credit model as an extension of the Rasch model to handle polytomous items, or items that allow for partially correct responses (e.g., open-ended items). As noted, all Missouri EOC item calibrations used the dichotomous Rasch model for SR items and the partial credit model for open-ended items (polytomous items).

7.3 Scaling and Equating

IRT pre-equating involves scaling item parameters and equating test forms based on field-test data before the forms are administered operationally. Note, however, that for the 2008–2009 year, the forms were pre-equated retroactively (after the Spring 2009 operational administration) to allow for a one-time recentering of the item pool using Spring 2009 operational data. The following approach was used for pre-equating the EOC Assessments prior to removal of the PE items:

- 1a. Calibrate all 2008 standalone field-test forms concurrently without constraint.

- 2a. Establish the base scale through calibration of the Spring 2009 operational forms without constraint.
- 3a. Examine the stability of the common items from the two calibrations (i.e., the operational form items).
- 4a. Recenter the 2008 item bank to the 2009 base scale.
- 5a. Place the 2009 embedded field test items onto the 2009 operational scale.
- 6a. Perform fixed calibrations on the Summer 2009, Fall 2009, and Spring 2010 operational forms.
- 7a. Place the 2010 embedded field test items onto the 2009 operational scale.

For detailed explanations on steps 1a through 7a of this approach, please refer to the *MO EOC Technical Report 2009-2010*. The following steps were completed in response to removal of the PE items.

1. Re-establish the base scale (with PE items removed) through calibrations of the Spring 2009 operational forms without constraint.
2. Determine the mean shift from the original calibrations (PE items included) and those obtained from step 1.
3. Determine new theta cuts and the scaling constants.
4. Perform fixed calibrations on the Summer 2010, Fall 2010, and Spring 2011 operational forms.

7.3.1 Step 1: Reestablish the Base Scale

In response to the removal of the writing prompt and performance events from the operational assessments, the base scale was reestablished on the Spring 2009 operational test form. This was accomplished by removing the PE/WP items from the calibrations. After removing the items, calibrations of the Spring 2009 operational forms were executed freely, without constraint.

7.3.2 Step 2: Use the Mean Difficulty Shift to Recenter All Items in the Item Bank

The recalibration of the items on the Spring 2009 operational forms yielded different difficulty values. The item bank was recentered to the new scale by determining the mean shift of the difficulty values from the form administered in Spring 2009 to the Spring 2009 with the PE/WP items removed. Table 7.2 lists the mean shift values for each of the test forms.

Table 7.2: Mean Shift Values for the Spring 2009 Operational Test

Subject	Mean Rasch with PE/WPs	Mean Rasch without PE/WPs	Mean Shift
English II	0.0021	0.0000	0.0021
Algebra I	-0.0057	0.0000	-0.0057
Biology	-0.0700	0.0000	-0.0700

The mean shift value was then added to the difficulty value computed in Winsteps for each item in its respective item bank.

7.3.3 Step 3: Determine New Theta Cuts and the Scaling Constants.

For the English II and Algebra I assessments new theta cuts were computed by adding the mean shift values in Table 7.2 to the theta cuts from Spring 2009. New scaling constants were then computed based on these adjusted theta values. For the Biology assessment, the Missouri TAC recommended that alternate equating methods be researched. The equipercentile equating process proved to be the most appropriate in the case of Biology. The process involved selecting raw score cuts from the Spring 2010 raw score distributions without PE items that most closely matched the performance level percentages from the Spring 2010 raw score distributions with PE items.

7.3.4 Step 4: Perform Fixed Calibrations on the Summer 2010, Fall 2010, and Spring 2011 Operational Forms.

To place the 2010–2011 operational forms onto the 2009 scale, an anchored item calibration was performed by fixing the parameters with the estimates resulting from Step 2 above. Figures 7.2 to 7.4 show the TCCs for the three operational forms (Summer, Fall, and Spring) for each content area. The TCCs generally show the three forms to be similar (differences were within 5% of the range of test scores) across the full range of ability.

Figure 7.2: TCCs for Three Alternate Forms for English II for the 2011 Test Administration

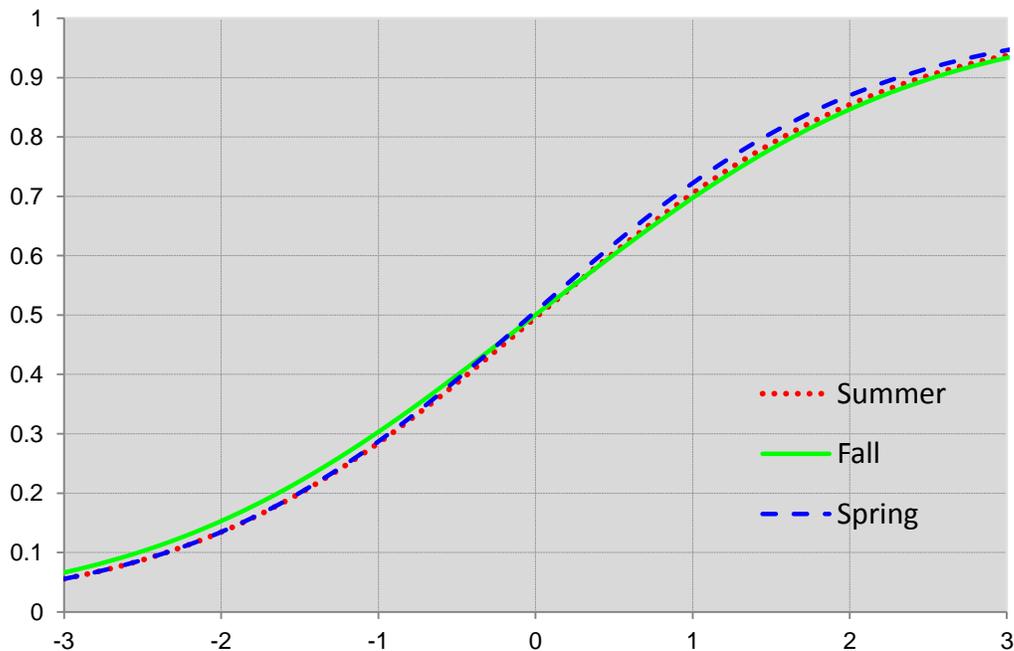


Figure 7.3: TCCs for Three Alternate Forms for Algebra I for the 2011 Test Administration

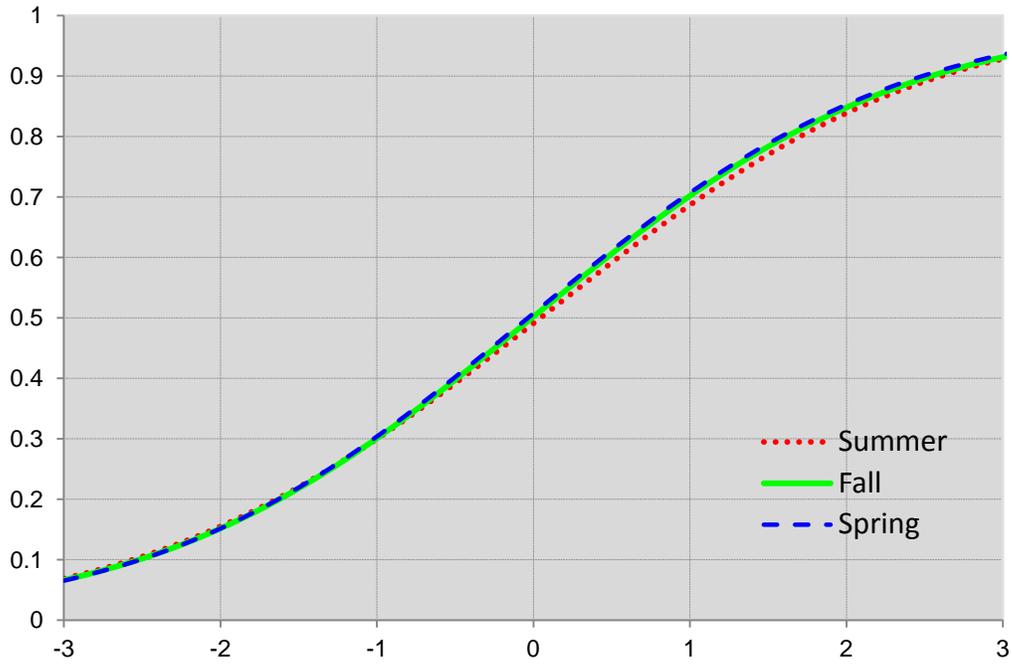


Figure 7.4: TCCs for Three Alternate Forms for Biology for the 2011 Test Administration

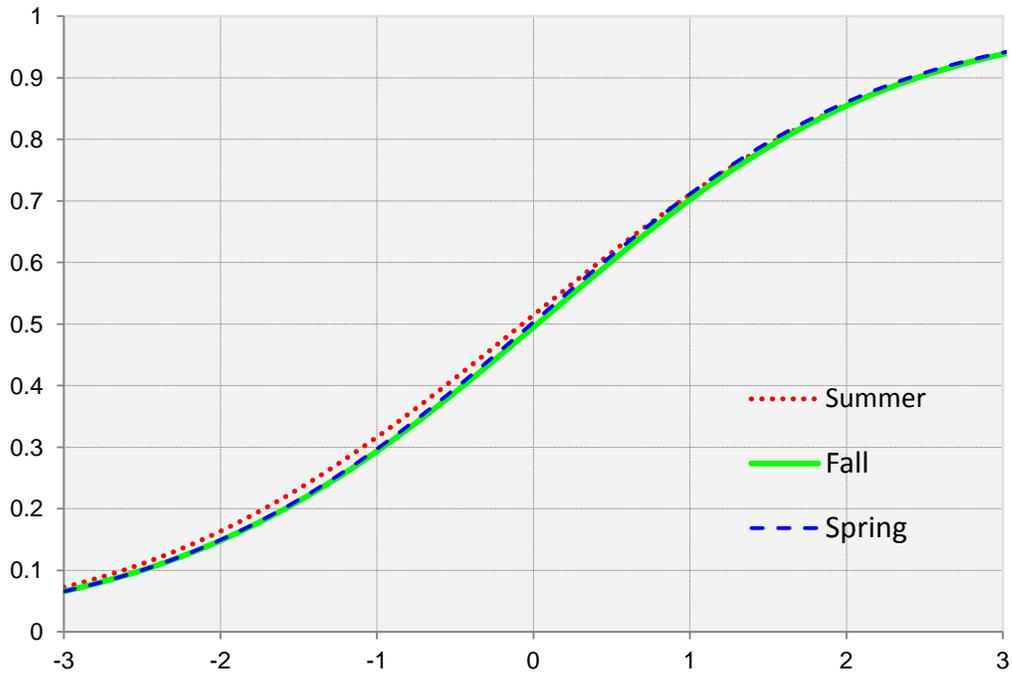


Table 7.3 provides a comparison of classical item statistics for each content area for the item pool and for the Summer 2010, Fall 2010, and Spring 2011 operational forms based on the 2009–2010 operational test administration. The comparison includes the percentage of items with p -values less than 0.3 and point-biserial correlations less than 0.1. Items with values below these criteria are typically considered low performing and are excluded from operational forms. However, such items may be included if the item pool is limited or if content considerations justify keeping an item. For example, an item may have poor field-test statistics because of examinee motivational issues or because content is not currently being taught. Examination of the summary statistics in Table 7.3 generally supports test development efforts in selecting the highest-quality items for inclusion in each operational form. Summary statistics for the Spring 2011 operational administration are provided in Table 7.4.

Table 7.3: Comparison of 2010 Item Pool with 2010–2011 Operational Test Forms

Subject	Item Set	% p -Value < .3	% Point-Biserial < .1
English II	Phase I Pool	1.32	1.17
	Summer	0.00	0.00
	Fall	2.13	0.00
	Spring	0.00	0.00
Algebra I	Phase I Pool	14.50	6.62
	Summer	8.51	0.00
	Fall	14.89	2.13
	Spring	8.57	0.00
Biology	Phase I	2.99	1.50
	Summer	0.00	2.13
	Fall	2.13	2.13
	Spring	2.13	0.00

Table 7.4: Summary Statistics for the Spring 2011 Operational Administration

Content	Total Items	Total Points	Minimum	Maximum	Mean (Raw Score)	SD (Raw Score)
English II	35	35	1	35	24.63	6.421
Algebra I	35	35	1	35	21.44	6.614
Biology	35	35	1	35	23.07	6.211

Because the Rasch model is the basis of all scoring and scaling analyses associated with the EOC Assessments, the utility of the results from the Spring 2011 administration depends on the degree to which the assumptions of the model are met as well as the degree to which the test data fit the model. As noted in Section 7.3.2, the assumptions of the Rasch model are that (1) the data are unidimensional, and (2) the data have the quality of local independence, meaning that responses to one item do not depend on responses to

another item. The analyses below address these assumptions and include evaluations of the dimensionality and local independence of the data, as well as fit indices.

7.3.5 Assessing Unidimensionality of the Data

WINSTEPS provides a residual-based, unrotated principal components analysis (PCA) that can be used to assess the unidimensionality assumption of the Rasch model. The purpose of the analysis is to reveal contrasts between opposing factors by showing the variance explained by factors not accounted for by the Rasch model. That is, the Rasch dimension is removed first, and the residual variance is then analyzed. Consequently, with this analysis, one does not want to identify a second dimension that accounts for a practically significant amount of residual variance.

Ideally, additional factors will be at the “noise” level, implying that there are no other shared dimensions in the data. Because the WINSTEPS standardized residuals are modeled to have unit normal distributions, which are independent, a PCA of these residuals should look similar to a PCA of random normal deviates. Simulation studies (such as Smith and Miao, 1994) indicate that the largest component in a set of random normal deviates would have an eigenvalue of about 1.4, which represents a small percentage of variance explained (i.e., less than 5%).

Table 7.5 shows the results of the PCA for the Spring 2011 operational form for each content area. For each analysis, the secondary dimension has an eigenvalue representing fewer than three items (less than 5% of the total variance) and, therefore, is of little practical import.

Table 7.5: Results of the PCA for the Spring 2011 Operational Tests

Content	Total Units (Items)	Second Dimension Eigenvalue	Second Dimension % of Total Variance Explained	% of Unexplained Variance	Second Dimension % of Unexplained Variance
English II	35	1.4	4.1	63.7	2.6
Algebra I	35	1.5	4.3	49.2	2.1
Biology	35	1.5	4.4	57.9	2.5

7.3.6 Assessing Local Independence of the Data

Based on the PCA, WINSTEPS also provides standardized residual correlations that can be used to assess the local independence assumption of the Rasch model. The purpose of the analysis is to detect dependency between pairs of items. Figures 7.4 to 7.6 provide screen shots from WINSTEPS Table 23.99 (Linacre, 2006b) for each content area from the Spring 2011 operational test administration. Results of these analyses generally support the assumption of local independence. More specifically, values for standardized residual correlations were generally low (i.e., had absolute values below .10), indicating little dependency between pairs of items.

Figure 7.4: Standardized Residual Correlations from the Spring 2011 Administration for English II

TABLE 23.99 Spring 2011 Missouri EOC Local Calibra MOELA11.OUT Oct 19 13:57 2011
 INPUT: 63682 PERSONS 47 ITEMS MEASURED: 63682 PERSONS 35 ITEMS 70 CATS 3.64.2

```

-----
LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT ITEMS
-----+-----+
|RESIDUL| ENTRY      | ENTRY      |
|CORRELN|NUMBER ITEM |NUMBER ITEM |
|-----+-----+-----|
| .10 | 37 IO037 | 41 IO041 |
|-----+-----+-----|
| -.10 | 10 IO010 | 36 IO036 |
| -.09 | 26 IO026 | 37 IO037 |
| -.08 | 12 IO012 | 45 IO045 |
| -.08 | 27 IO027 | 45 IO045 |
| -.08 | 7 IO007 | 26 IO026 |
| -.08 | 37 IO037 | 45 IO045 |
| -.07 | 26 IO026 | 31 IO031 |
| -.07 | 12 IO012 | 36 IO036 |
| -.07 | 10 IO010 | 26 IO026 |
|-----+-----+-----|
  
```

Figure 7.5: Standardized Residual Correlations from the Spring 2011 Administration for Algebra I

TABLE 23.99 Spring 2011 Missouri EOC Local Calibra MOMAT11.OUT Oct 19 13:58 2011
 INPUT: 63997 PERSONS 47 ITEMS MEASURED: 63997 PERSONS 35 ITEMS 70 CATS 3.64.2

```

-----
LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT ITEMS
-----+-----+
|RESIDUL| ENTRY      | ENTRY      |
|CORRELN|NUMBER ITEM |NUMBER ITEM |
|-----+-----+-----|
| -.11 | 15 IO015 | 36 IO036 |
| -.10 | 11 IO011 | 15 IO015 |
| -.09 | 1 IO001 | 36 IO036 |
| -.09 | 15 IO015 | 18 IO018 |
| -.08 | 1 IO001 | 18 IO018 |
| -.08 | 5 IO005 | 15 IO015 |
| -.08 | 3 IO003 | 15 IO015 |
| -.08 | 1 IO001 | 11 IO011 |
| -.07 | 21 IO021 | 35 IO035 |
| -.07 | 1 IO001 | 35 IO035 |
|-----+-----+-----|
  
```

Figure 7.6: Standardized Residual Correlations from the Spring 2011 Administration for Biology

TABLE 23.99 Spring 2011 Missouri EOC Local Calibra MOSCI11.OUT Oct 19 13:59 2011
 INPUT: 62068 PERSONS 47 ITEMS MEASURED: 62068 PERSONS 35 ITEMS 70 CATS 3.64.2

```

-----
LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT ITEMS
-----
|RESIDUL| ENTRY      | ENTRY      |
|CORRELN|NUMBER ITEM |NUMBER ITEM |
|-----+-----+-----|
| .10 | 18 I0018 | 27 I0027 |
| .09 | 30 I0030 | 37 I0037 |
| .08 | 13 I0013 | 15 I0015 |
| .08 | 5 I0005 | 31 I0031 |
|-----+-----+-----|
| -.10 | 26 I0026 | 34 I0034 |
| -.09 | 21 I0021 | 27 I0027 |
| -.08 | 26 I0026 | 33 I0033 |
| -.08 | 27 I0027 | 33 I0033 |
| -.08 | 27 I0027 | 34 I0034 |
| -.07 | 15 I0015 | 32 I0032 |
-----
  
```

7.3.7 Assessing Data Fit to the Model

WINSTEPS provides two statistics for indicating how well the data fit the Rasch model. Infit (inlier-sensitive or information-weighted fit) is sensitive to aberrations in item response patterns at the examinee’s ability level. High infit statistics indicate unexpected responses to items that are well-targeted at the examinee’s ability. Low infit statistics, while not a threat to measurement, may indicate over-fit of the data to the model (resulting in Guttman-like patterns) that may result in artificially inflated reliability statistics. Outfit (outlier-sensitive fit) is sensitive to outliers (in other words, to aberrant responses to items with difficulty far from a person’s ability). High outfit values may indicate lucky guessing or careless mistakes. Relatively speaking, extremely high infit values are believed to be a greater threat to the measurement process than extreme outfit values.

Infit and outfit can be expressed as a mean square (MS) statistic or on a standardized metric (*z*). Both should be considered because they provide different perspectives: MS values are more oriented toward practical significance, while standardized values are more oriented toward statistical significance. Fit statistics expressed as mean squares (statistically, a chi-square statistic divided by its degrees of freedom) show the degree of practical distortion in the measurement. The expected value is 1.0, with values less than 1.0 indicating overfitting items (too predictable) and values greater than 1.0 indicating underfitting items (unpredictability, too much noise). Rules of thumb regarding “practically significant” MS fit values vary. Wright and Linacre (1994) suggest that reasonable MS fit values range from 0.8 to 1.2 for SR items. Others believe that reasonable test results can be achieved with values from 0.5 to 1.5. Riverside Publishing

has typically considered values outside the range of 0.7 to 1.3 to be outside the range of acceptable fit.

Fit statistics expressed as *z*-scores (standardized unit normal deviates) offer a means to statistically test model fit. Standardized fit statistics show the degree of statistical improbability in the data (i.e., its significance) if the data actually do fit the model. The expected value of standardized fit statistics is 0.0, with values significantly less than 0.0 indicating too much predictability and values significantly greater than 0.0 indicating lack of predictability. Also, *z*-scores may be affected by sample sizes. For example, in a large sample, the test of interest might show a statistically significant difference. In practice, the difference might not be important.

Tables 7.6 to 7.8 provide summary statistics, including summary fit statistics, for the Spring 2011 operational test calibrations. The evaluation of fit values, specifically MS infit, yielded these results: Infit values for English II ranged from 0.84 to 1.21, values for Algebra I ranged from 0.83 to 1.32, and values for Biology ranged from 0.83 to 1.18. The fit values and output files are based on the local runs using WINSTEPS version 3.64.2. Tables 7.9 to 7.11 provide Rasch difficulties and item fit statistics for all items from the Spring 2011 operational test calibrations.

Table 7.6: Summary Statistics for the Spring 2011 Operational Test Calibrations for English II

Statistic	Rasch Difficulty Estimate	Item Mean	Infit		Outfit		Point-Biserial
			<i>MS</i>	Standardized	<i>MS</i>	Statistic	
# of Items	35	35	35	35	35	35	35
Mean	0.00	0.70	1.00	-0.76	0.99	-0.60	0.36
<i>SD</i>	0.69	0.11	0.09	8.14	0.16	8.49	0.08
Minimum	-1.68	0.42	0.84	-9.90	0.62	-9.90	0.19
Percentiles							
10	-0.91	0.59	0.90	-9.90	0.79	-9.90	0.25
25	-0.35	0.65	0.95	-9.90	0.87	-9.90	0.32
50	0.16	0.69	0.99	-2.41	0.98	-3.08	0.36
75	0.37	0.77	1.04	9.90	1.07	9.06	0.41
90	0.66	0.84	1.10	9.90	1.20	9.90	0.47
Maximum	1.51	0.91	1.21	9.90	1.33	9.90	0.50

Table 7.7: Summary Statistics for the Spring 2011 Operational Test Calibrations for Algebra I

Statistic	Rasch Difficulty Estimate	Item Mean	Infit		Outfit		Point-Biserial
			<i>MS</i>	Standardized	<i>MS</i>	Standardized	
# of Items	35	35	35	35	35	35	35
Mean	0.00	0.61	0.99	-1.36	1.02	-0.76	0.36
<i>SD</i>	1.05	0.18	0.10	8.15	0.18	8.28	0.10
Minimum	-1.84	0.16	0.83	-9.90	0.70	-9.90	0.10
Percentiles							
10	-1.35	0.33	0.89	-9.90	0.85	-9.90	0.23
25	-0.56	0.52	0.94	-9.90	0.92	-9.70	0.31
50	-0.18	0.66	1.00	-0.77	0.98	-2.61	0.37
75	0.50	0.72	1.05	5.95	1.14	9.90	0.42
90	1.51	0.83	1.09	9.90	1.23	9.90	0.47
Maximum	2.69	0.89	1.32	9.90	1.45	9.90	0.52

Table 7.8: Summary Statistics for the Spring 2011 Operational Test Calibrations for Biology

Statistic	Rasch Difficulty Estimate	Item Mean	Infit		Outfit		Point-Biserial
			<i>MS</i>	Standardized	<i>MS</i>	Statistic	
# of Items	35	35	35	35	35	35	35
Mean	0.00	0.66	0.99	-0.90	0.98	-0.37	0.34
<i>SD</i>	0.87	0.15	0.09	8.61	0.17	8.73	0.08
Minimum	-1.75	0.24	0.83	-9.90	0.59	-9.90	0.18
Percentiles							
10	-1.07	0.50	0.89	-9.90	0.74	-9.90	0.23
25	-0.45	0.55	0.94	-9.90	0.90	-9.90	0.28
50	-0.05	0.68	0.99	-2.86	0.99	-1.12	0.33
75	0.61	0.75	1.05	9.90	1.09	9.90	0.40
90	0.90	0.84	1.11	9.90	1.20	9.90	0.43
Maximum	2.29	0.90	1.18	9.90	1.28	9.90	0.49

Table 7.9: Item Statistics for the Spring 2011 Operational Test Calibrations for English II

Item Number	Item Type	Rasch Difficulty Estimate	N	MS Infit	Standardized Infit	MS Outfit	Standardized Outfit
1	SR	-0.2757	63192	0.99	-2.41	0.92	-8.85
2	SR	-1.4363	63191	0.91	-9.34	0.77	-9.90
3	SR	0.0872	63189	1.01	2.47	1.01	1.85
4	SR	-0.8950	63191	0.99	-0.73	1.08	5.90
5	SR	-0.9158	63191	0.98	-2.68	1.01	0.43
6	SR	0.3187	63190	1.05	9.90	1.05	7.54
7	SR	-0.2903	63192	0.94	-9.90	0.84	-9.90
8	SR	0.1506	63191	1.06	9.90	1.06	7.36
9	SR	0.0384	63190	1.06	9.90	1.10	9.90
10	SR	0.3169	63190	0.90	-9.90	0.85	-9.90
11	SR	-0.4998	63191	1.02	3.24	1.04	3.95
12	SR	-0.8994	63190	0.84	-9.90	0.62	-9.90
24	SR	-0.8380	63190	0.91	-9.90	0.83	-9.90
25	SR	0.0073	63190	1.00	-0.33	0.97	-3.94
26	SR	0.6598	63188	1.21	9.90	1.33	9.90
27	SR	-0.2610	63191	0.86	-9.90	0.76	-9.90
28	SR	0.7394	63190	1.01	4.15	1.02	3.44
29	SR	1.5062	63189	1.04	9.90	1.12	9.90
30	SR	0.2738	63187	1.02	5.46	1.07	9.33
31	SR	0.2812	63189	0.90	-9.90	0.84	-9.90
32	SR	0.3218	63191	0.98	-5.43	0.97	-5.20
33	SR	0.3567	63192	0.99	-3.54	0.96	-6.96
34	SR	-0.1375	63191	1.12	9.90	1.28	9.90
35	SR	0.1575	63192	0.96	-9.90	0.88	-9.90
36	SR	0.3737	63191	1.18	9.90	1.32	9.90
37	SR	0.4386	63190	0.89	-9.90	0.82	-9.90
38	SR	0.2099	63190	0.95	-9.90	0.94	-8.83
39	SR	-0.4085	63185	0.98	-2.76	1.16	9.90
40	SR	0.6529	63190	0.99	-1.58	0.98	-3.08
41	SR	0.6613	63192	0.97	-8.82	0.95	-9.29
43	SR	-1.6796	63189	0.95	-4.29	0.93	-3.49
44	SR	0.6617	63188	1.04	9.90	1.05	8.79
45	SR	0.2677	63190	1.17	9.90	1.23	9.90
46	SR	0.9828	63192	1.06	9.90	1.07	9.90
47	SR	-0.9274	63190	0.90	-9.90	0.76	-9.90

Table 7.10: Item Statistics for the Spring 2011 Operational Test Calibrations for Algebra I

Item Number	Item Type	Rasch Difficulty Estimate	N	MS Infit	Standardized Infit	MS Outfit	Standardized Outfit
1	SR	0.8482	63810	1.17	9.90	1.25	9.90
2	SR	-0.5764	63809	0.95	-9.90	0.95	-5.57
3	SR	0.4442	63809	0.88	-9.90	0.85	-9.90
4	SR	-0.6040	63810	0.95	-9.90	0.95	-5.22
5	SR	0.9048	63805	0.91	-9.90	0.89	-9.90
10	SR	0.0464	63808	1.01	2.67	1.01	2.13
11	SR	-0.3387	63808	0.84	-9.90	0.76	-9.90
12	SR	-1.4145	63805	0.94	-8.60	0.94	-4.30
13	SR	-0.3925	63808	0.96	-9.90	0.91	-9.90
14	SR	-0.3088	63806	0.94	-9.90	0.91	-9.90
15	SR	0.4318	63806	1.32	9.90	1.45	9.90
16	SR	2.6915	63808	1.10	9.90	1.32	9.90
17	SR	1.2788	63808	1.13	9.90	1.19	9.90
18	SR	-1.2249	63810	0.86	-9.90	0.70	-9.90
19	SR	-0.3178	63808	0.97	-6.53	0.93	-9.04
20	SR	-1.4037	63806	1.04	5.89	1.16	9.90
21	SR	-0.0920	63807	1.01	3.37	1.03	4.06
26	SR	0.1953	63808	0.97	-9.46	0.94	-9.49
27	SR	0.4539	63807	1.00	1.34	0.99	-1.23
28	SR	-0.2073	63809	1.00	-0.77	0.98	-3.08
29	SR	1.7669	63808	1.07	9.90	1.19	9.90
30	SR	-0.0505	63810	1.01	3.51	1.00	-0.26
31	SR	-0.1816	63808	0.90	-9.90	0.85	-9.90
32	SR	1.7650	63807	1.01	3.06	1.19	9.90
33	SR	-1.7713	63809	0.95	-6.25	0.88	-7.25
34	SR	-1.0831	63809	0.94	-9.90	0.92	-6.90
35	SR	0.5525	63808	0.93	-9.90	0.92	-9.90
36	SR	-0.5385	63806	0.83	-9.90	0.72	-9.90
37	SR	-1.2652	63806	1.00	-0.06	1.01	0.90
38	SR	0.9438	63806	0.99	-2.59	1.01	1.49
43	SR	-0.2296	63808	1.05	9.90	1.07	9.79
44	SR	-1.8384	63807	1.05	6.01	1.42	9.90
45	SR	1.6610	63808	1.06	9.90	1.19	9.90
46	SR	-0.2835	63811	1.00	0.44	0.98	-2.61
47	SR	0.1382	63811	1.08	9.90	1.12	9.90

Table 7.10: Item Statistics for the Spring 2011 Operational Test Calibrations for Biology

Item Number	Item Type	Rasch Difficulty Estimate	N	MS Infit	Standardized Infit	MS Outfit	Standardized Outfit
1	SR	-1.0330	61801	0.89	-9.90	0.75	-9.90
2	SR	0.1279	61802	1.04	9.90	1.06	8.61
3	SR	-1.0723	61801	1.01	1.53	1.07	5.37
4	SR	-0.8905	61799	0.98	-3.59	0.99	-1.25
5	SR	-0.2514	61802	1.05	9.90	1.03	4.23
10	SR	-0.2078	61802	0.94	-9.90	0.91	-9.90
11	SR	1.0704	61802	0.99	-2.86	1.01	2.87
12	SR	0.0241	61801	1.13	9.90	1.23	9.90
13	SR	-0.2104	61801	0.97	-7.47	0.89	-9.90
14	SR	1.4137	61800	1.07	9.90	1.11	9.90
15	SR	0.3670	61802	0.91	-9.90	0.88	-9.90
16	SR	-0.8691	61800	0.97	-5.53	0.97	-2.90
17	SR	0.7934	61802	0.94	-9.90	0.94	-9.90
18	SR	-1.7521	61800	0.87	-9.90	0.66	-9.90
19	SR	0.1646	61802	1.06	9.90	1.07	9.90
20	SR	0.1676	61800	1.00	-1.20	0.99	-1.12
21	SR	2.2862	61801	1.00	0.52	1.21	9.90
26	SR	-0.4766	61802	0.85	-9.90	0.72	-9.90
27	SR	-1.5149	61801	0.83	-9.90	0.59	-9.90
28	SR	-0.0650	61799	0.90	-9.90	0.84	-9.90
29	SR	0.4463	61801	0.98	-7.30	0.97	-4.60
30	SR	-1.0592	61800	0.93	-9.58	0.87	-9.90
31	SR	-0.4201	61800	1.08	9.90	1.10	9.90
32	SR	0.5233	61801	1.16	9.90	1.21	9.90
33	SR	0.7869	61802	1.13	9.90	1.18	9.90
34	SR	0.9375	61797	1.18	9.90	1.28	9.90
35	SR	0.8429	61801	1.02	5.56	1.02	4.60
36	SR	-0.0520	61800	0.94	-9.90	0.90	-9.90
37	SR	-1.3220	61802	0.89	-9.90	0.73	-9.90
38	SR	-0.2550	61801	1.05	9.90	1.14	9.90
43	SR	-0.1200	61801	0.96	-8.81	0.93	-9.90
44	SR	0.7049	61802	0.93	-9.90	0.92	-9.90
45	SR	-0.2032	61800	1.01	2.56	1.06	6.90
46	SR	0.7440	61801	1.02	4.71	1.02	3.85
47	SR	0.3738	61801	1.09	9.90	1.17	9.90

7.3.8 Establish Scaling Transformations

Total scores for the EOC Assessments were reported in scale scores with a range of 100–250. A scale score of 200 represents the cut point between Basic and Proficient, and a scale score of 225 represents the cut point between Proficient and Advanced. The scale score ranges are displayed in Table 7.12.

Table 7.12: Scale Score Ranges for EOC Assessment Achievement Levels

EOC Assessment	Achievement Level	Scale Score Range
English II	Below Basic	100-179
	Basic	180-199
	Proficient	200-224
	Advanced	225-250
Algebra I	Below Basic	100-176
	Basic	177-199
	Proficient	200-224
	Advanced	225-250
Biology	Below Basic	100-177
	Basic	178-199
	Proficient	200-224
	Advanced	225-250

To produce these scale score ranges, linear transformations were applied to theta estimates and scale scores. The following formula was used to obtain the slopes and intercepts for the transformation functions:

$$sc(y) = \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1} \right] y + \left\{ (sc(y_1) - \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1} \right] \theta_1) \right\},$$

where θ_1 and θ_2 are person parameter estimates that correspond to the cut score points, and $sc(y_1)$ and $sc(y_2)$ are scale score points. This formula was adopted from Kolen and Brennan (2004, p. 337). For the Spring 2009 base scale, $sc(y_1)$ was 200 and $sc(y_2)$ was 225. Slopes and intercepts of the transformation functions are summarized in Table 7.13. These same slopes and intercepts will be applied to all future forms for each content area.

Table 7.13: Summary of Slopes and Intercepts of Theta to Scale Score Transformation Functions by Content Area

	Basic			Proficient			Advanced			Slope	Intercept
	Raw Score	Theta	Scale Score	Raw Score	Theta	Scale Score	Raw Score	Theta	Scale Score		
English II	12	-0.7087	180	21	0.5045	200	29	2.0339	225	16.35	191.75
Algebra I	12	-0.7941	177	20	0.3649	200	28	1.6174	225	19.96	192.72
Biology	13	-0.5531	178	22	0.6289	200	30	1.9934	225	18.33	188.47

In addition to the above scaling transformation, the following rules were also applied:

- The raw score cut (e.g., for Proficient) was selected as the lowest raw score associated with a rounded scale score of 200. The same strategy was also followed for a scale score of 225.
- If there was no raw score associated with a rounded scale score of 200, the raw score with the highest scale score below 200 was selected as the cut score and assigned a scale score of 200. For example, if two consecutive raw scores were associated with rounded scale scores of 198 and 201, the scale score of 198 was moved up to 200. The same strategy was also followed for a scale score of 225.
- Scale scores below 100 were rounded up to 100.
- Scale scores above 250 were rounded down to 250.
- For each test, for a perfect raw score, the scale score was set to 250.

Tables 7.14 to 7.22 provide the raw score to scale score conversions for Summer 2010, Fall 2010, and Spring 2011, respectively.

Table 7.14: Raw Score to Scale Score Conversions for Summer 2010, English II

Raw Score	Scale Score	<i>CSEM</i>
0	111	30
1	131	17
2	143	12
3	151	10
4	156	9
5	160	8
6	164	8
7	167	7
8	170	7
9	173	7
10	176	6
11	178	6
12	180	6
13	183	6
14	185	6
15	187	6
16	189	6
17	191	6
18	193	6
19	195	6
20	197	6
21	200	6
22	202	6
23	204	6
24	206	6
25	209	6
26	212	7
27	214	7
28	218	7
29	221	8
30	225	8
31	229	9
32	235	10
33	243	12
34	250	17
35	250	30

Table 7.15: Raw Score to Scale Score Conversions for Summer 2010, Algebra I

Raw Score	Scale Score	<i>CSEM</i>
0	100	37
1	113	21
2	128	15
3	138	13
4	145	11
5	151	10
6	156	10
7	160	9
8	164	9
9	168	8
10	171	8
11	174	8
12	178	8
13	181	8
14	184	8
15	186	8
16	189	8
17	192	8
18	195	8
19	198	8
20	201	8
21	204	8
22	206	8
23	209	8
24	213	8
25	216	8
26	219	8
27	223	9
28	227	9
29	231	10
30	236	10
31	241	11
32	248	13
33	250	15
34	250	21
35	250	37

Table 7.16: Raw Score to Scale Score Conversions for Summer 2010, Biology

Raw Score	Scale Score	<i>CSEM</i>
0	100	34
1	115	19
2	129	14
3	137	11
4	143	10
5	149	9
6	153	9
7	157	8
8	161	8
9	164	8
10	167	7
11	170	7
12	173	7
13	176	7
14	178	7
15	181	7
16	183	7
17	186	7
18	188	7
19	191	7
20	194	7
21	196	7
22	199	7
23	201	7
24	204	7
25	207	7
26	210	8
27	214	8
28	217	8
29	221	9
30	225	9
31	230	10
32	237	11
33	245	14
34	250	19
35	250	34

Table 7.17: Raw Score to Scale Score Conversions for Fall 2010, English II

Raw Score	Scale Score	<i>CSEM</i>
0	109	30
1	129	17
2	141	12
3	149	10
4	154	9
5	159	8
6	163	8
7	166	7
8	169	7
9	172	7
10	175	7
11	177	6
12	180	6
13	182	6
14	185	6
15	187	6
16	189	6
17	191	6
18	193	6
19	196	6
20	198	6
21	200	6
22	203	6
23	205	6
24	207	6
25	210	7
26	213	7
27	215	7
28	219	7
29	225	8
30	226	8
31	230	9
32	236	10
33	243	12
34	250	17
35	250	30

Table 7.18: Raw Score to Scale Score Conversions for Fall 2010, Algebra I

Raw Score	Scale Score	<i>CSEM</i>
0	100	37
1	116	20
2	131	15
3	140	12
4	147	11
5	152	10
6	157	9
7	161	9
8	165	9
9	168	8
10	171	8
11	177	8
12	178	8
13	180	8
14	183	7
15	186	7
16	189	7
17	191	7
18	194	7
19	197	7
20	200	7
21	202	7
22	205	8
23	208	8
24	211	8
25	214	8
26	218	8
27	221	9
28	225	9
29	229	10
30	234	10
31	240	11
32	247	13
33	250	15
34	250	21
35	250	37

Table 7.19: Raw Score to Scale Score Conversions for Fall 2010, Biology

Raw Score	Scale Score	<i>CSEM</i>
0	100	34
1	117	19
2	131	14
3	139	12
4	146	10
5	151	9
6	155	9
7	159	8
8	163	8
9	166	8
10	169	7
11	172	7
12	175	7
13	178	7
14	180	7
15	183	7
16	185	7
17	188	7
18	190	7
19	193	7
20	195	7
21	197	7
22	200	7
23	203	7
24	205	7
25	208	7
26	211	7
27	214	8
28	218	8
29	225	9
30	226	9
31	231	10
32	237	11
33	245	14
34	250	19
35	250	34

Table 7.20: Raw Score to Scale Score Conversions for Spring 2011, English II

Raw Score	Scale Score	<i>CSEM</i>
0	111	30
1	131	17
2	143	12
3	151	10
4	156	9
5	160	8
6	164	8
7	167	7
8	170	7
9	173	7
10	175	6
11	178	6
12	180	6
13	182	6
14	184	6
15	186	6
16	188	6
17	190	6
18	192	6
19	194	6
20	196	6
21	200	6
22	201	6
23	203	6
24	205	6
25	207	6
26	210	7
27	213	7
28	216	7
29	219	8
30	225	8
31	227	9
32	232	10
33	240	12
34	250	17
35	250	30

Table 7.21: Raw Score to Scale Score Conversions for Spring 2011, Algebra I

Raw Score	Scale Score	<i>CSEM</i>
0	100	37
1	115	20
2	130	15
3	139	12
4	146	11
5	151	10
6	156	9
7	160	9
8	164	9
9	167	8
10	171	8
11	174	8
12	177	8
13	180	8
14	183	7
15	185	7
16	188	7
17	191	7
18	193	7
19	196	7
20	200	7
21	202	7
22	204	8
23	207	8
24	210	8
25	213	8
26	217	8
27	220	9
28	225	9
29	229	10
30	233	10
31	239	11
32	246	13
33	250	15
34	250	21
35	250	37

Table 7.22: Raw Score to Scale Score Conversions for Spring 2011, Biology

Raw Score	Scale Score	<i>CSEM</i>
0	100	34
1	119	19
2	133	14
3	141	11
4	147	10
5	152	9
6	157	9
7	160	8
8	164	8
9	167	7
10	170	7
11	173	7
12	176	7
13	178	7
14	181	7
15	183	7
16	185	7
17	188	7
18	190	7
19	193	7
20	195	7
21	197	7
22	200	7
23	202	7
24	205	7
25	208	7
26	211	7
27	214	8
28	217	8
29	221	9
30	225	9
31	230	10
32	236	11
33	245	14
34	250	19
35	250	34

CHAPTER 8: REPORTING

8.1 Introduction

The purpose of reporting assessment data is to communicate test results to students, their parents, and their teachers. The Missouri End-of-Course (MO EOC) Assessment reports provide useful information for determining the performance of students in a particular school and classroom. These reports help describe students' knowledge of a given set of expectations, allowing educators to determine specific instructional needs, measure student mastery toward post-secondary readiness, provide evidence of accountability for Missouri and national programs, and evaluate educational programs. Additionally, districts may use locally designed assessments aligned to the Show-Me Content Standards and Course-Level Expectations (CLEs) to provide more detailed information for each student in specific test areas.

Paper reports are generated for all assessment windows following the Spring administration; therefore, for the Summer 2010, Fall 2010, and Spring 2011 assessments, the paper reports were generated and distributed following the Spring 2011 operational administration. However, teachers may access their students' raw scores for items shortly after the district's testing materials have been received for processing in each assessment window.

For each testing event, Riverside Publishing converts each student's raw score points earned into an EOC scale score, as described in Chapter 7: Scaling and Equating. A student receives an EOC scale score when he or she has made a valid attempt in any session. EOC scale scores range in value from 100 to 250. The EOC scale score determines the student's achievement level. For all content areas, a scale score of 200 to 224 is considered Proficient, and a scale score of 225 and above is considered Advanced. The cut score for Basic varies by content area. Each achievement level represents standards of performance for each assessed content area (English II, Algebra I, and Biology). Achievement-level scores describe what students can do in terms of the content and skills assessed. These scores provide a way to compare test results with standards of academic performance. Panels drawn from Missouri's educational, business, and professional communities recommended the raw score cuts (based on the Spring 2009 test forms) to be used for each achievement level. These cuts were then reviewed and adopted by the Missouri State Board of Education. For more information on how the achievement levels were set, refer to Chapter 3: Achievement-Level Setting.

No test provides a perfect measure of a student's ability. This situation is expected because all tests have a known standard error of measurement (*SEM*). The *SEM* represents the amount of variability that can be expected in a student's test score due to the inherent imprecision of the test. For example, if the student were tested again, he or she would likely obtain a slightly different score. The range for this new score is provided as a standard error (*SE*) and gives an indication of the margin of error for the reported scale score.

8.2 Individual Student Report

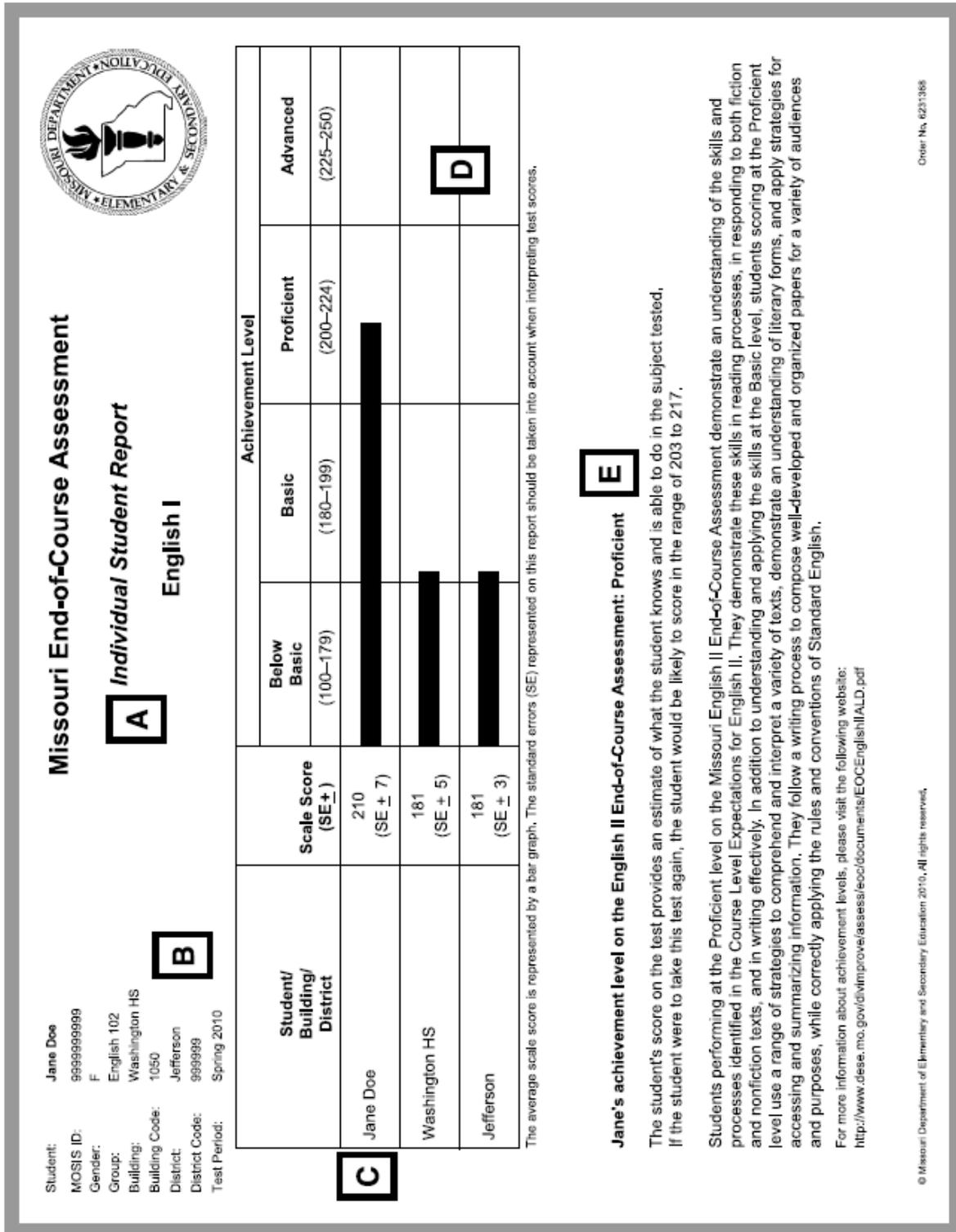
The 2010-2011 Individual Student Report provides information about performance on the EOC Assessment, describing the results in terms of four levels of achievement in a content area. It is used for measuring and reflecting an individual student's mastery toward post-secondary readiness for the content area. It is used in instructional planning as a point of reference during a parent-teacher conference and for permanent record keeping. Teachers are informed that other sources of information should be used along with this report when determining the student's areas of strength or need.

On the report, achievement-level scores describe what students can do in terms of the CLEs for the content and skills assessed by the EOC Assessment. A student at the Proficient or Advanced level has met the standard.

A sample of the 2010-2011 Individual Student Report appears in Figure 8.1. A brief description of selected parts of the report follows:

- A. The heading of the Individual Student Report includes the content area for the results being presented. A separate report is produced for each content area tested.
- B. The Student Information section contains the biographic data for the individual student taking the assessment. Identifying information, including the MOSIS ID, gender, building, and district, is listed, followed by the test period.
- C. The individual student's results are presented numerically as a three-digit scale score with the *SE*. An accompanying bar graph to the right of the scale score illustrates the achievement level obtained by the student. Achievement levels (whether Advanced, Proficient, Basic, or Below Basic) are based on the scale score ranges listed beneath the Achievement Level heading in the table.
- D. The mean scale scores for the student's building and district are displayed in the two rows below the student's individual results. The mean scale score, with an associated *SE*, and the bar graph provide a way to view the individual's results in contrast to the group's results for the content area during the same test period.
- E. The narrative describes the student performance characteristics corresponding to the obtained level of achievement. The text is specific to the content area tested. At the bottom of the narrative is a URL for a website that provides additional information for all achievement levels for the content area.

Figure 8.1: Individual Student Report

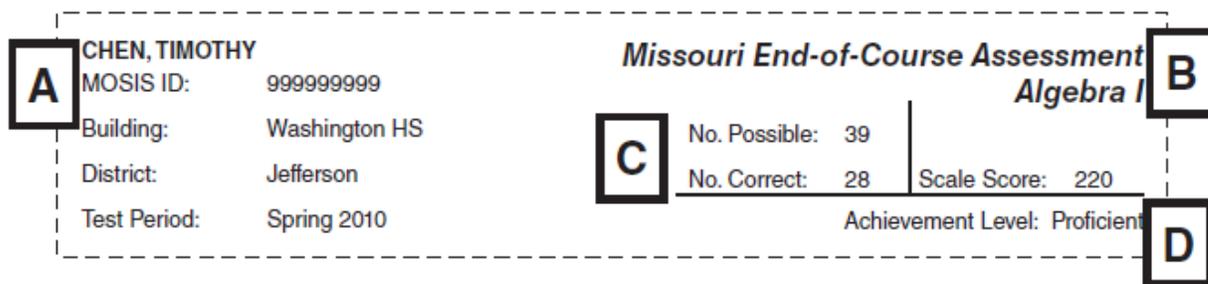


8.3 Student Score Label

The 2010-2011 Student Score Label provides a summary of a student's results on the EOC Assessment. A separate label is produced for each content area tested. The individual label provides the student's biographic data, raw score, scale score, and achievement level. The labels have adhesive backing so they can be easily transferred onto the student record folders. A sample label is shown below in Figure 8.2. A brief description of selected parts of the label follows:

- A. The student's name and identifying information are provided on the left side of the label.
- B. The upper right side of the label shows the content area tested. If a student has results for more than one content area, the next label is printed below the first one.
- C. The middle of the label has the Number Possible and the student's raw score (Number Correct). A corresponding column to the right of these data contains the raw score's associated Scale Score.
- D. The student's achievement level is displayed in the lower right corner below the scores.

Figure 8.2: Student Score Label



8.4 Online Crystal Reports

Schools and districts are able to access summary level reports through the online Crystal Reports tool. This tool allows district and school administrators to create on-the-fly reports containing information relevant to their data needs. There are several reporting options available through the Crystal Reports tool, including administrative reports, adequate yearly progress (AYP) reports, achievement level reports, content standard reports, and item analysis reports.

For each subreport, a user selects various filters such as year, grade/content area, and level of reporting (state, district, or school) to create the desired report. For the Content Standard Reports, the user may also disaggregate results by various subgroups (e.g., race, disability).

A detailed discussion of all available reports is beyond the scope of this document. Only those reports that are first-level analyses of MO EOC data will be discussed. The Achievement Level-5 reports will not be discussed, as these are summaries of the pre-No Child Left Behind (NCLB) testing program. In addition, the AYP reports and some of the

Administrative Reports, including the High School Career Education Student Summary and the Level Not Determined, will not be discussed.

The Crystal Reports tool is accessed through the Missouri Department of Elementary and Secondary Education (DESE) website. Each school and/or district is assigned a user name and password to access the site.

8.4.1 Administrative Reports

These reports provide student-level test data. Based on only the MO EOC Assessment results, four reports are generated: MO EOC Scale Score Summary, MO EOC Student Demographic, Student Achievement Level, and Student Report.

MO EOC Scale Score Summary: This report lists each student in the school or district along with his or her MOSIS ID, testing year, content area, grade level, MO EOC scale score, and achievement level.

MO EOC Student Demographic: This report lists all students in the school or district along with their date of birth (DOB), content area, MOSIS ID, district ID, and relevant demographic information, including if the student has been in the district for less than a year, if the student has been in the building for less than a year, if the student is limited English proficiency (LEP), the student's race, if the student qualifies for free and reduced lunch (SES), if the student has an individualized education program (IEP), if the student is an English-language learner (ELL)/LEP who has been in the school for less than one year and in the country for less than three years, if the student is an LEP/ELL Title 3, the number of months the LEP/ELL student has been in the U.S., the student's disability diagnosis, and if the student is Title 1.

Student Achievement Level: This report lists all students in a school or district along with the year of testing, content area, grade-level, achievement level, and MOSIS ID.

Student Report: For each school or district, this report contains the following information: student name, DOB, district student number, MOSIS ID, content area tested, grade level, achievement level, and scale score for each content area tested.

8.4.2 Achievement Level 4

These reports contain summary information on school or district performance in terms of the four MO EOC achievement levels. There are two types of achievement level reports: Achievement Level 4 Chart and Achievement Level 4 Report.

Achievement Level 4 Chart: This report charts the percentage of students classified as Proficient or Advanced on each MO EOC Assessment. State-level, district-level, and/or school-level performance may be displayed on the chart.

Achievement Level 4 Report: This report summarizes the number and percentage of students in each achievement level. This report is comprised of 10 columns: Total, content area, grade, year, number of accountable (ACC) students, number of reportable (REP) students, number and percentage of students classified in the Basic (B) achievement level, number and percentage of students classified in the Proficient (P) achievement level, number and percentage of students classified in the Advanced (A) achievement level, and mean MO EOC scale score. The first column, Total, shows if

aggregate or disaggregated information is being displayed. A key to the abbreviations is found in the bottom left corner.

8.4.3 Content Standard

The content standard reports summarize information about the content standards.

Content Standards Report: This report has 14 columns: content area, grade level, category/type, year, percentage of points earned on content standard 1 (CS-1), points possible (PP) on CS-1, percentage of points earned on CS-2, PP on CS-2, percentage of points earned on CS-3, PP on CS-3, percentage of points earned on CS-4, PP on CS-4, percentage of points earned on CS-5, and PP on CS-5. The category/type column indicates if the data are aggregated or disaggregated.

Content Standards Detail: This report shows the percentage of points each student achieved on each content standard within a particular content area.

8.4.4 Item Analysis Expanded

This set of reports provides detailed item-level results for the school or district, aggregated either by the content standard or the process standard.

Content Standard IBD EX: The Content Standard Item Benchmark Descriptor (IBD) Extended (EX) report contains item-level detail aggregated by content standard. The report is comprised of 11 columns: school code (SC), grade level (GR), standard number and description (desc.), code for the course-level expectation (CLE), description of the CLE, depth of knowledge (DOK) of the item, session/item number where the item was in the operational test, question type (QT), points possible for the item, the average points (avg pts) earned by students in the district on that item, and percentage of points earned by students in the district on that item.

CHAPTER 9: SUMMARY STATISTICS

9.1 Introduction

This chapter provides descriptive statistics for the number correct raw score and for scale scores for each of the three Missouri End-of-Course (EOC) Assessments from the Summer 2010, Fall 2010, and Spring 2011 administrations. Statistics include N counts, means, standard deviations (SD), minimum and maximum values, and a variety of data disaggregations.

9.2 Descriptive Statistics for Total Raw Score

Descriptive statistics for total raw score are summarized in Table 9.1 by test administration and content area.

Table 9.1: Descriptive Statistics for Total Raw Score

Test Period	Subject	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	476	7	34	18.60	6.373
	Algebra I	1,126	3	35	17.22	5.977
	Biology	384	4	34	18.80	6.119
Fall 2010	English II	1,833	4	35	22.18	6.796
	Algebra I	2,741	1	35	20.14	8.031
	Biology	2,391	2	35	21.78	7.175
Spring 2011	English II	63,682	1	35	24.63	6.421
	Algebra I	63,997	1	35	21.44	6.614
	Biology	62,068	1	35	23.07	6.211

9.3 Descriptive Statistics for Total Raw Score by Cluster

Table 9.2 summarizes the number correct raw score by test administration, content area, and cluster.

Table 9.2: Descriptive Statistics for Total Raw Score by Test Administration, Content Area, and Cluster

Test Period	Subject	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>	
Summer 2010	English II	Reading	476	4	29	15.37	5.772
		Writing	476	0	5	3.23	1.263
	Algebra I	Number and Operations	1126	0	9	4.94	2.136
		Algebraic Relationships	1126	1	17	7.10	2.745
		Data and Probability	1126	0	9	5.19	2.179
	Biology	Characteristics and Interactions of Living Organisms	384	3	21	10.35	3.918
		Changes in Ecosystems and Interactions of Organisms with their Environments	384	1	13	8.45	2.868
	Fall 2010	English II	Reading	1833	2	30	18.67
Writing			1833	0	5	3.50	1.281
Algebra I		Number and Operations	2741	0	8	4.89	2.260
		Algebraic Relationships	2741	0	19	10.32	4.556
		Data and Probability	2741	0	8	4.94	1.957
Biology		Characteristics and Interactions of Living Organisms	2391	2	22	13.15	4.805
		Changes in Ecosystems and Interactions of Organisms with their Environments	2391	0	13	8.63	2.849

Table 9.2: Descriptive Statistics for Total Raw Score by Test Administration, Content Area, and Cluster (continued)

Test Period	Subject	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>	
Spring 2011	English II	Reading	63682	1	30	21.08	5.711
		Writing	63682	0	5	3.56	1.187
	Algebra I	Number and Operations	63997	0	8	5.22	2.072
		Algebraic Relationships	63997	0	19	10.76	4.036
		Data and Probability	63997	0	8	5.46	1.414
	Biology	Characteristics and Interactions of Living Organisms	62068	0	22	13.68	4.116
		Changes in Ecosystems and Interactions of Organisms with their Environments	62068	0	13	9.39	2.638

9.4 Descriptive Statistics for Scale Scores by Test Period and Subject

Descriptive statistics of scale scores and percentage distributions of students' achievement levels are summarized in Tables 9.3 and 9.4. Table 9.3 summarizes student scale scores by each End-of-Course Assessment for the Summer 2010, Fall 2010, and Spring 2011 administrations. Table 9.4 lists the percentage and frequency of students in each achievement level.

Table 9.3: Scale Score Distributions for Each End-of-Course Assessment

Descriptive Statistics						
Test Period	Subject	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
Summer 2010	English II	476	167	250	195.17	15.411
	Algebra I	1,126	138	250	193.30	18.749
	Biology	384	143	250	191.16	17.635
Fall 2010	English II	1,833	154	250	204.99	18.535
	Algebra I	2,741	116	250	202.13	25.855
	Biology	2,391	131	250	201.95	21.843
Spring 2011	English II	63,682	131	250	209.86	17.927
	Algebra I	63,997	115	250	204.64	21.087
	Biology	62,068	119	250	204.79	18.544

Scale scores range from a minimum of 100 to a maximum of 250 for the three content areas administered in Summer 2010, Fall 2010, and Spring 2011. For English II, a minimum scale score of 180 is required to earn an achievement level of Basic. For Algebra I, a minimum scale score of 177 is required to earn an achievement level of Basic. For Biology, a minimum scale score of 178 is required to earn an achievement level of Basic. For all content areas, a scale score of 200 represents the minimum score to earn an achievement level of Proficient, and a scale score of 225 represents the minimum score to earn an achievement level of Advanced.

Table 9.4: Achievement-Level Distributions for Each End-of-Course Assessment

Test Period	Subject	Achievement	Frequency	Percentage
Summer 2010	English II	Below Basic	76	16.0
		Basic	213	44.7
		Proficient	168	35.3
		Advanced	19	4.0
		Total	476	100.0
	Algebra I	Below Basic	133	11.8
		Basic	562	49.9
		Proficient	340	30.2
		Advanced	91	8.1
Total		1,126	100.0	
Biology	Below Basic	89	23.2	
	Basic	160	41.7	
	Proficient	117	30.5	
	Advanced	18	4.7	
	Total	384	100.0	
Fall 2010	English II	Below Basic	145	7.9
		Basic	537	29.3
		Proficient	780	42.6
		Advanced	371	20.2
		Total	1,833	100.0
	Algebra I	Below Basic	353	12.9
		Basic	993	36.2
		Proficient	721	26.3
		Advanced	674	24.6
Total		2,741	100.0	
Biology	Below Basic	292	12.2	
	Basic	831	34.8	
	Proficient	760	31.8	
	Advanced	508	21.2	
	Total	2,391	100.0	

Table 9.4: Achievement-Level Distributions for Each End-of-Course Assessment (continued)

Test Period	Subject	Achievement	Frequency	Percentage
Spring 2011	English II	Below Basic	2,548	4.0
		Basic	13,463	21.1
		Proficient	30,712	48.2
		Advanced	16,959	26.6
		Total	63,682	100.0
	Algebra I	Below Basic	5,381	8.4
		Basic	18,914	29.6
		Proficient	26,590	41.5
		Advanced	13,112	20.5
		Total	63,997	100.0
	Biology	Below Basic	3,932	6.3
		Basic	19,250	31.0
		Proficient	29,029	46.8
		Advanced	9,857	15.9
		Total	62,068	100.0

9.5 Descriptive Statistics by Demographic Group

Descriptive statistics of scale scores and percentage distributions of students' achievement levels by demographic groups are summarized in Tables 9.5 through 9.20.

The demographic variables included are gender (Tables 9.5 and 9.13), ethnicity (Tables 9.6 and 9.14), migrant status (Tables 9.7 and 9.15), free and reduced lunch (FRL) (Tables 9.8 and 9.16), limited English proficient (LEP) (Tables 9.9 and 9.17), Title I (Tables 9.10 and 9.18), individualized education program (IEP) (Tables 9.11 and 9.19), and accommodations (Tables 9.12 and 9.20). Note that for certain cells in these tables (particularly for Fall and Summer where *n*-counts are low), when the total *n*-count is less than 30, the descriptive statistics are not reported.

Table 9.5: Scale Score Distributions by Demographic Group—Gender

Test Period	Subject	Gender	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
Summer 2010	English II	Female	169	167	243	195.37	15.681
		Male	300	167	250	195.18	15.296
	Algebra I	Female	507	151	250	192.35	17.893
		Male	587	138	250	193.87	19.084
	Biology	Female	172	143	245	189.13	17.448
		Male	207	153	250	193.07	17.710
Fall 2010	English II	Female	828	154	250	207.89	18.311
		Male	992	159	250	202.68	18.435
	Algebra I	Female	1284	116	250	203.18	25.544
		Male	1435	116	250	201.66	25.997
	Biology	Female	1106	146	250	201.60	21.021
		Male	1272	131	250	202.45	22.494
Spring 2011	English II	Female	31280	131	250	211.73	17.748
		Male	32284	131	250	208.11	17.904
	Algebra I	Female	32161	115	250	204.09	20.456
		Male	31680	115	250	205.25	21.676
	Biology	Female	30986	119	250	203.57	17.872
		Male	30955	119	250	206.08	19.079

Table 9.6: Scale Score Distributions by Demographic Group—Ethnicity

Test Period	Subject	Ethnicity	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
Summer 2010	English II	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	N/A	N/A	N/A	N/A	N/A
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	234	167	243	192.44	13.973
		Hispanic	N/A	N/A	N/A	N/A	N/A
		White (not Hispanic)	195	167	250	198.73	16.548
		Multi-racial	N/A	N/A	N/A	N/A	N/A
	Algebra I	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	N/A	N/A	N/A	N/A	N/A
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	473	138	231	185.77	14.878
		Hispanic	55	168	242	191.76	13.588
		White (not Hispanic)	513	156	250	200.55	19.036
		Multi-racial	N/A	N/A	N/A	N/A	N/A
	Biology	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	N/A	N/A	N/A	N/A	N/A
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	212	143	250	188.48	16.196
		Hispanic	N/A	N/A	N/A	N/A	N/A
		White (not Hispanic)	134	157	245	196.94	18.423
		Multi-racial	N/A	N/A	N/A	N/A	N/A

Table 9.6: Scale Score Distributions by Demographic Group—Ethnicity (continued)

Test Period	Subject	Ethnicity	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
Fall 2010	English II	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	N/A	N/A	N/A	N/A	N/A
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	342	159	250	196.40	15.638
		Hispanic	100	163	236	200.43	16.974
		White (not Hispanic)	1,317	159	250	207.79	18.661
		Multi-racial	N/A	N/A	N/A	N/A	N/A
	Algebra I	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	54	161	250	218.83	25.230
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	506	140	250	185.68	19.269
		Hispanic	124	131	250	201.74	25.406
		White (not Hispanic)	1,960	116	250	206.58	25.443
		Multi-racial	40	161	247	201.40	25.165
	Biology	American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
		Asian	89	163	250	218.57	21.875
		Pacific Islander	N/A	N/A	N/A	N/A	N/A
		Black (not Hispanic)	350	151	245	187.59	17.431
		Hispanic	79	131	250	199.92	24.087
		White (not Hispanic)	1,804	146	250	204.25	21.110
		Multi-racial	32	159	245	203.78	21.498

Table 9.6: Scale Score Distributions by Demographic Group—Ethnicity (continued)

Test Period	Subject	Ethnicity	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
Spring 2011	English II	American Indian/ Alaskan Native	309	167	250	207.31	17.132
		Asian	1,187	164	250	215.10	19.609
		Pacific Islander	60	173	240	207.50	17.559
		Black (not Hispanic)	9,815	131	250	200.24	15.779
		Hispanic	2,327	151	250	205.48	17.438
		White (not Hispanic)	49,093	131	250	211.93	17.623
		Multi-racial	614	164	250	210.23	16.688
	Algebra I	American Indian/ Alaskan Native	321	151	250	200.15	20.913
		Asian	1,197	139	250	215.18	22.925
		Pacific Islander	71	160	246	204.25	19.919
		Black (not Hispanic)	9,920	115	250	191.57	18.759
		Hispanic	2,319	139	250	200.59	19.946
		White (not Hispanic)	48,993	115	250	207.34	20.442
		Multi-racial	856	151	250	203.17	19.835
	Biology	American Indian/ Alaskan Native	313	141	250	203.27	17.485
		Asian	1,197	119	250	210.37	21.144
		Pacific Islander	76	164	245	197.96	18.580
		Black (not Hispanic)	9,897	133	250	192.70	16.848
		Hispanic	2,275	141	250	199.47	17.631
		White (not Hispanic)	47,331	119	250	207.47	17.718
		Multi-racial	696	152	250	207.89	18.622

Table 9.7: Scale Score Distributions by Demographic Group—Migrant Status

Test Period	Subject	Migrant	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	476	167	250	195.17	15.411
		Yes	N/A	N/A	N/A	N/A	N/A
	Algebra I	No	1,126	138	250	193.30	18.749
		Yes	N/A	N/A	N/A	N/A	N/A
	Biology	No	384	143	250	191.16	17.635
		Yes	N/A	N/A	N/A	N/A	N/A
Fall 2010	English II	No	1,832	154	250	205.00	18.521
		Yes	N/A	N/A	N/A	N/A	N/A
	Algebra I	No	2,741	116	250	202.13	25.855
		Yes	N/A	N/A	N/A	N/A	N/A
	Biology	No	2,390	131	250	201.96	21.845
		Yes	N/A	N/A	N/A	N/A	N/A
Spring 2011	English II	No	63,649	131	250	209.87	17.927
		Yes	33	175	232	198.24	13.029
	Algebra I	No	63,961	115	250	204.64	21.084
		Yes	36	160	250	191.22	22.022
	Biology	No	62,026	119	250	204.80	18.545
		Yes	42	167	221	194.24	13.884

Table 9.8: Scale Score Distributions by Demographic Group—Free and Reduced Lunch

Test Period	Subject	FRL	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	226	167	250	195.94	15.878
		Yes	250	167	250	194.46	14.973
	Algebra I	No	543	138	250	198.05	19.754
		Yes	583	138	250	188.88	16.597
	Biology	No	179	153	245	194.32	17.968
		Yes	205	143	250	188.40	16.902
Fall 2010	English II	No	1,016	154	250	209.34	18.813
		Yes	817	159	250	199.58	16.679
	Algebra I	No	1,633	116	250	209.08	25.839
		Yes	1,108	140	250	191.89	22.232
	Biology	No	1,554	146	250	207.64	21.275
		Yes	837	131	250	191.40	18.751
Spring 2011	English II	No	38,467	143	250	214.25	17.228
		Yes	25,215	131	250	203.17	16.873
	Algebra I	No	38,065	115	250	209.38	20.617
		Yes	25,932	115	250	197.67	19.798
	Biology	No	38,160	119	250	208.86	17.884
		Yes	23,908	119	250	198.30	17.711

Table 9.9: Scale Score Distributions by Demographic Group—Limited English Proficient

Test Period	Subject	LEP	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	466	167	250	195.20	15.452
		Yes	N/A	N/A	N/A	N/A	N/A
	Algebra I	No	1,091	138	250	193.63	18.787
		Yes	35	156	213	183.09	14.321
	Biology	No	378	143	250	191.38	17.675
		Yes	N/A	N/A	N/A	N/A	N/A
Fall 2010	English II	No	1,794	154	250	205.30	18.465
		Yes	39	163	225	190.28	15.791
	Algebra I	No	2,703	116	250	202.26	25.841
		Yes	38	131	250	192.74	25.484
	Biology	No	2,370	131	250	202.06	21.826
		Yes	N/A	N/A	N/A	N/A	N/A
Spring 2011	English II	No	62,778	131	250	210.08	17.875
		Yes	904	160	250	194.81	14.882
	Algebra I	No	62,945	115	250	204.79	21.060
		Yes	1052	139	250	195.52	20.695
	Biology	No	61,062	119	250	205.03	18.459
		Yes	1,006	119	250	190.41	17.993

Table 9.10: Scale Score Distributions by Demographic Group—Title I

Test Period	Subject	Title I	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	435	167	250	195.25	15.077
		Yes	41	167	250	194.29	18.790
	Algebra I	No	992	138	250	194.06	18.864
		Yes	134	145	242	187.64	16.895
	Biology	No	357	153	250	191.95	17.082
		Yes	N/A	N/A	N/A	N/A	N/A
Fall 2010	English II	No	1,703	154	250	205.88	18.385
		Yes	130	163	250	193.32	16.485
	Algebra I	No	2,584	116	250	203.18	25.918
		Yes	157	147	247	184.87	17.355
	Biology	No	2,305	131	250	202.55	21.712
		Yes	86	151	245	186.06	19.291
Spring 2011	English II	No	59,776	131	250	210.41	17.813
		Yes	3,906	143	250	201.46	17.561
	Algebra I	No	58,833	115	250	205.53	20.851
		Yes	5,164	115	250	194.44	21.079
	Biology	No	58,153	119	250	205.54	18.317
		Yes	3,915	133	250	193.63	18.311

Table 9.11: Scale Score Distributions by Demographic Group—Students with IEPs

Test Period	Subject	IEP	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	403	167	250	196.17	15.584
		Yes	73	167	243	189.60	13.180
	Algebra I	No	1,006	138	250	194.62	18.663
		Yes	120	138	225	182.23	15.607
	Biology	No	342	143	250	191.55	17.460
		Yes	42	153	237	187.95	18.917
Fall 2010	English II	No	1,644	154	250	206.70	18.035
		Yes	189	163	250	190.08	16.009
	Algebra I	No	2,548	116	250	203.65	25.647
		Yes	193	116	250	182.09	19.483
	Biology	No	2,167	131	250	203.11	21.705
		Yes	224	146	245	190.73	19.950
Spring 2011	English II	No	57,154	131	250	211.96	16.983
		Yes	6528	151	250	191.47	15.265
	Algebra I	No	59,209	115	250	206.09	20.524
		Yes	4788	115	250	186.67	19.635
	Biology	No	57,643	119	250	205.86	18.129
		Yes	4425	133	250	190.80	18.204

Table 9.12: Scale Score Distributions by Demographic Group—Students with Accommodations

Test Period	Subject	Accom.	N	Minimum	Maximum	Mean	SD
Summer 2010	English II	No	469	167	250	195.27	15.471
		Yes	N/A	N/A	N/A	N/A	N/A
	Algebra I	No	1,119	138	250	193.40	18.742
		Yes	N/A	N/A	N/A	N/A	N/A
	Biology	No	377	143	250	191.39	17.561
		Yes	N/A	N/A	N/A	N/A	N/A
Fall 2010	English II	No	1,786	154	250	205.30	18.426
		Yes	47	166	250	192.89	18.824
	Algebra I	No	2,667	116	250	202.71	25.797
		Yes	74	147	240	181.11	18.095
	Biology	No	2,351	131	250	202.10	21.813
		Yes	40	155	245	193.38	22.144
Spring 2011	English II	No	60,205	131	250	211.08	17.363
		Yes	3,477	151	250	188.87	14.146
	Algebra I	No	62,485	115	250	205.13	20.908
		Yes	1,512	130	250	184.12	18.010
	Biology	No	60,519	119	250	205.19	18.387
		Yes	1,549	133	250	189.03	17.783

Table 9.13: Achievement-Level Distributions by Gender

Test Period	Subject	Gender	Achievement Level	Frequency	Percentage
Summer 2010	English II	Female	Below Basic	27	16.0
			Basic	74	43.8
			Proficient	58	34.3
			Advanced	10	5.9
	Total		169	100.0	
	Male	Below Basic	48	16.0	
		Basic	135	45.0	
		Proficient	108	36.0	
		Advanced	9	3.0	
		Total	300	100.0	
	Algebra I	Female	Below Basic	59	11.6
			Basic	265	52.3
Proficient			143	28.2	
Advanced			40	7.9	
Total	507		100.0		
Male	Below Basic	68	11.6		
	Basic	287	48.9		
	Proficient	186	31.7		
	Advanced	46	7.8		
	Total	587	100.0		
Biology	Female	Below Basic	44	25.6	
		Basic	71	41.3	
		Proficient	51	29.7	
		Advanced	6	3.5	
Total		172	100.0		
Male	Below Basic	42	20.3		
	Basic	88	42.5		
	Proficient	65	31.4		
	Advanced	12	5.8		
	Total	207	100.0		

Table 9.13: Achievement-Level Distributions by Gender (continued)

Test Period	Subject	Gender	Achievement Level	Frequency	Percentage
Fall 2010	English II	Female	Below Basic	45	5.4
			Basic	205	24.8
			Proficient	377	45.5
			Advanced	201	24.3
			Total	828	100.0
		Male	Below Basic	99	10.0
	Basic	328	33.1		
	Proficient	395	39.8		
	Advanced	170	17.1		
	Total	992	100.0		
	Algebra I	Female	Below Basic	155	12.1
			Basic	435	33.9
			Proficient	380	29.6
			Advanced	314	24.5
			Total	1,284	100.0
		Male	Below Basic	187	13.0
	Basic	548	38.2		
	Proficient	340	23.7		
Advanced	360	25.1			
Total	1,435	100.0			
Biology	Female	Below Basic	119	10.8	
		Basic	406	36.7	
		Proficient	363	32.8	
		Advanced	218	19.7	
		Total	1,106	100.0	
	Male	Below Basic	169	13.3	
Basic	417	32.8			
Proficient	396	31.1			
Advanced	290	22.8			
Total	1,272	100.0			

Table 9.13: Achievement-Level Distributions by Gender (continued)

Test Period	Subject	Gender	Achievement Level	Frequency	Percentage
Spring 2011	English II	Female	Below Basic	856	2.7
			Basic	6,026	19.3
			Proficient	15,068	48.2
			Advanced	9,330	29.8
			Total	31,280	100.0
		Male	Below Basic	1,674	5.2
	Basic	7,386	22.9		
	Proficient	15,604	48.3		
	Advanced	7,620	23.6		
	Total	32,284	100.0		
	Algebra I	Female	Below Basic	2,548	7.9
			Basic	9,820	30.5
			Proficient	13,737	42.7
			Advanced	6,056	18.8
			Total	32,161	100.0
		Male	Below Basic	2,793	8.8
	Basic	9,039	28.5		
	Proficient	12,804	40.4		
Advanced	7,044	22.2			
Total	31,680	100.0			
Biology	Female	Below Basic	1,924	6.2	
		Basic	10,502	33.9	
		Proficient	14,320	46.2	
		Advanced	4,240	13.7	
		Total	30,986	100.0	
	Male	Below Basic	1,969	6.4	
Basic	8,700	28.1			
Proficient	14,677	47.4			
Advanced	5,609	18.1			
Total	30,955	100.0			

Table 9.14: Achievement-Level Distribution by Ethnicity

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Summer 2010	English II	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	43	18.4
			Basic	118	50.4
			Proficient	66	28.2
			Advanced	7	3.0
			Total	234	100.0
		Hispanic	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		White (not Hispanic)	Below Basic	27	13.8
			Basic	73	37.4
			Proficient	84	43.1
Advanced	11		5.6		
Total	195		100.0		
Multi-racial	Below Basic	N/A	N/A		
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Summer 2010	Algebra I	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	89	18.8
			Basic	276	58.4
			Proficient	104	22.0
			Advanced	4	0.8
			Total	473	100.0
		Hispanic	Below Basic	2	3.6
			Basic	36	65.5
			Proficient	16	29.1
			Advanced	1	1.8
			Total	55	100.0
		White (not Hispanic)	Below Basic	27	5.3
			Basic	209	40.7
			Proficient	200	39.0
Advanced	77		15.0		
Total	513		100.0		
Multi-racial	Below Basic	N/A	N/A		
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Summer 2010	Biology	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	52	24.5
			Basic	101	47.6
			Proficient	53	25.0
			Advanced	6	2.8
			Total	212	100.0
		Hispanic	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		White (not Hispanic)	Below Basic	23	17.2
			Basic	45	33.6
			Proficient	56	41.8
Advanced	10		7.5		
Total	134		100.0		
Multi-racial	Below Basic	N/A	N/A		
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Fall 2010	English II	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	48	14.0
			Basic	140	40.9
			Proficient	139	40.6
			Advanced	15	4.4
			Total	342	100.0
		Hispanic	Below Basic	9	9.0
			Basic	37	37.0
			Proficient	41	41.0
			Advanced	13	13.0
			Total	100	100.0
		White (not Hispanic)	Below Basic	84	6.4
			Basic	332	25.2
			Proficient	566	43.0
Advanced	335		25.4		
Total	1,317		100.0		
Multi-racial	Below Basic	N/A	N/A		
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Fall 2010	Algebra I	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	4	7.4
			Basic	7	13.0
			Proficient	17	31.5
			Advanced	26	48.1
			Total	54	100.0
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	135	26.7
			Basic	259	51.2
			Proficient	87	17.2
			Advanced	25	4.9
			Total	506	100.0
		Hispanic	Below Basic	15	12.1
			Basic	47	37.9
			Proficient	33	26.6
			Advanced	29	23.4
			Total	124	100.0
		White (not Hispanic)	Below Basic	170	8.7
			Basic	641	32.7
			Proficient	566	28.9
Advanced	583		29.7		
Total	1,960		100.0		
Multi-racial	Below Basic	5	12.5		
	Basic	16	40.0		
	Proficient	10	25.0		
	Advanced	9	22.5		
	Total	5	12.5		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Fall 2010	Biology	American Indian/ Alaskan Native	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Asian	Below Basic	4	4.5
			Basic	14	15.7
			Proficient	25	28.1
			Advanced	46	51.7
			Total	89	100.0
		Pacific Islander	Below Basic	N/A	N/A
			Basic	N/A	N/A
			Proficient	N/A	N/A
			Advanced	N/A	N/A
			Total	N/A	N/A
		Black (not Hispanic)	Below Basic	95	27.1
			Basic	176	50.3
			Proficient	60	17.1
			Advanced	19	5.4
			Total	350	100.0
		Hispanic	Below Basic	12	15.2
			Basic	28	35.4
			Proficient	23	29.1
			Advanced	16	20.3
			Total	79	100.0
		White (not Hispanic)	Below Basic	168	9.3
			Basic	585	32.4
			Proficient	637	35.3
Advanced	414		22.9		
Total	1,804		100.0		
Multi-racial	Below Basic	3	9.4		
	Basic	9	28.1		
	Proficient	11	34.4		
	Advanced	9	28.1		
	Total	32	100.0		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Spring 2011	English II	American Indian/ Alaskan Native	Below Basic	11	3.6
			Basic	82	26.5
			Proficient	154	49.8
			Advanced	62	20.1
			Total	309	100.0
		Asian	Below Basic	32	2.7
			Basic	212	17.9
			Proficient	481	40.5
			Advanced	462	38.9
			Total	1,187	100.0
		Pacific Islander	Below Basic	4	6.7
			Basic	13	21.7
			Proficient	32	53.3
			Advanced	11	18.3
			Total	60	100.0
		Black (not Hispanic)	Below Basic	810	8.3
			Basic	3,532	36.0
			Proficient	4,562	46.5
			Advanced	911	9.3
			Total	9,815	100.0
		Hispanic	Below Basic	130	5.6
			Basic	635	27.3
			Proficient	1,149	49.4
			Advanced	413	17.7
			Total	2,327	100.0
		White (not Hispanic)	Below Basic	1,504	3.1
			Basic	8,787	17.9
			Proficient	23,906	48.7
Advanced	14,896		30.3		
Total	49,093		100.0		
Multi-racial	Below Basic	22	3.6		
	Basic	113	18.4		
	Proficient	329	53.6		
	Advanced	150	24.4		
	Total	614	100.0		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Spring 2011	Algebra I	American Indian/ Alaskan Native	Below Basic	35	10.9
			Basic	116	36.1
			Proficient	120	37.4
			Advanced	50	15.6
			Total	321	100.0
		Asian	Below Basic	59	4.9
			Basic	226	18.9
			Proficient	444	37.1
			Advanced	468	39.1
			Total	1197	100.0
		Pacific Islander	Below Basic	7	9.9
			Basic	23	32.4
			Proficient	26	36.6
			Advanced	15	21.1
			Total	71	100.0
		Black (not Hispanic)	Below Basic	2,030	20.5
			Basic	4,342	43.8
			Proficient	2,978	30.0
			Advanced	570	5.7
			Total	9,920	100.0
		Hispanic	Below Basic	238	10.3
			Basic	823	35.5
			Proficient	916	39.5
			Advanced	342	14.7
			Total	2,319	100.0
		White (not Hispanic)	Below Basic	2,868	5.9
			Basic	12,998	26.5
			Proficient	21,636	44.2
Advanced	11,491		23.5		
Total	48,993		100.0		
Multi-racial	Below Basic	65	7.6		
	Basic	271	31.7		
	Proficient	378	44.2		
	Advanced	142	16.6		
	Total	856	100.0		

Table 9.14: Achievement-Level Distribution by Ethnicity (continued)

Test Period	Subject	Ethnicity	Achievement Level	Frequency	Percentage
Spring 2011	Biology	American Indian/ Alaskan Native	Below Basic	19	6.1
			Basic	98	31.3
			Proficient	161	51.4
			Advanced	35	11.2
			Total	313	100.0
		Asian	Below Basic	68	5.7
			Basic	277	23.1
			Proficient	519	43.4
			Advanced	333	27.8
			Total	1,197	100.0
		Pacific Islander	Below Basic	10	13.2
			Basic	26	34.2
			Proficient	34	44.7
			Advanced	6	7.9
			Total	76	100.0
		Black (not Hispanic)	Below Basic	1,731	17.5
			Basic	4,717	47.7
			Proficient	3,024	30.6
			Advanced	425	4.3
			Total	9,897	100.0
		Hispanic	Below Basic	207	9.1
			Basic	890	39.1
			Proficient	962	42.3
			Advanced	216	9.5
			Total	2,275	100.0
		White (not Hispanic)	Below Basic	1,796	3.8
			Basic	12,964	27.4
			Proficient	23,904	50.5
Advanced	8,667		18.3		
Total	47,331		100.0		
Multi-racial	Below Basic	30	4.3		
	Basic	181	26.0		
	Proficient	347	49.9		
	Advanced	138	19.8		
	Total	696	100.0		

Table 9.15: Achievement-Level Distribution—Migrant

Test Period	Subject	Migrant	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	76	16.0
			Basic	213	44.7
			Proficient	168	35.3
			Advanced	19	4.0
			Total	476	100.0
		Yes	Below Basic	N/A	N/A
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		
	Algebra I	No	Below Basic	133	11.8
			Basic	562	49.9
Proficient			340	30.2	
Advanced			91	8.1	
Total			1,126	100.0	
Yes		Below Basic	N/A	N/A	
Basic	N/A	N/A			
Proficient	N/A	N/A			
Advanced	N/A	N/A			
Total	N/A	N/A			
Biology	No	Below Basic	89	23.2	
		Basic	160	41.7	
		Proficient	117	30.5	
		Advanced	18	4.7	
		Total	384	100.0	
	Yes	Below Basic	N/A	N/A	
Basic	N/A	N/A			
Proficient	N/A	N/A			
Advanced	N/A	N/A			
Total	N/A	N/A			

Table 9.15: Achievement-Level Distribution—Migrant (continued)

Test Period	Subject	Migrant	Achievement Level	Frequency	Percentage
Fall 2010	English II	No	Below Basic	144	7.9
			Basic	537	29.3
			Proficient	780	42.6
			Advanced	371	20.3
			Total	1,832	100.0
		Yes	N/A	N/A	
	Algebra I	No	Below Basic	353	12.9
			Basic	993	36.2
			Proficient	721	26.3
			Advanced	674	24.6
			Total	2,741	100.0
		Yes	N/A	N/A	
	Biology	No	Below Basic	292	12.2
			Basic	830	34.7
			Proficient	760	31.8
			Advanced	508	21.3
			Total	2,390	100.0
		Yes	N/A	N/A	

Table 9.15: Achievement-Level Distribution—Migrant (continued)

Test Period	Subject	Migrant	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	2,547	4.0
			Basic	13,448	21.1
			Proficient	30,696	48.2
			Advanced	16,958	26.6
			Total	63,649	100.0
	English II	Yes	Below Basic	1	3.0
			Basic	15	45.5
			Proficient	16	48.5
			Advanced	1	3.0
			Total	33	100.0
	Algebra I	No	Below Basic	5,370	8.4
			Basic	18,902	29.6
Proficient			26,579	41.6	
Advanced			13,110	20.5	
		Total	63,961	100.0	
Algebra I	Yes	Below Basic	11	30.6	
		Basic	12	33.3	
		Proficient	11	30.6	
		Advanced	2	5.6	
		Total	36	100.0	
Biology	No	Below Basic	3,928	6.3	
		Basic	19,229	31.0	
		Proficient	29,012	46.8	
		Advanced	9,857	15.9	
		Total	62,026	100.0	
Biology	Yes	Below Basic	4	9.5	
		Basic	21	50.0	
		Proficient	17	40.5	
		Advanced	42	100.0	
		Total	4	9.5	

Table 9.16: Achievement-Level Distribution—FRL

Test Period	Subject	FRL	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	35	15.5
			Basic	97	42.9
			Proficient	82	36.3
			Advanced	12	5.3
			Total	226	100.0
			Yes	Below Basic	41
	Basic	116		46.4	
	Proficient	86		34.4	
	Advanced	7		2.8	
	Total	250		100.0	
	Algebra I	No		Below Basic	46
			Basic	232	42.7
Proficient			197	36.3	
Advanced			68	12.5	
Total			543	100.0	
Yes			Below Basic	87	14.9
	Basic	330	56.6		
	Proficient	143	24.5		
	Advanced	23	3.9		
	Total	583	100.0		
	Biology	No	Below Basic	35	19.6
Basic			65	36.3	
Proficient			69	38.5	
Advanced			10	5.6	
Total			179	100.0	
Yes			Below Basic	54	26.3
	Basic	95	46.3		
	Proficient	48	23.4		
	Advanced	8	3.9		
	Total	205	100.0		

Table 9.16: Achievement-Level Distribution—FRL (continued)

Test Period	Subject	FRL	Achievement Level	Frequency	Percentage	
Fall 2010	English II	No	Below Basic	56	5.5	
			Basic	232	22.8	
			Proficient	438	43.1	
			Advanced	290	28.5	
				Total	1,016	100.0
	English II	Yes	Below Basic	89	10.9	
			Basic	305	37.3	
			Proficient	342	41.9	
			Advanced	81	9.9	
				Total	817	100.0
	Algebra I	No	Below Basic	136	8.3	
			Basic	463	28.4	
Proficient			482	29.5		
Advanced			552	33.8		
			Total	1,633	100.0	
Algebra I	Yes	Below Basic	217	19.6		
		Basic	530	47.8		
		Proficient	239	21.6		
		Advanced	122	11.0		
			Total	1,108	100.0	
Biology	No	Below Basic	115	7.4		
		Basic	429	27.6		
		Proficient	570	36.7		
		Advanced	440	28.3		
			Total	1,554	100.0	
Biology	Yes	Below Basic	177	21.1		
		Basic	402	48.0		
		Proficient	190	22.7		
		Advanced	68	8.1		
			Total	837	100.0	

Table 9.16: Achievement-Level Distribution—FRL (continued)

Test Period	Subject	FRL	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	834	2.2
			Basic	5,569	14.5
			Proficient	18,765	48.8
			Advanced	13,299	34.6
			Total	38,467	100.0
	Yes	Below Basic	1,714	6.8	
		Basic	7,894	31.3	
		Proficient	11,947	47.4	
		Advanced	3,660	14.5	
		Total	25,215	100.0	
Algebra I	No	Below Basic	1,939	5.1	
		Basic	9,116	23.9	
		Proficient	16,818	44.2	
		Advanced	10,192	26.8	
		Total	38,065	100.0	
Yes	Below Basic	3,442	13.3		
	Basic	9,798	37.8		
	Proficient	9,772	37.7		
	Advanced	2,920	11.3		
	Total	25,932	100.0		
Biology	No	Below Basic	1,314	3.4	
		Basic	9,463	24.8	
		Proficient	19,545	51.2	
		Advanced	7,838	20.5	
		Total	38,160	100.0	
Yes	Below Basic	2,618	11.0		
	Basic	9,787	40.9		
	Proficient	9,484	39.7		
	Advanced	2,019	8.4		
	Total	23,908	100.0		

Table 9.17: Achievement-Level Distribution—LEP

Test Period	Subject	LEP	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	75	16.1
			Basic	208	44.6
			Proficient	164	35.2
			Advanced	19	4.1
			Total	466	100.0
		Yes	Below Basic	N/A	N/A
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		
	Algebra I	No	Below Basic	125	11.5
			Basic	541	49.6
Proficient			334	30.6	
Advanced			91	8.3	
Total			1,091	100.0	
Yes		Below Basic	8	22.9	
Basic	21	60.0			
Proficient	6	17.1			
Advanced	35	100.0			
Total	8	22.9			
Biology	No	Below Basic	86	22.8	
		Basic	157	41.5	
		Proficient	117	31.0	
		Advanced	18	4.8	
		Total	378	100.0	
	Yes	Below Basic	N/A	N/A	
Basic	N/A	N/A			
Proficient	N/A	N/A			
Advanced	N/A	N/A			
Total	N/A	N/A			

Table 9.17: Achievement-Level Distribution—LEP (continued)

Test Period	Subject	LEP	Achievement Level	Frequency	Percentage
Fall 2010	English II	No	Below Basic	135	7.5
			Basic	518	28.9
			Proficient	772	43.0
			Advanced	369	20.6
			Total	1,794	100.0
	English II	Yes	Below Basic	10	25.6
			Basic	19	48.7
			Proficient	8	20.5
			Advanced	2	5.1
			Total	39	100.0
Algebra I	No	Below Basic	346	12.8	
		Basic	975	36.1	
		Proficient	713	26.4	
		Advanced	669	24.8	
		Total	2,703	100.0	
Algebra I	Yes	Below Basic	7	18.4	
		Basic	18	47.4	
		Proficient	8	21.1	
		Advanced	5	13.2	
		Total	38	100.0	
Biology	No	Below Basic	285	12.0	
		Basic	825	34.8	
		Proficient	754	31.8	
		Advanced	506	21.4	
		Total	2,370	100.0	
Biology	Yes	Below Basic	N/A	N/A	
		Basic	N/A	N/A	
		Proficient	N/A	N/A	
		Advanced	N/A	N/A	
		Total	N/A	N/A	

Table 9.17: Achievement-Level Distribution—LEP (continued)

Test Period	Subject	LEP	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	2,437	3.9
			Basic	13,018	20.7
			Proficient	30,410	48.4
			Advanced	16,913	26.9
			Total	62,778	100.0
		Yes	Below Basic	111	12.3
	Basic	445	49.2		
	Proficient	302	33.4		
	Advanced	46	5.1		
	Total	904	100.0		
	Algebra I	No	Below Basic	5,205	8.3
			Basic	18,474	29.3
Proficient			26,271	41.7	
Advanced			12,995	20.6	
Total			62,945	100.0	
Yes		Below Basic	176	16.7	
Basic	440	41.8			
Proficient	319	30.3			
Advanced	117	11.1			
Total	1,052	100.0			
Biology	No	Below Basic	3,702	6.1	
		Basic	18,786	30.8	
		Proficient	28,767	47.1	
		Advanced	9,807	16.1	
		Total	61,062	100.0	
	Yes	Below Basic	230	22.9	
Basic	464	46.1			
Proficient	262	26.0			
Advanced	50	5.0			
Total	1,006	100.0			

Table 9.18: Achievement-Level Distribution—Title I

Test Period	Subject	Title I	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	67	15.4
			Basic	196	45.1
			Proficient	156	35.9
			Advanced	16	3.7
			Total	435	100.0
	English II	Yes	Below Basic	9	22.0
			Basic	17	41.5
			Proficient	12	29.3
			Advanced	3	7.3
			Total	41	100.0
	Algebra I	No	Below Basic	107	10.8
			Basic	494	49.8
Proficient			302	30.4	
Advanced			89	9.0	
		Total	992	100.0	
Algebra I	Yes	Below Basic	26	19.4	
		Basic	68	50.7	
		Proficient	38	28.4	
		Advanced	2	1.5	
		Total	134	100.0	
Biology	No	Below Basic	73	20.4	
		Basic	155	43.4	
		Proficient	112	31.4	
		Advanced	17	4.8	
		Total	357	100.0	
Biology	Yes	Below Basic	N/A	N/A	
		Basic	N/A	N/A	
		Proficient	N/A	N/A	
		Advanced	N/A	N/A	
		Total	N/A	N/A	

Table 9.18: Achievement-Level Distribution—Title I (continued)

Test Period	Subject	Title I	Achievement Level	Frequency	Percentage
Fall 2010	English II	No	Below Basic	118	6.9
			Basic	482	28.3
			Proficient	737	43.3
			Advanced	366	21.5
			Total	1,703	100.0
	English II	Yes	Below Basic	27	20.8
			Basic	55	42.3
			Proficient	43	33.1
			Advanced	5	3.8
			Total	130	100.0
	Algebra I	No	Below Basic	312	12.1
			Basic	907	35.1
			Proficient	698	27.0
			Advanced	667	25.8
			Total	2,584	100.0
Algebra I	Yes	Below Basic	41	26.1	
		Basic	86	54.8	
		Proficient	23	14.6	
		Advanced	7	4.5	
		Total	157	100.0	
Biology	No	Below Basic	261	11.3	
		Basic	796	34.5	
		Proficient	745	32.3	
		Advanced	503	21.8	
		Total	2,305	100.0	
Biology	Yes	Below Basic	31	36.0	
		Basic	35	40.7	
		Proficient	15	17.4	
		Advanced	5	5.8	
		Total	86	100.0	

Table 9.18: Achievement-Level Distribution—Title I (continued)

Test Period	Subject	Title I	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	2,198	3.7
			Basic	12,115	20.3
			Proficient	29,037	48.6
			Advanced	16,426	27.5
			Total	59,776	100.0
	English II	Yes	Below Basic	350	9.0
			Basic	1,348	34.5
			Proficient	1,675	42.9
			Advanced	533	13.6
			Total	3,906	100.0
	Algebra I	No	Below Basic	4,328	7.4
			Basic	17,019	28.9
			Proficient	24,912	42.3
			Advanced	12,574	21.4
			Total	58,833	100.0
Algebra I	Yes	Below Basic	1,053	20.4	
		Basic	1,895	36.7	
		Proficient	1,678	32.5	
		Advanced	538	10.4	
		Total	5,164	100.0	
Biology	No	Below Basic	3,214	5.5	
		Basic	17,511	30.1	
		Proficient	27,814	47.8	
		Advanced	9,614	16.5	
		Total	58,153	100.0	
Biology	Yes	Below Basic	718	18.3	
		Basic	1,739	44.4	
		Proficient	1,215	31.0	
		Advanced	243	6.2	
		Total	3,915	100.0	

Table 9.19: Achievement-Level Distribution—IEP

Test Period	Subject	IEP	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	61	15.1
			Basic	171	42.4
			Proficient	153	38.0
			Advanced	18	4.5
			Total	403	100.0
		Yes	Below Basic	15	20.5
	Basic	42	57.5		
	Proficient	15	20.5		
	Advanced	1	1.4		
	Total	73	100.0		
	Algebra I	No	Below Basic	102	10.1
			Basic	492	48.9
Proficient			322	32.0	
Advanced			90	8.9	
Total			1,006	100.0	
Yes		Below Basic	31	25.8	
Basic	70	58.3			
Proficient	18	15.0			
Advanced	1	.8			
Total	120	100.0			
Biology	No	Below Basic	75	21.9	
		Basic	145	42.4	
		Proficient	105	30.7	
		Advanced	17	5.0	
		Total	342	100.0	
	Yes	Below Basic	14	33.3	
Basic	15	35.7			
Proficient	12	28.6			
Advanced	1	2.4			
Total	42	100.0			

Table 9.19: Achievement-Level Distribution—IEP (continued)

Test Period	Subject	IEP	Achievement Level	Frequency	Percentage
Fall 2010	English II	No	Below Basic	96	5.8
			Basic	444	27.0
			Proficient	741	45.1
			Advanced	363	22.1
			Total	1,644	100.0
		Yes	Below Basic	49	25.9
	Basic	93	49.2		
	Proficient	39	20.6		
	Advanced	8	4.2		
	Total	189	100.0		
	Algebra I	No	Below Basic	289	11.3
			Basic	901	35.4
Proficient			691	27.1	
Advanced			667	26.2	
Total			2,548	100.0	
Yes		Below Basic	64	33.2	
Basic	92	47.7			
Proficient	30	15.5			
Advanced	7	3.6			
Total	193	100.0			
Biology	No	Below Basic	228	10.5	
		Basic	743	34.3	
		Proficient	708	32.7	
		Advanced	488	22.5	
		Total	2,167	100.0	
	Yes	Below Basic	64	28.6	
Basic	88	39.3			
Proficient	52	23.2			
Advanced	20	8.9			
Total	224	100.0			

Table 9.19: Achievement-Level Distribution—IEP (continued)

Test Period	Subject	IEP	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	1,237	2.2
			Basic	10,266	18.0
			Proficient	28,962	50.7
			Advanced	16,689	29.2
			Total	57,154	100.0
		Yes	Below Basic	1,311	20.1
	Basic	3,197	49.0		
	Proficient	1,750	26.8		
	Advanced	270	4.1		
	Total	6,528	100.0		
	Algebra I	No	Below Basic	3,888	6.6
			Basic	16,839	28.4
Proficient			25,635	43.3	
Advanced			12,847	21.7	
Total			59,209	100.0	
Yes		Below Basic	1,493	31.2	
Basic	2,075	43.3			
Proficient	955	19.9			
Advanced	265	5.5			
Total	4,788	100.0			
Biology	No	Below Basic	2,876	5.0	
		Basic	17,281	30.0	
		Proficient	27,854	48.3	
		Advanced	9,632	16.7	
		Total	57,643	100.0	
	Yes	Below Basic	1,056	23.9	
Basic	1,969	44.5			
Proficient	1,175	26.6			
Advanced	225	5.1			
Total	4,425	100.0			

Table 9.20: Achievement-Level Distribution—Accommodations

Test Period	Subject	Accommodations	Achievement Level	Frequency	Percentage
Summer 2010	English II	No	Below Basic	76	16.2
			Basic	207	44.1
			Proficient	167	35.6
			Advanced	19	4.1
			Total	469	100.0
		Yes	Below Basic	N/A	N/A
	Basic	N/A	N/A		
	Proficient	N/A	N/A		
	Advanced	N/A	N/A		
	Total	N/A	N/A		
	Algebra I	No	Below Basic	131	11.7
			Basic	558	49.9
Proficient			339	30.3	
Advanced			91	8.1	
Total			1,119	100.0	
Yes		Below Basic	N/A	N/A	
Basic	N/A	N/A			
Proficient	N/A	N/A			
Advanced	N/A	N/A			
Total	N/A	N/A			
Biology	No	Below Basic	86	22.8	
		Basic	157	41.6	
		Proficient	116	30.8	
		Advanced	18	4.8	
		Total	377	100.0	
	Yes	Below Basic	N/A	N/A	
Basic	N/A	N/A			
Proficient	N/A	N/A			
Advanced	N/A	N/A			
Total	N/A	N/A			

Table 9.20: Achievement-Level Distribution—Accommodations (continued)

Test Period	Subject	Accommodations	Achievement Level	Frequency	Percentage
Fall 2010	English II	No	Below Basic	133	7.4
			Basic	516	28.9
			Proficient	769	43.1
			Advanced	368	20.6
			Total	1,786	100.0
			Yes	Below Basic	12
	Basic	21		44.7	
	Proficient	11		23.4	
	Advanced	3		6.4	
	Total	47		100.0	
	Algebra I	No		Below Basic	327
			Basic	958	35.9
			Proficient	710	26.6
			Advanced	672	25.2
			Total	2,667	100.0
			Yes	Below Basic	26
	Basic	35		47.3	
	Proficient	11		14.9	
Advanced	2	2.7			
Total	74	100.0			
Biology	No	Below Basic		280	11.9
		Basic	819	34.8	
		Proficient	749	31.9	
		Advanced	503	21.4	
		Total	2,351	100.0	
		Yes	Below Basic	12	30.0
Basic	12		30.0		
Proficient	11		27.5		
Advanced	5		12.5		
Total	40		100.0		

Table 9.20: Achievement-Level Distribution—Accommodations (continued)

Test Period	Subject	Accommodations	Achievement Level	Frequency	Percentage
Spring 2011	English II	No	Below Basic	1,722	2.9
			Basic	11,639	19.3
			Proficient	29,972	49.8
			Advanced	16,872	28.0
			Total	60,205	100.0
		Yes	Below Basic	826	23.8
	Basic	1,824	52.5		
	Proficient	740	21.3		
	Advanced	87	2.5		
	Total	3,477	100.0		
	Algebra I	No	Below Basic	4,861	7.8
			Basic	18,221	29.2
Proficient			26,343	42.2	
Advanced			13,060	20.9	
Total			62,485	100.0	
Yes		Below Basic	520	34.4	
Basic	693	45.8			
Proficient	247	16.3			
Advanced	52	3.4			
Total	1,512	100.0			
Biology	No	Below Basic	3,529	5.8	
		Basic	18,533	30.6	
		Proficient	28,663	47.4	
		Advanced	9,794	16.2	
		Total	60,519	100.0	
	Yes	Below Basic	403	26.0	
Basic	717	46.3			
Proficient	366	23.6			
Advanced	63	4.1			
Total	1,549	100.0			

CHAPTER 10: RELIABILITY

10.1 Introduction

The Missouri Department of Elementary and Secondary Education (DESE) is required by federal law to ensure that the instruments used to measure student achievement for school accountability provide reliable results. This chapter provides evidence that scores from the Missouri End-of-Course (EOC) Assessments measure student achievement in a reliable manner and that the size of the measurement error associated with reported test scores is reasonable, especially at the Proficient cut score.

10.2 Reliability and Measurement Error

10.2.1 Defining Reliability

Reliability refers to the consistency of student test scores. *Measurement error* refers to the random variability in the test scores. Both are indicators of the degree of precision in a test score. In general, measurement error and reliability are inversely related. When measurement error is large, reliability is small. Increasing reliability by minimizing measurement error is an important goal in the construction of any test.

Estimating the size of the measurement error associated with a true score is the key to estimating reliability. Errors in measurement can result from any of a multitude of factors, including environmental factors (e.g., testing conditions) and examinee factors (e.g., fatigue, stress). Feldt and Brennan (1989) note that “Quantification of the consistency and inconsistency in examinee performance constitutes the essence of reliability analysis” (p. 105). Classical test theory (CTT) provides a means for this quantification of examinee inconsistency (i.e., measurement error). This approach builds on the notion of an ideal error-free, or true, measurement score. Any observed measurement, such as test score X , is defined as a composite of true score, T , and its associated error:

$$X = T + \text{error.}$$

The definitions or assumptions in CTT lead to several important properties. For example, it can be demonstrated that observed score variance equals the sum of true score variance plus error variance:

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2.$$

The relationship among variance terms (i.e., $\sigma_x^2, \sigma_t^2, \sigma_e^2$) is critical to a more thorough understanding of important CTT concepts, including reliability and the standard error of measurement (*SEM*). For example, CTT equivalence reliability is defined as the correlation between observed scores on parallel forms, which is equal to

$$\rho_{x_1 x_2} = \sigma_t^2 / \sigma_x^2.$$

Reliability in CTT is thus conceptualized as true score variance divided by observed score variance. With just a few algebraic steps, the CTT definition of the *SEM* can be derived:

$$\sigma_e = \sigma_x \sqrt{1 - \rho_{x_1 x_2}}.$$

Although the conceptualizations of reliability and *SEM* are relatively straightforward, issues underlying the estimation of reliability are not.

10.2.2 Estimating Reliability

Reliability can be estimated via the correlation of scores on parallel forms (equivalence reliability) or from test-retest data (stability reliability), or it can be estimated from a single test administration (internal consistency reliability) using any one of a variety of techniques (e.g., Brown, 1910; Cronbach, 1951; Kuder and Richardson, 1937). A very popular index for describing internal consistency reliability based on a single test administration is Cronbach's coefficient alpha, which provides an estimate of reliability that is mathematically equivalent to the average of all possible split-half reliability estimates.

10.2.3 Sources of Measurement Error

As noted above, errors in measurement can result from environmental factors and examinee factors. To reduce other sources of measurement error, the scoring of student responses to selected response (SR) items was done electronically. Scoring error may result from improper coding or extraneous marks on scannable response sheets. The size of this sort of error is usually small and is controlled through standardized test administration procedures (including detailed instructions on how to fill out response sheets and how to erase extraneous markings) and quality control measures implemented during the scanning process.

From Spring 2008 through Spring 2010, the Performance Event (PE) and Writing Prompt (WP) items are susceptible to scoring error due to ambiguity in the scoring rubric and differences among raters. Rubrics were written to balance generality and specificity and to cover the range of student responses, while at the same time allowing raters to easily identify the response characteristics distinguishing each score category. To minimize rater error, the Assessment Resource Center (ARC) at the University of Missouri—the organization that conducted the hand scoring of the PE/WP items—thoroughly trained raters and monitored the scoring process. Only raters who met ARC's criteria for consistent scoring during training were retained as scorers.

10.3 Evidence of Raw-Score Internal Consistency

Consistency of individual student performance was estimated using Cronbach's coefficient alpha. As previously noted, coefficient alpha provides an estimate of reliability that is mathematically equivalent to the average of all possible split-half reliability estimates. Alpha is an appropriate index of internal consistency for use on untimed tests such as the MO EOC Assessments.

Separate analyses were performed for each EOC content area. For the Summer 2010, Fall 2010, and Spring 2011 test forms, PE items were not used and therefore only SR items were used in the computations. Cronbach's alpha can be interpreted as a lower bound to reliability and can be estimated using the following formula:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_{Y_i}^2}{\sigma_X^2} \right],$$

where n is the number of items, $\sigma_{Y_i}^2$ is the variance of item i , and σ_X^2 is the variance of the total score. Following this, SEM can be interpreted as "the square root of the average of the person-specific error variances of all examinees who participated in the reliability estimation experiment" (Traub, 1994, p. 114). SEM s were calculated using the following formula:

$$SEM = S_X \sqrt{1 - \alpha},$$

where S_X is the standard deviation of observed total scores. Tables 10.1 to 10.9 show the reliability coefficients (Cronbach's alpha) and SEM s based on the raw-score metric for the total population and for select student subgroups.

Table 10.1: Alpha Coefficients and Standard Errors of Measurement, English II, Summer 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	18.60	6.37	476	0.82	2.67
Gender					
Female	18.68	6.46	169	0.83	2.65
Male	18.61	6.33	300	0.82	2.69
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	N/A	N/A	N/A	N/A	N/A
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	17.51	5.94	234	0.79	2.70
Hispanic	N/A	N/A	N/A	N/A	N/A
White (not Hispanic)	20.02	6.66	195	0.84	2.63
Multi-racial	N/A	N/A	N/A	N/A	N/A
LEP					
No	18.61	6.39	466	0.83	2.67
Yes	N/A	N/A	N/A	N/A	N/A
IEP					
No	19.03	6.43	403	0.83	2.66
Yes	16.23	5.55	73	0.76	2.75
Migrant					
No	18.60	6.37	476	0.82	2.67
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	18.89	6.49	226	0.83	2.66
Yes	18.33	6.27	250	0.82	2.69
Title I					
No	18.66	6.27	435	0.82	2.68
Yes	18.00	7.41	41	0.88	2.54
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.2: Alpha Coefficients and Standard Errors of Measurement, English II, Fall 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	22.18	6.80	1,833	0.86	2.51
Gender					
Female	23.25	6.55	828	0.86	2.44
Male	21.32	6.88	992	0.86	2.56
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	N/A	N/A	N/A	N/A	N/A
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	19.09	6.25	342	0.82	2.64
Hispanic	20.55	6.61	100	0.85	2.58
White (not Hispanic)	23.18	6.70	1,317	0.87	2.46
Multi-racial	N/A	N/A	N/A	N/A	N/A
LEP					
No	22.30	6.76	1,794	0.86	2.51
Yes	16.54	6.34	39	0.83	2.61
IEP					
No	22.84	6.52	1,644	0.86	2.48
Yes	16.37	6.36	189	0.82	2.68
Migrant					
No	22.18	6.79	1,832	0.86	2.57
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	23.73	6.66	1,016	0.87	2.42
Yes	20.25	6.47	817	0.84	2.61
Title I					
No	22.52	6.70	1,703	0.86	2.49
Yes	17.74	6.47	130	0.83	2.70
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.3: Alpha Coefficients and Standard Errors of Measurement, English II, Spring 2011

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	24.63	6.42	63,682	0.86	2.43
Gender					
Female	25.29	6.15	31,280	0.85	2.38
Male	24.02	6.60	32,284	0.86	2.47
Ethnicity					
American Indian/ Alaskan Native	23.79	6.36	309	0.85	2.49
Asian	26.15	6.49	1,187	0.88	2.30
Pacific Islander	23.80	6.58	60	0.86	2.50
Black (not Hispanic)	21.15	6.39	9,815	0.83	2.62
Hispanic	23.07	6.54	2,327	0.85	2.52
White (not Hispanic)	25.39	6.16	49,093	0.85	2.39
Multi-racial	24.91	6.03	614	0.84	2.43
LEP					
No	24.72	6.39	62,778	0.86	2.42
Yes	18.86	6.22	904	0.81	2.70
IEP					
No	25.46	5.87	57,154	0.83	2.39
Yes	17.40	6.48	6,528	0.83	2.71
Migrant					
No	24.64	6.42	63,649	0.86	2.43
Yes	20.52	5.63	33	0.77	2.69
FRL					
No	26.21	5.81	38,467	0.84	2.33
Yes	22.23	6.56	25,215	0.85	2.57
Title I					
No	24.84	6.34	59,776	0.85	2.42
Yes	21.48	6.87	3,906	0.86	2.59
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.4: Alpha Coefficients and Standard Errors of Measurement, Algebra I, Summer 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	17.22	5.98	1,126	0.81	2.63
Gender					
Female	16.90	5.76	507	0.79	2.64
Male	17.42	6.06	587	0.81	2.63
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	N/A	N/A	N/A	N/A	N/A
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	14.82	4.86	473	0.69	2.69
Hispanic	16.71	4.50	55	0.64	2.72
White (not Hispanic)	19.56	5.99	513	0.82	2.56
Multi-racial	N/A	N/A	N/A	N/A	N/A
LEP					
No	17.33	5.98	1,091	0.81	2.63
Yes	13.86	4.74	35	0.67	2.73
IEP					
No	17.64	5.95	1,006	0.81	2.63
Yes	13.72	4.99	120	0.72	2.66
Migrant					
No	17.22	5.98	1,126	0.81	2.63
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	18.74	6.20	543	0.83	2.59
Yes	15.81	5.40	583	0.76	2.67
Title I					
No	17.46	6.01	992	0.81	2.63
Yes	15.46	5.46	134	0.76	2.66
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.5: Alpha Coefficients and Standard Errors of Measurement, Algebra I, Fall 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	20.14	8.03	2,741	0.91	2.48
Gender					
Female	20.49	7.92	1,284	0.90	2.48
Male	19.98	8.08	1,435	0.91	2.48
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	25.22	7.46	54	0.91	2.20
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	14.92	6.33	506	0.83	2.62
Hispanic	20.06	7.85	124	0.90	2.51
White (not Hispanic)	21.56	7.81	1,960	0.90	2.43
Multi-racial	19.98	8.03	40	0.90	2.51
LEP					
No	20.18	8.03	2,703	0.91	2.47
Yes	17.32	7.79	38	0.89	2.61
IEP					
No	20.62	7.95	2,548	0.90	2.46
Yes	13.81	6.20	193	0.82	2.61
Migrant					
No	20.14	8.03	2,741	0.91	2.48
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	22.32	7.86	1,633	0.91	2.40
Yes	16.93	7.15	1,108	0.87	2.59
Title I					
No	20.48	8.03	2,584	0.91	2.46
Yes	14.67	5.73	157	0.79	2.65
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.6: Alpha Coefficients and Standard Errors of Measurement, Algebra I, Spring 2011

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	21.44	6.61	63,997	0.86	2.47
Gender					
Female	21.30	6.45	32,161	0.85	2.47
Male	21.60	6.76	31,680	0.87	2.46
Ethnicity					
American Indian/ Alaskan Native	19.96	6.66	321	0.86	2.53
Asian	24.52	6.76	1,197	0.88	2.30
Pacific Islander	21.38	6.38	71	0.85	2.51
Black (not Hispanic)	17.24	6.22	9,920	0.83	2.59
Hispanic	20.17	6.41	2,319	0.85	2.52
White (not Hispanic)	22.32	6.33	48,993	0.85	2.44
Multi-racial	21.01	6.28	856	0.84	2.50
LEP					
No	21.49	6.60	62,945	0.86	2.47
Yes	18.46	6.68	1,052	0.85	2.58
IEP					
No	21.92	6.40	59,209	0.85	2.46
Yes	15.55	6.41	4,788	0.84	2.60
Migrant					
No	21.44	6.61	63,961	0.86	2.47
Yes	17.00	7.01	36	0.87	2.54
FRL					
No	22.93	6.31	38,065	0.85	2.41
Yes	19.25	6.44	25,932	0.84	2.56
Title I					
No	21.73	6.51	58,833	0.86	2.46
Yes	18.15	6.87	5,164	0.86	2.56
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.7: Alpha Coefficients and Standard Errors of Measurement, Biology, Summer 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	18.80	6.12	384	0.82	2.61
Gender					
Female	18.10	6.06	172	0.81	2.63
Male	19.46	6.13	207	0.82	2.59
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	N/A	N/A	N/A	N/A	N/A
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	17.87	5.66	212	0.78	2.64
Hispanic	N/A	N/A	N/A	N/A	N/A
White (not Hispanic)	20.82	6.33	134	0.84	2.54
Multi-racial	N/A	N/A	N/A	N/A	N/A
LEP					
No	18.88	6.13	378	0.82	2.61
Yes	N/A	N/A	N/A	N/A	N/A
IEP					
No	18.93	6.04	342	0.81	2.61
Yes	17.71	6.70	42	0.85	2.64
Migrant					
No	18.80	6.12	384	0.82	2.61
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	19.94	6.24	179	0.83	2.57
Yes	17.80	5.85	205	0.80	2.64
Title I					
No	19.09	5.95	357	0.81	2.61
Yes	N/A	N/A	N/A	N/A	N/A
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.8: Alpha Coefficients and Standard Errors of Measurement, Biology, Fall 2010

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	21.78	7.18	2391	0.88	2.50
Gender					
Female	21.72	6.96	1106	0.87	2.50
Male	21.90	7.34	1272	0.89	2.49
Ethnicity					
American Indian/ Alaskan Native	N/A	N/A	N/A	N/A	N/A
Asian	26.81	6.37	89	0.88	2.18
Pacific Islander	N/A	N/A	N/A	N/A	N/A
Black (not Hispanic)	16.85	6.19	350	0.82	2.66
Hispanic	21.06	7.71	79	0.90	2.48
White (not Hispanic)	22.60	6.87	1804	0.87	2.48
Multi-racial	22.41	7.06	32	0.88	2.42
LEP					
No	21.82	7.16	2370	0.88	2.50
Yes	N/A	N/A	N/A	N/A	N/A
IEP					
No	22.17	7.07	2167	0.88	2.48
Yes	18.01	7.06	224	0.86	2.61
Migrant					
No	21.78	7.18	2390	0.88	2.50
Yes	N/A	N/A	N/A	N/A	N/A
FRL					
No	23.69	6.76	1554	0.87	2.42
Yes	18.24	6.56	837	0.84	2.63
Title I					
No	21.98	7.11	2305	0.88	2.49
Yes	16.36	6.81	86	0.86	2.59
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

Table 10.9: Alpha Coefficients and Standard Errors of Measurement, Biology, Spring 2011

Group	Mean Raw Score	SD Raw Score	N Count	Reliability	SEM
All Students	23.07	6.21	62068	0.84	2.48
Gender					
Female	22.69	6.07	30986	0.83	2.51
Male	23.47	6.31	30955	0.85	2.46
Ethnicity					
American Indian/ Alaskan Native	22.68	5.92	313	0.82	2.53
Asian	24.66	6.65	1197	0.87	2.37
Pacific Islander	20.63	6.65	76	0.85	2.62
Black (not Hispanic)	18.81	6.22	9897	0.82	2.65
Hispanic	21.27	6.21	2275	0.83	2.58
White (not Hispanic)	24.02	5.77	47331	0.82	2.44
Multi-racial	24.08	6.02	696	0.84	2.43
LEP					
No	23.15	6.17	61062	0.84	2.48
Yes	17.91	6.56	1006	0.83	2.68
IEP					
No	23.45	6.01	57643	0.83	2.47
Yes	18.03	6.64	4425	0.84	2.66
Migrant					
No	23.07	6.21	62026	0.84	2.48
Yes	19.50	5.36	42	0.75	2.66
FRL					
No	24.46	5.74	38160	0.82	2.41
Yes	20.84	6.29	23908	0.83	2.59
Title I					
No	23.33	6.09	58153	0.84	2.47
Yes	19.08	6.62	3915	0.84	2.64
Accommodations					
No	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	N/A	N/A	N/A

10.4 Conditional Standard Error Estimates for Scale Scores

The overall *SEM* in Tables 10.1 to 10.9 represents the standard deviation of projected replications of the testing procedure averaged over all students. In contrast, conditional standard errors of measurement (*CSEMs*) are conditioned on the ability of the student. Rasch-based *CSEMs* ($CSEM(\theta)$) for each scale score are defined as the reciprocal of the square root of the test information function ($I(\theta)$) at the point on the ability continuum that corresponds to each scale score (Hambleton and Swaminathan, 1985):

$$CSEM(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

CSEMs are especially useful for characterizing measurement precision in the neighborhood of score levels used for decision making, such as cut scores at various achievement levels. The *CSEMs* for the Proficient cut scores for the MO EOC Assessments are presented in Table 10.10. *CSEMs* for other scale scores are reported in Chapter 7 of this technical report. Note that *CSEMs* are smaller in the middle of the score distribution than at the extremes. This pattern is expected for item response theory (IRT) based *CSEMs*. The value for all *CSEMs* was either 6 or 7 scale-score points. These values reflect a reasonable amount of measurement error at the Proficient cut for making adequate yearly progress (AYP) determinations for federal accountability.

Table 10.10: *CSEMs* at the Proficient Cut Score

Test Event	Subject	SS Cut*	CSEM
Summer 2010	English II	200	6
	Algebra I	200	8
	Biology	200	7
Fall 2010	English II	200	6
	Algebra I	200	7
	Biology	200	7
Spring 2011	English II	200	6
	Algebra I	200	7
	Biology	200	7

*See Tables 7.13 through 7.21 in Chapter 7 for the *CSEM* at each scale score.

10.5 Reliability of Classifications

The reliability of student achievement-level classifications (i.e., Below Basic, Basic, Proficient, and Advanced) was evaluated using a computer program developed by Huynh (1979). This FORTRAN program is based on the beta-binomial model that also provides SEs for the consistency estimates. Classification consistency refers to the degree to which each student's achievement level can be replicated and is similar to the traditional test-retest or equivalent forms reliability. Using the maximum possible score, mean, standard deviation, and KR-21 reliability estimate, the program computes parameters (α , β) for the beta-binomial distribution. Kappa indices, which estimate the level of improvement in

decision consistency beyond chance when test data are used, are then computed (Huynh, 1979).

Tables 10.11 and 10.12 show the results of the classification and decision consistency analyses for Summer 2010, Fall 2010, and Spring 2011 administrations. As noted above, the raw agreement index is a classification consistency index that estimates the percentage of examinees who would (hypothetically) be assigned to the same achievement level if the same test was administered a second time or an equivalent test was administered under the same conditions. The agreement consistency indices (p) for the EOC Assessments were generally in the mid 60s to lower 70s. These values reflect classification agreement consistency for the four performance categories: Below Basic, Basic, Proficient, and Advanced. Had a student been regarded as “pass” if his or her achievement level was Proficient or Advanced and as “fail” if his or her achievement level was Below Basic or Basic, the agreement consistency indices would have been 10 to 15 percent higher, as indicated in Tables 10.13 and 10.14. The latter classification accuracy is directly related to determining the accuracy of proficiency classifications for the No Child Left Behind Act (NCLB).

Table 10.11: Classification Consistency Coefficients

Year	N (Items)	Raw Cut Scores			Mean	SD	Kappa	SE (κ)
		Basic	Proficient	Advanced				
Summer 2010								
English II	35	12	21	30	18.60	6.37	0.47	0.0107
Algebra I	35	11	19	27	17.22	5.98	0.42	0.0078
Biology	35	14	22	30	18.80	6.12	0.44	0.0134
Fall 2010								
English II	35	12	21	29	22.18	6.80	0.50	0.0057
Algebra I	35	11	20	28	20.14	8.03	0.55	0.0044
Biology	35	13	22	29	21.78	7.18	0.51	0.0050
Spring 2011								
English II	35	12	21	30	24.63	6.42	0.52	0.0010
Algebra I	35	12	20	28	21.44	6.61	0.47	0.0010
Biology	35	13	22	30	23.07	6.21	0.47	0.0010

Table 10.12: Raw Agreement Consistency Coefficients

Year	N (Items)	Raw Cut Scores			Mean	SD	p	SE (p)
		Basic	Proficient	Advanced				
Summer 2010								
English II	35	12	21	30	18.60	6.37	0.66	0.0022
Algebra I	35	11	19	27	17.22	5.98	0.62	0.0017
Biology	35	14	22	30	18.80	6.12	0.63	0.0042
Fall 2010								
English II	35	12	21	29	22.18	6.80	0.66	0.0021
Algebra I	35	11	20	28	20.14	8.03	0.68	0.0023
Biology	35	13	22	29	21.78	7.18	0.65	0.0023
Spring 2011								
English II	35	12	21	30	24.63	6.42	0.69	0.0003
Algebra I	35	12	20	28	21.44	6.61	0.64	0.0004
Biology	35	13	22	30	23.07	6.21	0.65	0.0002

Table 10.13: Classification Consistency Coefficients (Two Classification Categories)

Year	N (Items)	Raw Cut Scores	Mean	SD	Kappa	SE (κ)
		Proficient/ Advanced				
Summer 2010						
English II	35	21	18.60	6.37	0.61	0.0140
Algebra I	35	19	17.22	5.98	0.58	0.0098
Biology	35	22	18.80	6.12	0.58	0.0170
Fall 2010						
English II	35	21	22.18	6.80	0.66	0.0063
Algebra I	35	20	20.14	8.03	0.72	0.0042
Biology	35	22	21.78	7.18	0.68	0.0051
Spring 2011						
English II	35	21	24.63	6.42	0.65	0.0012
Algebra I	35	20	21.44	6.61	0.64	0.0011
Biology	35	22	23.07	6.21	0.62	0.0012

Table 10.14: Raw Agreement Consistency Coefficients (Two Classification Categories)

Year	N (Items)	Raw Cut Scores	Mean	SD	p	SE (p)
		Proficient/ Advanced				
Summer 2010						
English II	35	21	18.60	6.37	0.81	0.0065
Algebra I	35	19	17.22	5.98	0.79	0.0047
Biology	35	22	18.80	6.12	0.81	0.0073
Fall 2010						
English II	35	21	22.18	6.80	0.84	0.0029
Algebra I	35	20	20.14	8.03	0.86	0.0021
Biology	35	22	21.78	7.18	0.84	0.0025
Spring 2011						
English II	35	21	24.63	6.42	0.87	0.0004
Algebra I	35	20	21.44	6.61	0.83	0.0005
Biology	35	22	23.07	6.21	0.82	0.0006

CHAPTER 11: VALIDITY

11.1 Introduction

According to the *Standards* (AERA, APA, and NCME, 1999), “Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing program. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees” (p. 17). While this chapter summarizes evidence that supports claims about the validity of Missouri End-of-Course (MO EOC) Assessment scores, many other parts of this technical report also provide appropriate evidence for validity. Some of this evidence is cross-referenced below for added convenience. The procedural and empirical evidence available, along with the rationale presented below, provides support for the standards-based interpretations of the MO EOC Assessments.

This chapter begins with a brief review of important federal statutes that require the MO EOC Assessments and explains the purposes and intended uses of test scores, suggesting the value implications of these assessments for schools, teachers, students, and parents. Validity evidence related to test content is presented in terms of the adequacy and appropriateness of the EOC Assessments for measuring progress on the Missouri content standards. Then, validity evidence based on the internal structure of the MO EOC Assessments is provided through a correlational analysis of MO EOC Assessment content clusters. References to specific standards are provided where appropriate.

11.2 Federal Authority for School Accountability

The United States Department of Education bases accountability on a school’s achievement of adequate yearly progress (AYP) in reading/language arts and mathematics. AYP determinations refer to the minimum improvement required of each school and district during the course of one year. For Missouri high schools and school districts, AYP is set in terms of the percentage of all students, and all student groups of sufficient size, scoring Proficient or above on the MO EOC tests in English II and Algebra I.

11.3 Purpose and Intended Uses of Test Scores

The *Standards* state that “Validation logically begins with an explicit statement of the proposed interpretation of the test scores, along with a rationale for the relevance of the interpretation to the proposed use” (AERA, APA, and NCME 1999).¹⁰ The MO EOC Assessments were developed for the following purposes and uses:

- Measuring and reflecting students’ mastery toward post-secondary readiness
- Identifying students’ strengths and weaknesses
- Communicating expectations for all students

¹⁰ **Standard 1.2:** The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intended to assess should be clearly described (p. 17).

- Serving as the basis for state and national accountability plans
- Evaluating programs

The valid interpretation and appropriate use of MO EOC Assessment scores are supported in a variety of ways, including the training and consultation provided by personnel of the Missouri Department of Elementary and Secondary Education (DESE) and publications such as the *Test Examiner's Manual*, *Guide to Interpreting Results*, and this technical report. The training and documentation provided to test users help them better administer, understand, and use test score results.

11.4 MO EOC Assessment Scores

The MO EOC Assessment scores are scaled in several ways: raw-score points, item response theory (IRT) derived scale scores, and achievement level (based on scale-score cuts). Missouri actively promotes the use of achievement-level results, reporting them annually on each assessment at the student, school, district, and state levels. Individual student and average scale scores are also used, but they play a secondary role and are generally interpreted with reference to their distance from achievement-level cut points. Test results are reported for students as a whole as well as by student group, including gender, ethnicity, migrant status, free and reduced lunch (FRL) status, English language proficiency, Title I, Individualized Education Program (IEP) status, and accommodations used during testing. Scores are reported to schools and districts in annually published reports (for more information, see Chapter 8: Reporting).

The MO EOC Assessment score indicates that an individual student performs at the Below Basic, Basic, Proficient, or Advanced level in a given content area. Achievement-level descriptors provide details about the content expectations that students at each level meet or exceed. No stakes for teachers are attached to student-level scores by the state. Teachers are counseled to interpret individual student scores only in the context of other assessment results and their own experience.

11.5 Content-Related Evidence of Validity

Baker and Linn (2002) suggest that “Two questions are central in the evaluation of content aspects of validity. Is the definition of the content domain to be assessed adequate and appropriate? Does the test provide an adequate representation of the content domain the test is intended to measure?” (p. 6). The following sections help answer these two very important questions and also address Standard 1.6¹¹ of the *Standards* (AERA, APA, NCME, 1999), which specifically relates to the definition and development of test content.

¹¹ **Standard 1.6:** When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified (p. 18).

11.5.1 Appropriateness of Content Definition

In 1993, the Missouri legislature passed the Outstanding Schools Act (Senate Bill 380), requiring the State Board of Education to adopt challenging academic performance standards that define the skills and competencies necessary for students to successfully advance through the public school system, prepare for post-secondary education and the workplace, and participate as citizens in a democratic society. The Missouri State Board of Education formally adopted the academic standards known as the Show-Me Standards in January 1996.

In addition to mandating the development of rigorous academic standards, the Outstanding Schools Act of 1993 required the development and implementation of a comprehensive, primarily performance-based assessment program to measure student proficiency in the knowledge, skills, and competencies identified in the standards. Upon adoption of the standards in 1996, Missouri began developing the Missouri Assessment Program (MAP).

In January 2007, the Missouri State Board of Education approved a plan to replace the MAP for high school students, beginning in August of the 2008–2009 school year, with EOC Assessments in English II, Algebra I, and Biology. The intent was to provide MO EOC Assessments that are an integral part of the statewide assessment system and, as such, are a logical extension of MAP tests at the elementary and middle grade levels.

11.5.2 Adequacy of Content Representation

Adequacy of the content representation of the MO EOC Assessments is critically important because the tests must provide an indication of student progress toward achieving the knowledge and skills identified in the Missouri Course-Level Expectations (CLEs), and they must fulfill the requirements of the No Child Left Behind Act.

Adequate representation of the content domains defined in the CLEs is assured through the use of a test blueprint and a carefully documented test construction process. CLEs and the Show-Me Standards are taken into consideration in the writing of selected response. Each assessment must align with and proportionally represent the subdomains of the test blueprint. Evidence to support the content validity of the MO EOC Assessments was provided in Chapter 2: Test Development through the documentation of the test specifications and blueprints, item-writing processes, and item-review processes.

Additional evidence to support the content validity of the MO EOC Assessments was provided in Tables 2.1 through 2.3 in Chapter 2: Test Development and also in Chapter 4: Item Analysis. Chapter 2 outlined the target strand and CLE point distributions on the English II, Algebra I, and Biology operational forms. All forms administered in 2009–2010 met the point ranges specified in the blueprints. In addition, Riverside Publishing strove to equitably represent the strands on each assessment by balancing CLE and sub-CLE coverage according to the targets outlined in the test specifications and by matching item format to the requirements of the content and standards descriptions.

11.6 Validity Evidence Based on the Internal Structure of the MO EOC Assessments

Standard 1.11¹² pertains to the relationships between the parts of the test. One way to study patterns of relationships to provide evidence supporting the inferences made from test scores is the multitrait, multimethod matrix. Tables 11.1 through 11.3 summarize Pearson correlation coefficients among test domains and clusters for English II, Algebra I, and Biology. The correlations between clusters within each assessment are in the moderate to moderately high range, suggesting strong relationships between the clusters. Note that the high correlations between cluster scores and total assessment scores are inflated due to the overlap of items.

Table 11.1: Pearson Correlation Coefficients Between Domains and Clusters for English II

		Reading	Writing
Fall 2008	English II	0.97	0.80
	Reading		0.63
	Writing		
Spring 2009	English II	0.98	0.73
	Reading		0.57
	Writing		
Summer 2009	English II	0.98	0.65
	Reading		0.47
	Writing		
Fall 2009	English II	0.98	0.76
	Reading		0.60
	Writing		
Spring 2010	English II	0.98	0.75
	Reading		0.60
	Writing		
Summer 2010	English II	0.98	0.55
	Reading		0.39
	Writing		
Fall 2010	English II	0.99	0.66
	Reading		0.53
	Writing		
Spring 2011	English II	0.99	0.66
	Reading		0.53
	Writing		

¹² **Standard 1.11:** If the rationale for a test use or interpretation depends on premises about the relationships among parts of the test, evidence concerning the internal structure of the test should be provided (p. 20).

Table 11.2: Pearson Correlation Coefficients Between Domains and Clusters for Algebra I

		Number and Operations	Algebraic Relationships	Data and Probability
Fall 2008	Algebra I	0.81	0.95	0.80
	Number and Operations		0.66	0.56
	Algebraic Relationships			0.66
	Data and Probability			
Spring 2009	Algebra I	0.82	0.95	0.79
	Number and Operations		0.66	0.58
	Algebraic Relationships			0.64
	Data and Probability			
Summer 2009	Algebra I	0.79	0.95	0.73
	Number and Operations		0.64	0.45
	Algebraic Relationships			0.58
	Data and Probability			
Fall 2009	Algebra I	0.82	0.94	0.80
	Number and Operations		0.66	0.58
	Algebraic Relationships			0.63
	Data and Probability			
Spring 2010	Algebra I	0.79	0.96	0.76
	Number and Operations		0.64	0.52
	Algebraic Relationships			0.62
	Data and Probability			
Summer 2010	Algebra I	0.84	0.85	0.85
	Number and Operations		0.54	0.64
	Algebraic Relationships			0.56
	Data and Probability			
Fall 2010	Algebra I	0.88	0.96	0.85
	Number and Operations		0.76	0.69
	Algebraic Relationships			0.73
	Data and Probability			
Spring 2011	Algebra I	0.85	0.95	0.73
	Number and Operations		0.69	0.54
	Algebraic Relationships			0.57
	Data and Probability			

Table 11.3: Pearson Correlation Coefficients Between Domains and Clusters for Biology

		Characteristics and Interactions	Changes in Ecosystems	Inquiry
Fall 2008	Biology	0.91	0.83	0.87
	Characteristics and Interactions		0.72	0.63
	Changes in Ecosystems			0.57
	Inquiry			
Spring 2009	Biology	0.90	0.83	0.86
	Characteristics and Interactions		0.67	0.61
	Changes in Ecosystems			0.58
	Inquiry			
Summer 2009	Biology	0.86	0.85	0.89
	Characteristics and Interactions		0.65	0.59
	Changes in Ecosystems			0.64
	Inquiry			
Fall 2009	Biology	0.91	0.86	0.91
	Characteristics and Interactions		0.72	0.70
	Changes in Ecosystems			0.69
	Inquiry			
Spring 2010	Biology	0.89	0.83	0.90
	Characteristics and Interactions		0.65	0.66
	Changes in Ecosystems			0.63
	Inquiry			
Summer 2010	Biology	0.93	0.86	
	Characteristics and Interactions		0.62	
	Changes in Ecosystems			
	Inquiry			
Fall 2010	Biology	0.96	0.89	
	Characteristics and Interactions		0.74	
	Changes in Ecosystems			
	Inquiry			
Spring 2011	Biology	0.95	0.87	
	Characteristics and Interactions		0.68	
	Changes in Ecosystems			
	Inquiry			

11.7 Discriminant Validity Evidence for the MO EOC Assessments

The Standards for Educational and Psychological Testing (1999) states the following regarding convergent and divergent validity: “Relationships between test scores and other measures intended to assess similar constructs provide convergent evidence, whereas relationships between test scores and measures purportedly of different constructs provide discriminant evidence.” (pg. 14). The Missouri End-of Course assessments were designed to measure different constructs as shown by both the standards they assess and the content coverage detailed in the test blueprints.

The data presented below show evidence of divergent validity for Phase I content areas using scale scores. The data sets used for the analysis were drawn from the Spring 2011 operational test administration. The students in the data sets were merged using Missouri’s unique student identification number. Any student who took at least two of the three operational tests was included in the correlations. Table 11.4 shows the Pearson correlation coefficients between scale scores and the N count for each correlation.

The results shown in Table 11.4 contain evidence of divergent validity. Evidence of divergent validity is supported by the lower correlations between content areas that measure dissimilar constructs. For example, the correlation between English II and Algebra I (0.575), which is moderately low.

For English II and Biology, challenging language and reading on both tests could account for the higher correlation value (0.685). This correlation value is still lower than the tests measuring a similar construct and are in the range of the correlations among high school MAP subject area tests (the precursor to MAP End-of-Course assessments) as reported in the *Missouri Assessment Program Technical Report, 2008* (Missouri Department of Elementary and Secondary Education, 2008).

Table 11.4 Pearson Correlation Between Phase I Assessments

	English II	Algebra I	Biology
English II	--	0.575 (N=11059)	0.685 (N=33060)
Algebra I		--	0.640 (N=15476)
Biology			--

Table 11.5 provides more evidence of discriminant validity with correlations between Phase I and Phase II subjects. Evidence of discriminant validity is seen when comparing correlations between the similar contents of Algebra I and Geometry (0.709), Algebra I and Algebra II (0.770), English I and English II (0.681) and the dissimilar contents of Algebra I and English I (0.532), Algebra II and English II (0.535) and Geometry and English II (0.536).

Table 11.5 Pearson Correlation between Phase I and Phase II Assessments

	English I	Algebra II	Geometry	Government	History
English II	0.681 (N=420)	0.535 (N=8154)	0.536 (N=13562)	0.725 (N=8212)	0.656 (N=7270)
Algebra I	0.532 (N=20349)	0.770 (N=152)	0.709 (N=379)	0.618 (N=8558)	0.512 (N=9406)
Biology	0.676 (N=7137)	0.605 (N=8105)	0.588 (N=11086)	0.743 (N=7890)	0.675 (N=8640)

11.8 Additional Validity Evidence for the MO EOC Assessments

Validity evidence related to other standards is described below.

Standard 1.5¹³ relates to the characteristics of the sample of examinees from which validity evidence is inferred. The sample of examinees from which the validity evidence for the MO EOC Assessments was obtained is described in detail in Chapter 9: Summary Statistics, which includes tables with descriptive statistics for raw score, scale score, and achievement-level distributions. Statistics include *n*-counts, means, standard deviations, minimum and maximum values, and a variety of data disaggregations.

Standard 1.7¹⁴ relates to human judgment at various points in the test development and reporting process. For the MO EOC Assessments, human judgment was especially prevalent during the standard-setting process. Chapter 3: Achievement-Level Setting contains detailed information about the standard-setting procedures used for the MO EOC Assessments, including the selection process for and characteristics of the standard-setting participants. From Spring 2008 through Spring 2010, PE and WP items were hand scored. Chapter 6: Scanning, Scoring, and Quality Control Procedures contains detailed information about the processes involved with Assessment Resource Center’s hand scoring of the PE/WP items, including scorer selection, training, qualifications, and quality-control measures.

Standard 1.13¹⁵ relates to the conditions under which the data used to support validity claims were collected. Chapter 5: Test Administration contains information about how

¹³ **Standard 1.5:** The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics (p. 18).

¹⁴ **Standard 1.7:** When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth (p. 19).

¹⁵ **Standard 1.13:** When validity evidence includes statistical analyses of test results, either alone or together with data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions.

data were gathered in both the online and paper-and-pencil administrations, including the testing environment, materials distribution and security, Test Examiner training, student preparation, and allowable accommodations.

11.9 Summary

Validity is not an all-or-nothing property of a test; rather, validity evidence must be documented for a specific purpose and in the context of how the test scores will be interpreted and used. Much of the information contained in this technical report is, in and of itself, documentation of the validity of the MO EOC Assessments for their stated purpose. This chapter provides a summary of the evidence presented elsewhere in the manual and provides some additional types of validity evidence relevant to the content and internal structure of the assessments.

The overall technical quality of the EOC Assessments was sound. The Spring 2008 standalone field test, the Spring 2009 embedded field test, and the Spring 2010 embedded field test produced pools of technically sound items, with a 91% retention rate after psychometric and content criteria were applied. From those pools, Riverside Publishing was able to assemble forms that were psychometrically very similar, and that similarity helped support the pre-equating model that is in place. Application of item response theory (IRT) pre-equating resulted in perfect or nearly congruent raw-to-scale score conversions between the Summer, Fall, and Spring forms at the proficiency level cuts.

Post-administration test analyses supported the technical quality of the MO EOC Assessments. Evaluations of IRT model assumptions supported the use of the Rasch model for all tests. Test reliabilities ranged from 0.81 to 0.91 across the content areas for the 2010–2011 test forms. Conditional standard errors of measurement (*CSEMs*) were between 6 and 8 scale score points at the cut scores. The item analyses also showed that the MO EOC Assessments have sound psychometric properties. The *p*-value ranges were sufficiently broad, indicating that the items do measure achievement across a broad range of difficulty. Most of the items had discrimination values $> .15$, and only three items had a value $< .10$. Speededness was not a factor in students' test performance. Item bias analyses conducted on the pools further indicated that items were functioning equivalently for different gender and ethnic groups.

Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test performance (p. 20).

REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). 1999. *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Baker, E. L., and R. L. Linn. 2002. *Validity issues for accountability systems*. Technical report 585. Los Angeles: Center for the Study of Evaluation.
- Brown, W. 1910. Some experimental results in the correlation of mental abilities. *British Journal of Psychology* 3:296–322.
- Council of Chief State School Officers. 2003. *Quality control checklist for processing, scoring, and reporting*.
http://www.ccsso.org/Documents/2003/Quality_Control_Checklist_Scoring_2003.pdf (accessed November 5, 2009).
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.
- Dorans, N. J., and P. W. Holland. 1993. DIF detection and description: Mantel-Haenszel and standardization. In *Differential item functioning*, ed. P. W. Holland and H. Wainer, 35–66. Hillsdale, NJ: Erlbaum.
- Feldt, L. S., and R. L. Brennan. 1989. Reliability. In *Educational measurement*, 3rd ed., ed. R. L. Linn, 105–146. New York: Macmillan.
- Hambleton, R. K., and H. Swaminathan. 1985. *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff.
- Huynh, H. 1979. Computational and statistical inference for two reliability indices based on the beta-binomial model. *Journal of Educational Statistics* 4:231–246.
- Kolen, M. J., and R. L. Brennan. 2004. *Test equating, scaling, and linking: Methods and practices*, 2d ed. New York: Springer-Verlag.
- Kuder, G. F., and M. W. Richardson. 1937. The theory of the estimation of test reliability. *Psychometrika* 2:151–160.
- Linacre, J. M. 2006a. *A user's guide to WINSTEPS Rasch-model computer programs*. Chicago: Winsteps.
- Linacre, J. M. 2006b. WINSTEPS v3.64.
- Linacre, J. M., and B. D. Wright. 1989. The equivalence of Rasch, PROX, and Mantel-Haenszel. *Transactions of the Rasch Measurement Special Interest Group* 3:1–3.
- Liu, Y., and D. Mix. 2006. Categorizations of WINSTEPS DIF output. Paper presented at the American Educational Research Association conference, San Francisco, California.
- Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute* 22:719–748.

- Masters, G. N. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47:149–174.
- Miller, G. E., O. Rotou, and J. S. Twing. 2004. Evaluation of the 0.3 logits screening criterion in common item equating. *Journal of Applied Measurement* 5(2):172–177.
- Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Schulz, E. M., C. Perlman, W. K. Rice, and B. D. Wright. 1996. An empirical comparison of Rasch and Mantel-Haenszel procedures for assessing differential item functioning. In *Objective measurement: Theory into practice*, Vol. 3, ed. G. Engelhard and M. Wilson, 65–84. Norwood, NJ: Ablex Publishing.
- Smith R. M., and C. Y. Miao. 1994. Assessing unidimensionality for Rasch measurement. In *Objective measurement: Theory into practice*, Vol. 2, ed. M. Wilson, 316–328. Norwood, NJ: Ablex Publishing.
- Thompson, S. J., C. J. Johnstone, and M.L. Thurlow. 2002. *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
<http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html> (accessed December 21, 2009).
- Traub, R. E. 1994. *Reliability for the social sciences*. London: Sage Publications.
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wright, B. D., and M. H. Stone. 1979. *Best test design*. Chicago: MESA Press.

APPENDIX A: DEMOGRAPHIC CHARACTERISTICS OF STANDARD-SETTING PARTICIPANTS

Appendix Table A.1: English II

Region	District	Gender	Position	Ethnicity	% Minority*	% Free and Reduced Lunch*
Heart of Missouri	Jefferson City	F	Classroom Teacher	White	43%	25%
Heart of Missouri	Lincoln University	M	Non-Teacher Educator	White	NA	NA
Heart of Missouri	Not Available	M	Non-School	White	NA	NA
Kansas City	Blue Springs	F	Classroom Teacher	White	21%	19%
Kansas City	Lee's Summit	F	Classroom Teacher	White	13%	19%
Kansas City	North Kansas City	F	Classroom Teacher	White	41%	30%
Kansas City	Park Hill	M	Non-Teacher Educator	White	22%	22%
Northwest	St. Joseph	F	Classroom Teacher	White	56%	17%
Southeast	Jackson	F	Classroom Teacher	White	29%	4%
Southwest	Neosho	F	Classroom Teacher	White	56%	17%
St. Louis	Afton	F	Classroom Teacher	White	32%	16%
St. Louis	Rockwood	F	Classroom Teacher	White	13%	17%
West Central	Raymore-Peculiar	F	Classroom Teacher	White	21%	14%

*Percent minority and percent free and reduced lunch refers to the population of the district represented by the panelist. NA = Not available.

Appendix Table A.2: Algebra I

Region	District	Gender	Position	Ethnicity	% Minority*	% Free and Reduced Lunch*
Heart of Missouri	Jefferson City	M	Classroom Teacher	Asian/PI	43%	25%
Heart of Missouri	Keytesville	M	Non-Teacher Educator	White	46%	1%
Heart of Missouri	Moberly	F	Classroom Teacher	White	56%	13%
Kansas City	Center 58	F	Non-Teacher Educator	White	67%	86%
Kansas City	Kearney	M	Non-Teacher Educator	White	11%	4%
Kansas City	Lee's Summit	M	Classroom Teacher	White	13%	19%
Northwest	Hamilton	F	Classroom Teacher	NA	42%	4%
South Central	Saint Clair	F	Classroom Teacher	White	45%	3%
Southeast	North St. Francis County	F	Classroom Teacher	White	51%	2%
Southwest	Neosho	M	Classroom Teacher	White	56%	17%
Southwest	Nixa	F	Classroom Teacher	White	32%	7%
Southwest	Springfield	F	Non-Teacher Educator	White	45%	14%
St. Louis	Northwest	F	Classroom Teacher	White	34%	2%
St. Louis	Rockwood	F	Classroom Teacher	NA	13%	17%
West Central	Sherwood Cass	M	Classroom Teacher	White	47%	3%

*Percent minority and percent free and reduced lunch refers to the population of the district represented by the panelist. NA = Not available.

Appendix Table A.3: Biology

Region	District	Gender	Position	Ethnicity	% Minority*	% Free and Reduced Lunch*
Heart of Missouri	Fayette	F	Classroom Teacher	White	44%	13%
Heart of Missouri	Lincoln University	M	Non-Teacher Educator	White	NA	NA
Kansas City	Independence	M	Non-Teacher Educator	White	55%	25%
Kansas City	Kansas City	M	Classroom Teacher	Black	81%	85%
Northeast	North Shelby	F	Classroom Teacher	White	37%	0%
Northwest	Maryville	F	Classroom Teacher	White	29%	6%
Northwest	St. Joseph	M	Classroom Teacher	White	56%	17%
South Central	Maries County	M	Classroom Teacher	White	42%	2%
South Central	Waynesville	F	Classroom Teacher	Black	39%	39%
Southeast	Jackson	M	Classroom Teacher	White	29%	4%
Southwest	Branson	M	Classroom Teacher	White	48%	13%
Southwest	Carl Junction	F	Classroom Teacher	White	35%	6%
Southwest	Mansfield	F	Classroom Teacher	White	60%	6%
St. Louis	Clayton	M	Classroom Teacher	White	16%	24%
St. Louis	Ferguson-Florissant	M	Classroom Teacher	Asian/PI	60%	81%

*Percent minority and percent free and reduced lunch refers to the population of the district represented by the panelist. NA = Not available.

APPENDIX B: STANDARD-SETTING SESSION AGENDA

Missouri EOC Achievement-Level Setting Agenda Capitol Plaza Hotel and Convention Center Jefferson City, Missouri—November 3–5, 2008

(NOTE: Times are approximate.)

Monday, November 3

Morning

7:30–8:30	Registration and Breakfast
8:30–9:15	Welcome, Introductions, Logistics, and Overview of Missouri’s EOC Assessments (DESE)
9:15–9:35	Overview of the Standard-Setting Sessions (Questar Assessment)
9:35–10:00	Introduction to Achievement-Level Descriptors (ALDs) (Questar Assessment)
10:00–10:15	Break
10:15–11:15	Setting Performance Standards—General Process
11:15–12:15	“Experience” the Assessments
12:15–1:30	Lunch

Afternoon

1:30–3:15	Definitions and Description of Performance Standards
3:15–3:30	Break
3:30–4:30	Orientation to the Specific Standard-Setting Methodology
4:30–4:45	Questions and Dismissal for the Day

Tuesday, November 4

Morning

7:30–8:30	Breakfast
8:30–9:15	Review of Day 1 Activities and Discussions
9:15–10:15	Preparation for Round 1 of Judgments
10:15–10:30	Break
10:30–12:00	Round 1 Judgments
12:00–1:15	Lunch

Afternoon

1:15–1:45	Review of Round 1 Issues and Problems
1:45–3:15	Feedback and Discussion of Round 1 Judgments
3:15–3:30	Break
3:30–3:45	Preparation for Round 2 Judgments
3:45–5:00	Round 2 Judgments

Wednesday, November 5

Morning

7:45–8:45	Breakfast
8:45–9:45	Review of Round 2 Judgments
9:45–10:00	Break
10:00–10:45	Preparation for Final Judgments
10:45–12:30	Final Round of Judgments and Evaluation
12:30–1:15	Lunch

Afternoon

1:15–2:15	Final review of ALDs and Session Wrap-up
-----------	--

APPENDIX C: OPENING SESSION POWERPOINT PRESENTATION

**Standard Setting
Overview**

Missouri
Algebra I, Biology & English II
EOC Assessments

November, 2008

Setting Performance Standards

- *Who's Involved?* State and contractor roles
- *Why Questar?* Who's facilitating? Our role
- *Why you?* Individually & collectively:

You are the *experts*.
You represent various audiences.

Session Outline - Day 1

- I. What is "standard setting" - in general and for the EOC Assessments?
- II. Describe the performance "categories"; refine achievement level descriptors (ALDs)
- III. Review & discuss the actual EOC test;
- IV. The "Angoff procedure" – how it works

Setting Performance Standards

- *Who's Involved?* State and contractor roles
- *Why Questar?* Who's facilitating? Our role
- *Why you?* Individually & collectively:

You are the *experts*.
You represent various audiences.
You are *judges*, not psychometricians.
You are *advisors*, not policy makers

Setting Performance Standards

- *Who's Involved?* State and contractor roles
- *Why Questar?* Who's facilitating? Our role:
Not content experts, but facilitators

Groundrules

NO DISCUSSIONS about the *EOC* program or its underlying content standards

OR

Groundrules

NO DISCUSSIONS about the EOC program
OR

- why to set standards
- the philosophy of educational assessment
- why these particular tasks/assessments
- why a particular procedure is being used

What IS Standard Setting?

- another frame of reference to interpret test scores (“how good is *good*”?)
- a routine, daily activity

Groundrules

NO DISCUSSIONS about the EOC program
OR

- why to set standards
- the philosophy of educational assessment
- why these particular tasks/assessments
- why a particular procedure is being used

Confidentiality of all materials & discussions.

What IS Standard Setting?

- another frame of reference to interpret test scores (“how good is *good*”?)
- a routine, daily activity
- true “criterion-referencing”

Groundrules

NO DISCUSSIONS about the EOC program
OR

- why to set standards
- the philosophy of educational assessment
- why these particular tasks/assessments
- the fairness of assessing special students
- why a particular procedure is being used

Confidentiality of all materials & discussions.

All discussions should be *as a group*.

What IS Standard Setting?

- another frame of reference to interpret test scores (“how good is *good*”?)
- a routine, daily activity
- true “criterion-referencing”
- a *semi*-quantitative, *semi*-standardized, socio-political judgment process

What IS Standard Setting?

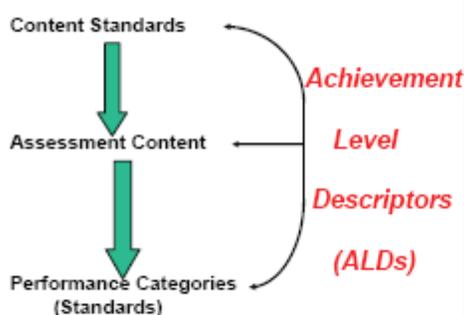
- just a frame of reference for test scores
- a routine, daily activity
- true “criterion-referencing”
- essentially, a judgment process

• **NOT** “science” !

Advice on Setting Standards

- Set demanding, but *attainable* standards

Critical Elements of An Assessment System



Advice on Setting Standards

- Set demanding, but *attainable* standards
- What “*should be*” probably shouldn’t disregard what “*is*”

4 Keys to Being a Great Judge:

1. **Judgments** vs. Data
2. “**Should**” vs. “Will”
3. Consider **ALL Missouri** students who took this EOC assessment
4. Think of **threshold** students, not *all* who are Proficient

Advice on Setting Standards

- Set demanding, but *attainable* standards
- What “*should be*” probably shouldn’t disregard what “*is*”
- Focus on **concrete** behaviors, skills, responses

Advice on Setting Standards

- Set demanding, but *attainable* standards
- What "*should be*" probably shouldn't disregard what "*is*"
- Focus on *concrete* behaviors, skills, responses
- (*for M-C items*) Item difficulty resides in the answer choices, not the item "stem"

Missouri EOC Achievement-Level Setting

English II
End-of-Course Assessment
November 3—5, 2008

Advice on Setting Standards

- Set demanding, but *attainable* standards
- Don't disregard what "*is*"
- Focus on the *concrete*
- (*for M-C items*) Item difficulty resides in the answer choices, not the item "stem"
- (*for constructed-response items*) Judge the *response quality*, not the task difficulty.

"Housekeeping"

- Security Forms
- Judges' Numbers
- Break and lunch locations
- General agenda for the day

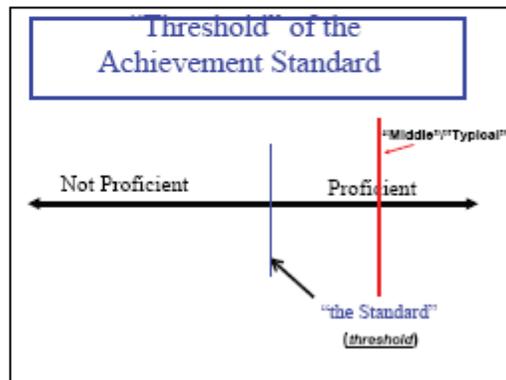
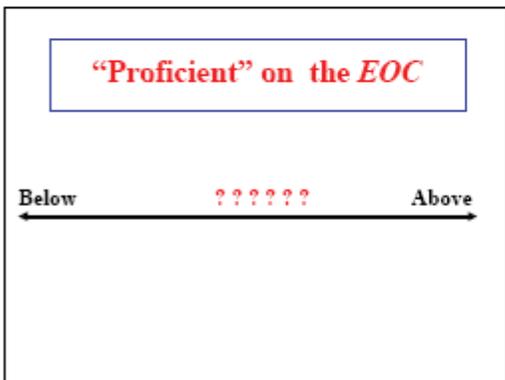
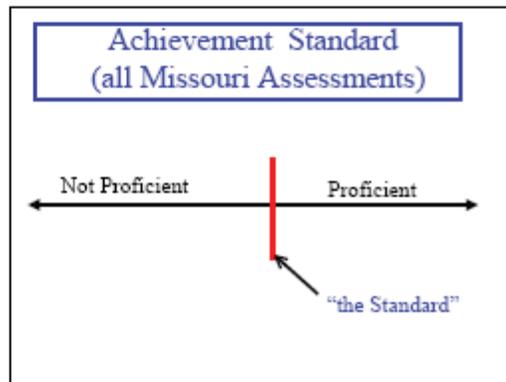
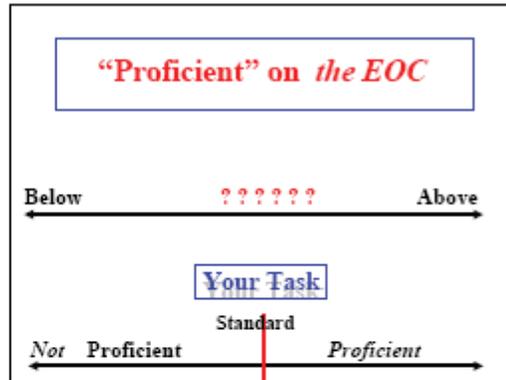
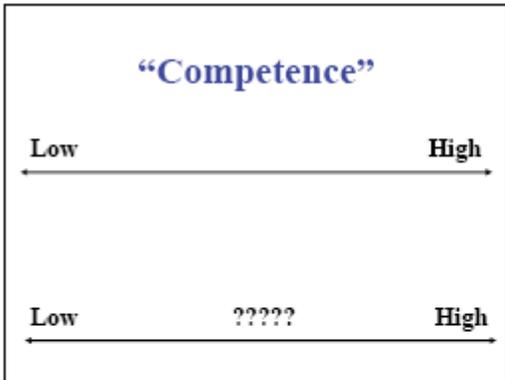
Advice on Setting Standards

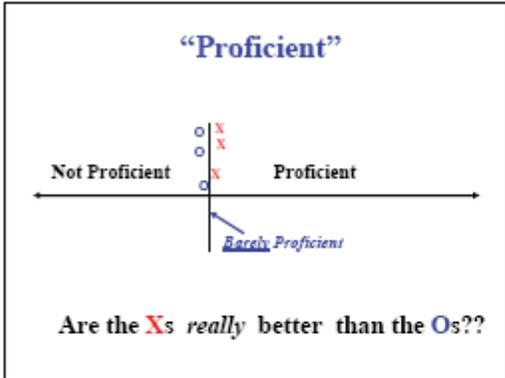
- Set demanding, but *attainable* standards
- What "*should be*" shouldn't disregard what "*is*"
- Focus on the *concrete*
- Remember the type of item you're judging

Use your best judgment !!

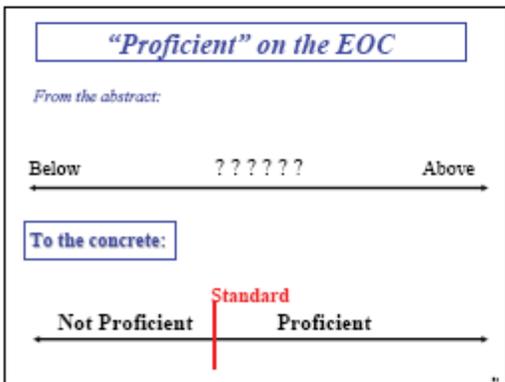
Session Outline - Day 1

- I. What is "standard setting" - in general, and for the EOC Assessments?
- II. Describe the performance "categories"; refine Achievement Level Descriptors (ALDs)
- III. Review & discuss the actual EOC test
- IV. The "Angoff procedure" – how it works



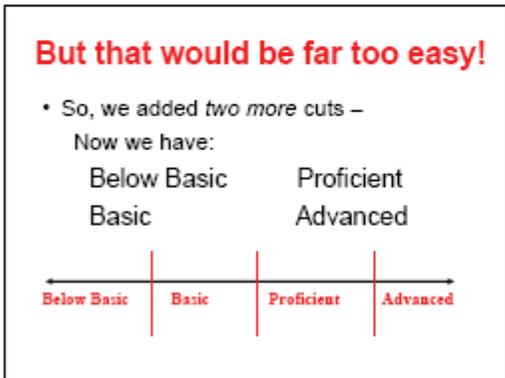


- ### Achievement Level Descriptors ALDs
- Start with the “labels”:
1. **Below Basic**
 2. **Basic**
 3. **Proficient**
 4. **Advanced**



Problem:

What do these *general*
descriptions of achievement levels
mean concretely for EOC
students in each content area?



- ### Key Elements of the ALDs
- **Advanced**
 - Demonstrate thorough understanding
 - Demonstrate higher-level skills
 - Consistently apply a variety of strategies
 - **Proficient**
 - Demonstrate understanding of skills and processes
 - Use a range of strategies
 - **Basic**
 - Demonstrate incomplete understanding
 - Demonstrate skills inconsistently
 - Use some strategies
 - **Below Basic**
 - Demonstrate little understanding
 - Demonstrate skills inconsistently or incorrectly
 - Use few strategies

Don't Forget the Assessment !

Why? Standards are set on the actual EOC assessments, not in general

What to do? "Be" a student
Think about each item / task

Think about: Skill(s) / behaviors / expectations tapped
Basic, Proficient, Advanced
"Threshold" students

ASK: How well SHOULD a student who is **JUST barely Proficient** be able to do this?

Achievement Level Descriptors ALDs – Your Task

- Action verbs, e.g., determine, analyze, evaluate, utilize, identify, compare, describe, etc.
- Qualifiers, e.g., adjectives and adverbs that describe:
 - Differences in amount (most, some, few, etc.)
 - Degree of
 - understanding (thorough, partial, etc.)
 - frequency (consistently, rarely, etc.)
 - effectiveness (highly, moderately, somewhat)

Now that you've seen the "tasks"...
let's debrief.

- Return to the general descriptors.
- Think about the tasks and items on the assessment.
- Which activities seemed to be hard (Advanced?) and easy (Basic?)?
These are the grounding of your work to establish standards. The descriptions "define" the categories and should anchor your judgments.



Key Elements of the ALDs

- **Advanced**
 - Demonstrate thorough understanding
 - Demonstrate higher-level skills
 - Consistently apply a variety of strategies
- **Proficient**
 - Demonstrate understanding of skills and processes
 - Use a range of strategies
- **Basic**
 - Demonstrate incomplete understanding
 - Demonstrate skills inconsistently
 - Use some strategies
- **Below Basic**
 - Demonstrate little understanding
 - Demonstrate skills inconsistently or incorrectly
 - Use few strategies

Achievement Level Descriptors ALDs – Your Task

- ALDs probably should be *broader* than any specific assessment.
- ALDs should be *descriptive*, not definitional.
- ALDs "anchor" the standards, as they describe the behaviors of students whose performances "fit" each category

Describe the assessed students *concretely*

Beginning with the *Proficient* category, describe the assessed students *concretely*.

What do they know? What can they do?
What skills do they possess in order to demonstrate this behavior?
What does the skill look like?
What are examples?
What behaviors/actions "fit" a certain category?

"Angoff" – What to Do ?

- Read each (MC) item in the test. Think about what is assessed/required.
- Conceptualize 100 "just barely" *Proficient* students all across the state who took this EOC.
- For each item, decide what percent of "barely *Proficient*" students should answer correctly.

"Angoff Procedure" for Setting Performance Standards

- A way, not *the way* to establish performance standards
- Recommended by the state's TAC
- Preferred procedure when statewide data are not available
- Requires judgments about each item on the assessment

"Angoff" – What to Do ?

- For each item, decide what percent of "barely *Proficient*" students should answer correctly.
- Repeat the decision for "barely *Advanced*" and "barely *Basic*."
- After making the 3 judgments about an item, move to the next item.

"Angoff Procedure" for Setting Performance Standards

- For each test item, simply judge the percent of students in each performance category who should answer correctly.
- You can expect **NO** students to answer correctly, all students, or somewhere in between.
- In general – *maybe without exception?* – you should expect *Basic* students to perform less well than *Proficient* students, and less well yet than *Advanced* students.

"Angoff" – What to Do ?

- Read each (MC) item in the test. Think about what is assessed/required.
- Conceptualize 100 "just barely" *Proficient* students all across the state who took this EOC
- For each item, decide what percent of the "barely *Proficient*" students should answer correctly.
- Repeat the decision for "barely *Basic*" & "barely *Advanced*." Move to the next item.
- The 100 kids in each group aren't identical in skill/background/instruction and don't all know the same things, so the decision can't be all-or-none.

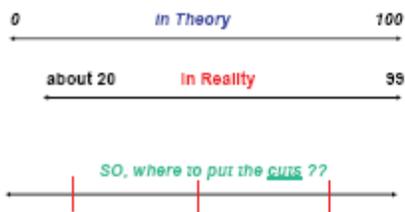
“Item Difficulty” Values

- The values you’re going to work with are often termed “p-values.”
- They’re the proportion of students answering the item correctly.
- Let’s look at what these values “mean.”

Judgments for the Constructed-Response Item

- The *mechanics* differ, but the *intent* is the same – how should *borderline* students do?
- Think of the same 100 “threshold” students at each performance level.
- Decide what their **AVERAGE SCORE** should be on the constructed-response item rubric.

“Item Difficulty” Values



What About the Constructed-Response Item?

- Think of the same 100 “threshold” students at each performance level.
- Decide what their average score should be on the writing prompt (0-4).
- For example, perhaps you expect the following averages:

<u>Basic</u>	<u>Proficient</u>	<u>Advanced</u>
1.0	2.5	3.5

Judgments for the Constructed-Response Item

- The English II EOC assessment includes one constructed-response (CR) item – a writing prompt.
- Making your judgments for this item differs from the process used for the MC items.
- How to “judge” this item:

What About the Constructed-Response Item?

- Think of the same 100 “threshold” students at each performance level. Decide what their average score should be on the constructed-response item.
- E.g. perhaps you expect the following averages:

<u>Basic</u>	<u>Proficient</u>	<u>Advanced</u>
1.0	2.5	3.5

- Enter these averages on your Rating Sheet.

Session Overview - Day 2

- I. Review Day 1 activities and outcomes.
- II. Practice the Angoff procedure.
- III. Round 1 of independent judgments
- IV. Feedback & discussion of Round 1
- V. Round 2 ratings -- *reconsider* Round 1

ISSUES:

Should / Ought
What *just* separates “Below Basic”
from “Basic”?
Basic from “Proficient”?
Proficient from “Advanced”?
Threshold Students
All Assessed EOC Students in Missouri

Practice Activity: What to Do?

Think about:

The *item* – what’s measured, intentionally or not?

The *curriculum* – Is this taught? Will it be?

The *performance category* – what does it mean?

Threshold students

How students *should* perform. What % should answer this item correctly

Record *three* judgments about each item.

Jot down any notes, questions, reactions as you work.

“Rules” for Ratings

- Anonymity
- Independence
- Don’t persevere -- Make a best guess.
- Find the “neighborhoods”; then refine.

Reminders for Round 1

- In a group of 100 students, all of whom are *just barely* Proficient, how many *should* answer this item correctly?
- Don’t think of a “clearly Proficient” student. Focus on the cut score, and someone who *barely* makes it into the category. *Above* the cut, but *just* above – at the threshold.
- Remember the ALDs – they’re your anchors.

Marking Your Judgments

- For MC items –
 - What percent of *barely* Proficient students should answer this item correctly?
- For CR items –
 - What should be the average score of *barely* Proficient students on this scale?
- Then, ask the same questions of *barely* Advanced and *barely* Basic.

Marking Your Judgments

- Record your Judge Number.
- For *MC Items*, bubble in 3 numbers per row/item.
Just Proficient, just Advanced, just Basic.
- For *CR Item*, remember - judge average performance by students in each category.
- If you change your mind, erase completely.
- Double check your form before turning it in.

Are these data helpful?

- Sorta. What's the issue??
- Small samples. *More importantly*, from a field test. However, it's all we have.
 - o Data tell how students *DID* perform.
 - o Data **CANNOT** tell how students **SHOULD** perform, **NOR** how those who demonstrate a particular level of competence perform.

ISSUES:

Should / Ought
What *just* separates "Below Basic"
from "Basic"?
Basic from "Proficient"?
Proficient from "Advanced"?
Threshold Students
All Assessed EOC Students in Missouri

Why Reratings?

- You are now a *different* judge.
 - Consider the judgments & views of your peers.
 - Goal: NOT "consensus," but *reflection*
- YOU ARE NOW a better judge,**
because you are a *better-informed* judge.

Discuss the Preliminary Ratings

- **WHY????**
- Hearing from your peers helps you to:
 - become more comfortable with your judgments -- both the *how* and *where*.
 - **reconsider** your earlier judgments.

Reratings: What to Do?

1. Reflect on earlier ratings – yours & peers.
2. Reflect on the discussions we have had.
3. Consider expanding the "zones" around your earlier judgments.
4. Reconsider each judgment –how well *should* a barely Basic/Proficient/Advanced student *do* on this item?

Session Overview -- Day 3

- I. Round 2 feedback & discussion
- II. *Real* ratings & session evaluation
- III. Final review of the ALDs

Thanks for your all
your hard work, your
patience, your many
contributions, and
your generous gift of
time!

“How do I know if I’m right?”

- *There is no “right.”*
- Did you keep in mind:
 1. “*Should*”?
 2. The *threshold* student?
 3. What “*Below Basic*,” “*Basic*,”
“*Proficient*” & “*Advanced*” mean?
 4. *All* assessed students for this EOC?
 5. The discussions you’ve had?

APPENDIX D: DRAFT ALDS

Missouri End-of-Course Assessment Achievement-Level Descriptors—DRAFT

English II

Advanced: Students performing at the Advanced level on the Missouri-End-of-Course Assessment consistently demonstrate a thorough understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate higher level skills in reading processes, in responding to both fiction and nonfiction texts, and in writing effectively. In addition to understanding and applying the skills at the Proficient level, students scoring at the Advanced level use a range of strategies to comprehend and interpret a variety of texts, demonstrate a thorough understanding of literary forms, and consistently apply different strategies for accessing and summarizing information. They follow a writing process to compose well developed and organized papers for a variety of audiences and purposes while consistently and correctly applying the rules and conventions of Standard English.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning;
- Analyze the main idea and evaluate supporting details;
- Make sophisticated connections—compare, contrast, evaluate;
- Evaluate text features;
- Analyze complex figurative language and literary techniques;
- Draw insightful conclusions;
- Summarize and paraphrase ideas and information;
- Analyze story components and theme;
- Analyze literary elements;
- Evaluate reasoning, inferences, and sources;
- Evaluate proposed solutions;
- Evaluate accuracy and adequacy of evidence;
- Utilize organizational patterns;
- Evaluate the author’s point of view, viewpoint/perspective, and/ or purpose;
- Evaluate the author’s tone.

Writing

A student is able to write across genres a paper that

- Contains a strong controlling idea, along with an effective beginning, middle, and end;
- Uses paragraphing effectively;
- Progresses in a logical order and uses cohesive devices effectively;
- Addresses the topic clearly and provides specific and relevant details, reasons, and examples;
- Uses precise, vivid language in sentences that are clear and varied in structure;
- Effectively uses writing techniques;

- Shows complexity, freshness of thought, and individual perspective;
- Shows an awareness of audience and purpose;
- Contains few errors in Standard English and spelling.

A student is able to consistently and correctly apply the conventions of capitalization, punctuation, and standard usage.

Proficient: Students performing at the Proficient level on the Missouri-End-of-Course Assessment demonstrate an understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills in reading processes, in responding to both fiction and nonfiction texts, and in writing effectively. In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to comprehend and interpret a variety of texts, demonstrate an understanding of literary forms, and apply strategies for accessing and summarizing information. They follow a writing process to compose well developed and organized papers for a variety of audiences and purposes while correctly applying the rules and conventions of Standard English.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning;
- Identify the main idea and supporting details;
- Make connections—compare, contrast, evaluate;
- Analyze text features;
- Analyze figurative language and literary techniques;
- Draw accurate conclusions;
- Summarize and paraphrase ideas and information;
- Analyze story components and theme;
- Analyze literary elements;
- Analyze reasoning, inferences, and sources;
- Analyze proposed solutions;
- Analyze evidence and use of information;
- Utilize organizational patterns;
- Analyze author’s point of view, viewpoint/perspective, and/or purpose;
- Analyze the author’s tone.

Writing

A student is able to write across genres a paper that

- Contains a controlling idea, along with a clear beginning, middle, and end;
- Uses paragraphing appropriately;
- Progresses in a generally logical order and uses cohesive devices;
- Addresses the topic and provides details, reasons, and examples;
- Uses precise language in sentences that are clear in structure;
- Uses writing techniques;
- Shows some complexity, freshness of thought, and/or individual perspective;
- Shows awareness of audience and purpose;
- Contains some errors in Standard English and spelling.

A student is able to apply the conventions of capitalization, punctuation, and standard usage correctly.

Basic: Students performing at the Basic level on the Missouri-End-of-Course Assessment demonstrate an incomplete understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills inconsistently in reading processes, in responding to both fiction and nonfiction texts, and in writing. In addition to understanding and applying the skills at the Below Basic level, students scoring at the Basic level use some strategies to comprehend and interpret a variety of texts, demonstrate a partial understanding of literary forms, and inconsistently apply few strategies for accessing and summarizing information. They may follow a writing process to compose papers while inconsistently applying the rules of Standard English.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning;
- Identify the main idea and major details;
- Make simple connections—compare, contrast;
- Identify text features;
- Identify figurative language and literary techniques;
- Draw simple conclusions;
- Summarize and paraphrase basic ideas and information;
- Identify characters, plot, setting, and basic theme;
- Identify basic literary elements;
- Make simple inferences;
- Identify proposed solutions;
- Determine reliability of information;
- Identify organizational patterns;
- Identify author’s purpose; and point of view.

Writing

A student is able to write across genres a paper that

- Contains an idea, though it may lack focus, along with a beginning, middle, and end;
- Shows evidence of paragraphing;
- Progresses generally in a somewhat logical order and may use cohesive devices;
- Addresses the topic but relies on generalities rather than specifics;
- May use imprecise language in sentences that are generally clear in structure;
- May lack writing techniques;
- May lack complexity, freshness of thought, and individual perspective;
- Shows some awareness of audience and purpose;
- Contains errors in Standard English and spelling that may be distracting.

A student inconsistently applies the conventions of capitalization, punctuation, and standard usage.

Below Basic: Students performing at the Below Basic level on the Missouri-End-of-Course Assessment demonstrate little understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills inconsistently and/or incorrectly in reading processes, in responding to both fiction and nonfiction texts, and in writing. Students scoring at the Below Basic level use few strategies to comprehend and interpret texts, demonstrate little understanding of literary forms, and apply few strategies for accessing information. They may not follow a writing process to compose papers and/or incorrectly apply the rules and conventions of Standard English.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning;
- Identify the main idea and some details;
- Make simple connections;
- Identify simple text features;
- Identify figurative language;
- Identify characters, plot and setting;
- Determine literal meaning;
- Identify point of view.

Writing

A student is able to write across genres a paper that

- May contain an unfocused idea and may lack a beginning, middle, and/or end;
- May lack evidence of paragraphing;
- Does not progress in a logical order and lacks cohesion;
- May address the topic but lacks details;
- May use imprecise language in sentences that may be unclear in structure;
- Shows little evidence of writing techniques;
- Lacks complexity, freshness of thought, and individual perspective;
- Shows little or no awareness of audience or purpose;
- Contains repeated errors in Standard English and spelling that are distracting.

A student incorrectly applies the conventions of capitalization, punctuation, and standard usage.

Algebra I

Advanced: Students performing at the Advanced level on the Missouri Algebra I End-of-Course Assessment demonstrate a thorough understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in algebraic relationships. In addition to understanding and applying the skills at the Proficient level, students scoring at the Advanced level use a wide range of strategies to solve problems and demonstrate a thorough understanding of important mathematical content and concepts.

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using explicitly or recursively defined functions

- Describe the effects of parameter changes on exponential growth/decay and quadratic functions including intercepts
- Use symbolic algebra to represent and solve problems that involve quadratic relationships including equations and inequalities
- Describe and use algebraic manipulations, including factoring and apply properties of exponents to simplify expressions
- Use and solve equivalent forms of quadratic equations
- Use and solve systems of linear inequalities with 2 variables
- Analyze quadratic functions by investigating rates of change, intercepts, and zeros

Proficient: Students performing at the Proficient level on the Missouri Algebra I End-of-Course Assessment demonstrate an understanding of most Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to solve problems and demonstrate understanding of important mathematical content and concepts.

Number and Operations—Using numbers and operations, a student can

- Compare and order rational and irrational numbers, including finding their approximate locations on a number line
- Use real numbers and various models, drawings, etc. to solve problems

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using explicitly or recursively defined linear functions
- Compare and contrast various forms of representations of patterns
- Compare the properties of linear and nonlinear functions
- Describe the effects of parameter changes on linear functions including intercepts
- Use symbolic algebra to represent problems that involve linear relationships including equations and inequalities
- Describe and use algebraic manipulations, including rules of integer exponents to simplify expressions
- Use and solve equivalent forms of absolute value and linear equations
- Use and solve systems of linear equations with 2 variables
- Identify quantitative relationships and determine type(s) of functions that might model the situation to solve the problem
- Analyze linear functions by investigating rates of change, intercepts, and zeros

Data and Probability—Using data and probability, a student can

- Determine the distributions of the outcome of an experiment
- Use appropriate graphical representations of data
- Given one-variable quantitative data, display the distribution and describe its shape
- Apply statistical methods to measures of center to solve problems
- Given a scatterplot, determine an equation for a line of best fit

- Make conjectures about possible relationships between 2 characteristics of a sample on the basis of scatterplots of the data

Basic: Students performing at the Basic level on the Missouri Algebra I End-of-Course Assessment demonstrate an incomplete understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition to understanding and applying the skills at the Below Basic level, students scoring at the Basic level use some strategies to solve problems and demonstrate some understanding of important mathematical content and concepts.

Number and Operations—Using numbers and operations, a student can

- Compare and order rational numbers, including finding their approximate locations on a number line

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using recursively defined single operation functions
- Compare the properties of linear functions
- Use symbolic algebra to solve problems that involve linear relationships including equations and inequalities
- Describe and use algebraic manipulations, including order of operations to simplify expressions
- Use equivalent forms of linear equations
- Use and solve systems of linear equations with 2 variables

Data and Probability—Using data and probability, a student can

- Formulate questions and collect data about a characteristic

Below Basic: Students performing at the Below Basic level on the Missouri Algebra I End-of-Course Assessment demonstrate a limited understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. Students scoring at the Below Basic level use very few strategies to solve problems and demonstrate a limited understanding of important mathematical content and concepts.

Number and Operations—Using numbers and operations, a student can

- Compare and order rational numbers

Algebraic Relationships—Using algebraic relationships, a student can

- Identify a function as linear or nonlinear
- Use symbolic algebra to solve problems that involve 2 step linear equations

Data and Probability—Using data and probability, a student can

- Identify the sample space of an experiment
- Select appropriate graphical representation of data

- Determine measures of center

Biology

Advanced: Students performing at the Advanced level on the Missouri End-of-Course Assessment consistently demonstrate a thorough understanding of the Course-Level Expectations for Biology. They demonstrate these skills in ...

In addition to understanding and applying the skills at the Proficient level, students scoring at the Advanced level use a range of strategies to ...

Characteristics and Interactions of Living Organisms—A student can

- List of skills here
- Apply the law of conservation of mass and energy to a biochemical process
- Classify different ways to store energy and describe the transfer of energy in a food web
- Relate structure of organic compounds to their role in living systems
- Predict the movement of molecules across a selectively permeable membrane needed for a cell to maintain homeostasis
- Compare and contrast process used in movement of molecules across a semipermeable membrane—taking energy use into consideration
- Predict patterns of inheritance using Mendelian genetics, including sex-linked, in a monohybrid cross
- Relate the expression of genetic diseases in offspring to the genetic makeup of the parents

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- List of skills here
- Predict how populations within an ecosystem may change in response to changes in abiotic or biotic factors
- Predict the impact of changes within a food chain on energy use and flow
- Explain how natural selection is related to environmental changes or species adaptations
- Predict local and global effects on environmental resources when given a scenario describing natural phenomena

Scientific Inquiry—A student can

- List of skills here
- Use quantitative data to calculate results
- Communicate information from investigations in data tables and appropriate graphical forms
- Identify and justify constants and variables in a repeatable scientific investigation
- Design a repeatable multi-step scientific investigation
- Gather evidence in qualitative and quantitative forms
- Determine how technological advances can affect real-world situations

Proficient: Students performing at the Proficient level on the Missouri End-of-Course Assessment demonstrate an understanding of the Course-Level Expectations for Biology. They demonstrate these skills in ...

In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to...

Characteristics and Interactions of Living Organisms—A student can

- List of skills here
- Explain cell differentiation
- Explain the chemical and physical interactions between organelles as they carry out life processes
- Explain interrelationships between photosynthesis and respiration
- Determine factors that affect the processes of photosynthesis and respiration
- Explain how enzymes affect chemical reactions
- Explain homeostasis and its effect on cellular activities
- Identify the causes of mutations in DNA and explain the possible effects on the organism
- Describe transcription and translation in DNA and identify steps in the processes of mitosis and meiosis
- Explain the advantages and disadvantages of sexual and asexual reproduction within a population
- Describe diploid and haploid chromosome number
- Explain how daughter cells compare to the original parent cell (heredity information and number)
- Describe how new genetic combinations result in new heritable characteristics
- Explain how genotypes contribute to phenotypic variation within a species

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- List of skills here
- Identify and explain limiting factors (abiotic and biotic) that may affect carrying capacity
- Explain the impact a natural environmental event may have on the diversity of different species in an ecosystem
- Explain the impact human activity may have on the diversity of different species in an ecosystem
- Describe energy flow in a food web
- Explain the natural and/or human factors that may lead to the extinction of a species
- Identify the evidence found in the fossil records to support relationship among species over time

Scientific Inquiry—A student can

- List of skills here

- Formulate a testable hypothesis
- Identify constants and variables in an investigation
- Determine scientific conclusions based on observations
- Use patterns to extrapolate data to form conclusions
- Identify factors required to make investigative results reliable
- Analyze quantitative data
- Design scientific investigations consisting of at least three steps
- Identify technology used to collect data to increase scientific knowledge
- Explain why accurate records and replications are essential for experimental creditability
- Calculate percent and ratios from sets of data
- Communicate procedures and results of investigations
- Explain the importance of peer review of scientific findings

Basic: Students performing at the Basic level on the Missouri End-of-Course Assessment demonstrate an incomplete understanding of the Course-Level Expectations for Biology. They demonstrate these skills inconsistently in ...

In addition to understanding and applying the skills at the Below Basic level, students scoring at the Basic level use some strategies to ...

Characteristics and Interactions of Living Organisms—A student can

- List of skills here
- Identify and describe cell structures and functions
- Define organelles by their functions
- Describe the equation for photosynthesis and respiration
- Identify that the carbon that organisms use for growth comes from the carbon dioxide in the air (this probably needs a better word than growth, but this is a huge misconception that needs to be addressed)
- Explain how water is important to cells
- Use a Punnett square to show a simple monohybrid cross

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- List of skills here
- Describe interactions between organisms in a predator/prey relationship
- Explain how interactions within an ecosystem maintain balance
- Define carrying capacity of a population within an ecosystem
- Describe how a natural environmental event impacts diversity in an ecosystem
- Describe how human caused change impacts the diversity in an ecosystem
- Construct a simple food web
- Define species in terms of the ability to mate and reproduce
- Describe similarities in DNA between species
- Describe how adaptations may have provided a population an advantage for survival

- Explain how environmental factors can be agents of natural selection

Scientific Inquiry—A student can

- List of skills here
- Select appropriate investigation methods
- Use data to formulate an explanation
- Calculate average/mean for sets of data
- Identify possible effects of errors in data collection and calculations
- Identify and describe how scientific explanations have changed over time or as a result of new evidence (strand 8?)

Below Basic: Students performing at the Below Basic level on the Missouri End-of-Course Assessment demonstrate little understanding of the Course-Level Expectations for Biology. They demonstrate these skills inconsistently and/or incorrectly in ...

Students scoring at the Below Basic level inconsistently use some strategies to ...

Characteristics and Interactions of Living Organisms—A student can

- List of skills here
- Identify that all organisms progress through life cycles
- Identify that all organisms are made of cells
- Identify that water is important to cells (life?)
- Identify that all living organisms have DNA
- Identify that DNA carries inherited information

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- List of skills here
- Use a model to show that populations interact in an ecosystem
- Identify examples of adaptations resulting from natural selection

Scientific Inquiry—A student can

- List of skills here
- Identify a valid conclusion in an experiment
- Use simple tools to measure length, mass, and volume
- Communicate basic information from an experiment
- Construct a simple graph of independent variable versus dependent variable from given data
- Identify how humans impact the environment (strand 8)
- Identify one impact of technology on an environmental factor (also strand 8)

APPENDIX E: FINAL ALDS

Missouri End-of-Course Assessment Achievement-Level Descriptors—FINAL

English II

Achievement Levels

Advanced: Students performing at the Advanced level on the Missouri English II End-of-Course Assessment consistently demonstrate a thorough understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate higher-level skills in reading processes, in responding to both fiction and nonfiction texts, and in writing effectively. In addition to understanding and applying the skills at the Proficient level, students scoring at the Advanced level use a wide range of strategies to comprehend and interpret a variety of texts, demonstrate a thorough understanding of literary forms, and consistently apply different strategies for accessing and summarizing information. They follow a writing process to compose well-developed and organized papers for a variety of audiences and purposes, while consistently and correctly applying the rules and conventions of Standard English. Raw Score Cut: 33–39; Scale Score Cut: 225–250.

Proficient: Students performing at the Proficient level on the Missouri English II End-of-Course Assessment demonstrate an understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills in reading processes, in responding to both fiction and nonfiction texts, and in writing effectively. In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to comprehend and interpret a variety of texts, demonstrate an understanding of literary forms, and apply strategies for accessing and summarizing information. They follow a writing process to compose well-developed and organized papers for a variety of audiences and purposes, while correctly applying the rules and conventions of Standard English. Raw Score Cut: 24–32; Scale Score Cut: 200–224.

Basic: Students performing at the Basic level on the Missouri English II End-of-Course Assessment demonstrate an incomplete understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills inconsistently in reading processes, in responding to both fiction and nonfiction texts, and in writing. In addition to understanding and applying the skills at the Below Basic level, students scoring at the Basic level use some strategies to comprehend and interpret a variety of texts, demonstrate a partial understanding of literary forms, and inconsistently apply few strategies for accessing and summarizing information. They may follow a writing process to compose papers while inconsistently applying the rules and conventions of Standard English. Raw Score Cut: 15–23; Scale Score Cut: To be determined after operational data are complete.

Below Basic: Students performing at the Below Basic level on the Missouri English II End-of-Course Assessment demonstrate little understanding of the skills and processes identified in the Course-Level Expectations for English II. They demonstrate these skills inconsistently and/or incorrectly in reading processes, in responding to both fiction and

nonfiction texts, and in writing. Students scoring at the Below Basic level use few strategies to comprehend and interpret texts, demonstrate little understanding of literary forms, and apply few strategies for accessing information. They may not follow a writing process to compose papers and/or incorrectly apply the rules and conventions of Standard English. Raw Score Cut: 0–14; Scale Score Cut: To be determined after operational data are complete.

Achievement Descriptors

Advanced

Raw Score Cut: 33–39; Scale Score Cut: 225–250

Reading—In both fiction and nonfiction, a student can

- Determine vocabulary meaning
- Analyze the main idea and evaluate supporting details
- Make sophisticated connections—compare, contrast, evaluate
- Evaluate text features
- Analyze complex figurative language and literary techniques
- Draw insightful conclusions
- Summarize and paraphrase complex ideas and information
- Analyze literary elements
- Evaluate reasoning, inferences, and sources
- Evaluate proposed solutions
- Evaluate accuracy and adequacy of evidence
- Evaluate organizational patterns
- Evaluate the author’s point of view, viewpoint/perspective, and purpose
- Evaluate the author’s tone

Writing—A student is able to write across genres a paper that

- Contains a strong controlling idea, along with an effective beginning, middle, and end
- Uses paragraphing effectively
- Progresses in a logical order and uses cohesive devices effectively
- Addresses the topic clearly and provides specific and relevant details, reasons, and examples
- Uses precise, vivid language in sentences that are clear and varied in structure
- Effectively uses writing techniques
- Shows complexity, freshness of thought, and individual perspective
- Shows a clear awareness of audience and purpose
- Contains few errors in Standard English and spelling

A student is able to consistently and correctly apply the conventions of capitalization, punctuation, and standard usage.

Proficient

Raw Score Cut: 24–32; Scale Score Cut: 200–224

Reading—In both fiction and nonfiction, a student can

- Determine vocabulary meaning
- Identify the main idea and supporting details
- Make connections—compare, contrast, analyze
- Analyze text features
- Analyze figurative language and literary techniques
- Draw accurate conclusions
- Summarize and paraphrase ideas and information
- Analyze literary elements
- Analyze reasoning, inferences, and sources
- Analyze proposed solutions
- Analyze evidence and use of information
- Analyze organizational patterns
- Analyze the author’s point of view, viewpoint/perspective, and purpose
- Analyze the author’s tone

Writing—A student is able to write across genres a paper that

- Contains a controlling idea, along with a clear beginning, middle, and end
- Uses paragraphing appropriately
- Progresses in a generally logical order and uses cohesive devices
- Addresses the topic and provides details, reasons, and examples
- Uses precise language in sentences that are clear and show some variety in structure
- Uses writing techniques
- Shows some complexity, freshness of thought, and/or individual perspective
- Shows awareness of audience and purpose
- Contains some errors in Standard English and spelling

A student is able to apply the conventions of capitalization, punctuation, and standard usage correctly.

Basic

Raw Score Cut: 15–23; Scale Score Cut: To be determined after operational data are complete.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning
- Identify the main idea and major details
- Make simple connections—compare, contrast
- Identify text features
- Identify figurative language and literary techniques
- Draw basic/simple conclusions
- Summarize and paraphrase basic ideas and information
- Identify basic literary elements

- Make simple inferences
- Identify proposed solutions
- Determine reliability of information
- Identify organizational patterns
- Identify author’s purpose and point of view
- Identify author’s tone

Writing—A student is able to write across genres a paper that

- Contains an idea, though it may lack focus, along with a beginning, middle, and end
- Shows evidence of paragraphing
- Progresses generally in a somewhat logical order and may use cohesive devices
- Addresses the topic but relies on generalities rather than specifics
- May use imprecise language in sentences that are generally clear in structure
- May lack writing techniques
- May lack complexity, freshness of thought, and individual perspective
- Shows some awareness of audience and purpose
- Contains errors in Standard English and spelling that may be distracting

A student inconsistently applies the conventions of capitalization, punctuation, and standard usage.

Below Basic

Raw Score Cut: 0–14; Scale Score Cut: To be determined after operational data are complete.

Reading—In fiction and nonfiction, a student can

- Determine vocabulary meaning
- Identify the main idea and some details
- Make simple connections
- Identify simple text features
- Identify figurative language
- Identify characters, plot, and setting
- Determine literal meaning
- Identify point of view

Writing—A student is able to write across genres a paper that

- May contain an unfocused idea and may lack a beginning, middle, and/or end
- May lack evidence of paragraphing
- Does not progress in a logical order and lacks cohesion
- May address the topic but lacks details
- May use imprecise language in sentences that may be unclear in structure
- Shows little evidence of writing techniques
- Lacks complexity, freshness of thought, and individual perspective
- Shows little or no awareness of audience or purpose
- Contains repeated errors in Standard English and spelling that are distracting

A student incorrectly applies the conventions of capitalization, punctuation, and standard usage.

Algebra I

Achievement Levels

Advanced: Students performing at the Advanced level on the Missouri Algebra I End-of-Course Assessment demonstrate a thorough understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition to understanding and applying the skills at the Proficient level, students scoring at the Advanced level use a wide range of strategies to solve problems and demonstrate a thorough understanding of important mathematical content and concepts. Raw Score Cut: 31–39; Scale Score Cut: 225–250

Proficient: Students performing at the Proficient level on the Missouri Algebra I End-of-Course Assessment demonstrate an understanding of most Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to solve problems and demonstrate an understanding of important mathematical content and concepts. Raw Score Cut: 22–30; Scale Score Cut: 200–224

Basic: Students performing at the Basic level on the Missouri Algebra I End-of-Course Assessment demonstrate some understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition to understanding and applying the skills at the Below Basic level, students scoring at the Basic level use some strategies to solve problems and demonstrate some understanding of important mathematical content and concepts. Raw Score Cut: 13–21; Scale Score Cut: To be determined after operational data are complete.

Below Basic: Students performing at the Below Basic level on the Missouri Algebra I End-of-Course Assessment demonstrate a limited understanding of the Course-Level Expectations for Algebra I. They demonstrate these skills in number and operations, algebraic relationships, and data and probability. In addition, students scoring at the Below Basic level use very few strategies to solve problems and demonstrate a limited understanding of important mathematical content and concepts. Raw Score Cut: 0–12; Scale Score Cut: To be determined after operational data are complete.

Achievement Descriptors

Advanced

Raw Score Cut: 31–39; Scale Score Cut: 225–250

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using explicitly or recursively defined functions
- Describe the effects of parameter changes on exponential growth/decay and quadratic functions, including intercepts

- Use symbolic algebra to represent and solve problems that involve quadratic relationships, including equations and inequalities
- Describe and use algebraic manipulations, including factoring, and apply properties of exponents to simplify expressions
- Use and solve equivalent forms of quadratic and absolute value equations
- Identify quantitative relationships and determine type(s) of functions that might model the situation to solve a problem, including quadratic and exponential growth/decay
- Use and solve systems of linear inequalities with two variables
- Analyze quadratic functions by investigating rates of change, intercepts, and zeros

Proficient

Raw Score Cut: 22–30; Scale Score Cut: 200–224

Number and Operations—Using numbers and operations, a student can

- Compare and order rational and irrational numbers, including finding their approximate locations on a number line
- Use real numbers and various models, drawings, etc. to solve problems

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using explicitly or recursively defined linear functions
- Compare and contrast various forms of representations of patterns
- Compare and contrast the properties of linear and nonlinear functions
- Describe the effects of parameter changes on linear functions, including intercepts
- Use symbolic algebra to represent problems that involve linear relationships, including equations and inequalities
- Describe and use algebraic manipulations, including rules of integer exponents, to simplify expressions
- Use and solve equivalent forms of absolute value and linear equations
- Use and solve systems of linear equations with two variables
- Identify quantitative relationships that can be modeled by linear functions to solve a problem
- Analyze linear functions by investigating rates of change, intercepts, and zeros

Data and Probability—Using data and probability, a student can

- Use appropriate graphical representations of data
- Given one-variable quantitative data, display the distribution and describe its shape
- Apply statistical methods to measures of center to solve problems
- Given a scatterplot, determine an equation for a line of best fit
- Make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data

Basic

Raw Score Cut: 13–21; Scale Score Cut: To be determined after operational data are complete.

Number and Operations—Using numbers and operations, a student can

- Compare and order rational numbers, including finding their approximate locations on a number line

Algebraic Relationships—Using algebraic relationships, a student can

- Generalize patterns using recursively defined single-operation functions
- Compare the properties of linear functions
- Use symbolic algebra to solve problems that involve linear relationships, including equations and inequalities
- Describe and use algebraic manipulations, including order of operations, to simplify expressions
- Use equivalent forms of linear equations

Data and Probability—Using data and probability, a student can

- Determine the sample space of an experiment
- Formulate questions about a characteristic which include sample spaces and distributions

Below Basic

Raw Score Cut: 0–12; Scale Score Cut: To be determined after operational data are complete.

Number and Operations—Using numbers and operations, a student can

- Compare and order rational numbers

Algebraic Relationships—Using algebraic relationships, a student can

- Identify a function as linear or nonlinear
- Use symbolic algebra to solve problems that involve two-step linear equations

Data and Probability—Using data and probability, a student can

- Identify the sample space of an experiment
- Select appropriate graphical representations of data
- Determine measures of center

Biology

Achievement Levels

Advanced: Students performing at the Advanced level on the Missouri End-of-Course Assessment demonstrate a thorough understanding of the Course-Level Expectations for Biology. They demonstrate these skills in addition to understanding and applying the

skills at the Proficient level; students scoring at the Advanced level use a range of strategies. Raw Score Cut: 45–55; Scale Score Cut: 225–250

Proficient: Students performing at the Proficient level on the Missouri End-of-Course Assessment demonstrate an understanding of the Course-Level Expectations for Biology. They demonstrate these skills in addition to understanding and applying the skills at the Basic level; students scoring at the Proficient level use a range of strategies. Raw Score Cut: 32–44; Scale Score Cut: 200–224

Basic: Students performing at the Basic level on the Missouri End-of-Course Assessment demonstrate a partial understanding of the Course-Level Expectations for Biology. They demonstrate these skills in addition to understanding and applying the skills at the Below Basic level; students scoring at the Basic level use some strategies. Raw Score Cut: 18–31; Scale Score Cut: To be determined after operational data are complete.

Below Basic: Students performing at the Below Basic level on the Missouri End-of-Course Assessment demonstrate a limited understanding of the Course-Level Expectations for Biology. Students scoring at the Below Basic level use very few strategies and demonstrate a limited understanding of important Biological content and concepts. Raw Score Cut: 0–17; Scale Score Cut: To be determined after operational data are complete.

Achievement Descriptors

Advanced

Raw Score Cut: 45–55; Scale Score Cut: 225–250

Characteristics and Interactions of Living Organisms—A student can

- Predict the movement of molecules across a selectively permeable membrane needed for a cell to maintain homeostasis
- Compare and contrast process used in movement of molecules across a semipermeable membrane, taking energy use into consideration
- Predict patterns of inheritance, using Mendelian genetics, in a monohybrid cross

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- Predict how populations within an ecosystem may change in response to changes in abiotic or biotic factors
- Predict the impact of changes within in a food chain based on energy use and flow
- Explain how natural selection is related to environmental changes or species adaptations

Scientific Inquiry—A student can

- Use quantitative data to calculate results
- Communicate information from investigations in data tables and appropriate graphical forms
- Identify and justify constants and variables in a repeatable scientific investigation
- Design a repeatable multistep scientific investigation

- Recognize it is not always possible, for practical or ethical reasons, to control some conditions (e.g., when sampling or testing humans, when observing animal behaviors in nature)

Proficient

Raw Score Cut: 32–44; Scale Score Cut: 200–224

Characteristics and Interactions of Living Organisms—A student can

- Identify cell differentiation
- Explain the chemical and physical interactions between organelles as they carry out life processes
- Explain interrelationships between photosynthesis and respiration (reactant and product only)
- Determine factors that affect the processes of photosynthesis and respiration (excludes light intensity)
- Identify homeostasis and its effect on cellular activities
- Identify the causes of mutations in DNA and explain the possible effects on the organism
- Describe the chemical and structural properties of DNA
- Recognize that DNA codes for proteins, which are expressed as the heritable characteristics of an organism
- Compare the processes of mitosis and meiosis (excludes identification of steps)
- Explain the advantages and disadvantages of sexual and asexual reproduction within a population
- Identify diploid and haploid chromosome number
- Explain how daughter cells compare to the original parent cell
- Explain how genotypes contribute to phenotypic variation within a species

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- Identify and explain limiting factors (abiotic and biotic) that may affect carrying capacity
- Describe how a natural environmental event impacts diversity in an ecosystem
- Explain the impact human activity may have on the diversity of different species in an ecosystem
- Predict the energy flow in a food web
- Explain the natural and/or human factors that may lead to the extinction of a species
- Given a scenario describing an environmental change, hypothesize why a given species was unable to survive

Scientific Inquiry—A student can

- Formulate a testable hypothesis
- Identify constants and variables in an investigation

- Determine the appropriate tools and techniques to collect, analyze, and interpret data
- Determine scientific conclusion based on observations
- Identify factors required to make investigative results reliable
- Analyze quantitative data
- Design scientific investigations consisting of at least three steps
- Explain why accurate records and replications are essential for experimental creditability (includes peer review)
- Communicate procedures and results of investigations

Basic

Raw Score Cut: 18–31; Scale Score Cut: To be determined after operational data are complete.

Characteristics and Interactions of Living Organisms—A student can

- Identify and describe cell structures and functions
- Define organelles by their functions
- Explain how water is important to cells
- Use a Punnett square to show a simple monohybrid cross

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- Explain how interactions within an ecosystem maintain balance
- Explain the nature of interactions between organisms in predator/prey relationships and different symbiotic relationships (i.e., mutualism, commensalism, parasitism)
- Define carrying capacity of a population within an ecosystem
- Identify how adaptations may have provided a population an advantage for survival
- Identify the impact a natural environmental event may have on the diversity of different species in an ecosystem
- Explain how environmental factors can be agents of natural selection
- Explain the importance of reproduction to the survival of a species

Scientific Inquiry—A student can

- Select appropriate investigation methods (techniques only)
- Use data to formulate an explanation
- Calculate average/mean for sets of data
- Identify possible effects of errors in data collection and calculations

Below Basic

Raw Score Cut: 0–17; Scale Score Cut: To be determined after operational data are complete.

Characteristics and Interactions of Living Organisms—A student can

- Identify that all organisms progress through life cycles
- Identify that all organisms are made of cells
- Identify that water is important to cells
- Identify that all living organisms have DNA
- Identify that DNA carries inherited information

Changes in Ecosystems and Interactions of Organisms with Their Environments—A student can

- Describe interactions between organisms in a predator/prey relationship
- Use a model to show that populations interact in an ecosystem
- Identify examples of adaptations resulting from natural selection

Scientific Inquiry—A student can

- Identify a valid conclusion in an experiment
- Use simple tools to measure length, mass, and volume
- Communicate basic information from an experiment
- Construct a simple graph of independent variable versus dependent variable from given data

APPENDIX F: QUALIFYING TEST

EOC Assessment: E A B

Judge # _____

Pre-Standard-Setting Self-Evaluation Assessment for Judges of the Missouri EOC Assessments (PSSSEAJMEOCA)

Directions: Circle the letter next to your answer for each item. Don't copy from your neighbor; he/she hasn't been listening very closely.

1. Why are the Achievement-Level Descriptors such an integral part of the standard-setting process?
 - A. They provide an anchor that gives concrete meaning to the terms Basic, Proficient, and Advanced.
 - B. All students at a given performance level should possess all critical behaviors and understandings listed in the ALDs.
 - C. They define all of the items that are contained on the EOC.
 - D. They summarize all of the elements of the Course-Level Expectations for the course.
2. Which of these statements about standard setting is TRUE?
 - A. Panelists should use their best judgment to make their recommendations, but should rely more on various data to be provided during the sessions.
 - B. While the EOC assessments are given statewide, judges should make recommendations based on the unique characteristics of *their* districts since other panelists will focus on other district types.
 - C. A judge who concludes that the "proper" cut score for Proficient is 24 should make a final recommendation of 22 or 23 to account for errors that are present in any assessment.
 - D. Judges must consider both the "stem" *and* answer options in multiple-choice items in deciding the percent of students who should answer correctly.
3. Joe the Judge decided that about 50% of the typical Proficient children in Missouri taking the EOC assessment should answer Item 32 correctly. He coded 50% under Proficient on his Rating Form. What error did he make?
 - A. He should have coded 45% since some percent of special-needs students will take the assessment.
 - B. He should have considered *barely* Proficient, not *typical* Proficient, students.
 - C. He should reconsider his judgment, as 50% correct couldn't possibly be considered Proficient.
 - D. He made no error here. This was the correct procedure.

4. Judge Jan reviewed the performance event for her EOC and decided that the average score of borderline Proficient students should be 2 out of 4. What should she enter on her Rating Form?
- A. 50%, since 2 out of 4 is 50%
 - B. 2.5, since she decided that 2 was the minimum acceptable score
 - C. 1.5, since the minimum expected score should be somewhat lower than the average score
 - D. 2, since her judgment is that 2 should be the average score of the target group
5. Which of these sets of “Angoff” judgments for a multiple-choice item appears to be improper and why?

	Below Basic/Basic	Basic/ Proficient	Proficient/ Advanced
A.	25%	35%	40%
B.	80%	90%	100%
C.	50%	50%	55%
D.	40%	75%	95%

- A. A, because these are very low expectations for a multiple-choice item
- B. B, because it is unrealistic to expect students to score this well on a multiple-choice item
- C. C, because the judge doesn't expect higher-classified students to perform any better on the item than lower-classified students
- D. D, because the increase in percentages across the three groups is probably unrealistically large

APPENDIX G: PARTICIPANT EVALUATION

Missouri End-of-Course Standard Setting

EVALUATION FORM

This form contains six sections, five of which ask for feedback on specific aspects of this standard-setting meeting. The last section asks for general reactions to the standard-setting meeting. Please fill out each of these sections as completely as possible in order to provide information that will help in the improvement of similar meetings in the future. Your identification number is used for analysis purposes only. Your responses to these questions will be held in strict confidence and will be analyzed in conjunction with those of the other judges who participated in this meeting.

Judge's I.D. (optional) _____

Section I: Opening Training Sessions

The following statements seek your judgments about the Opening Sessions for the Missouri End-of-Course standard-setting meeting. Please circle the value on the scale under each statement that best characterizes your judgment.

1. The Opening Sessions provided adequate background information about the Missouri End-of-Course Assessments.

5	4	3	2	1
Completely		Somewhat		Not at all

2. The topics covered in the Opening Sessions were appropriate to providing a context for my role in this meeting.

5	4	3	2	1
Completely		Somewhat		Not at all

3. The content of the Opening Sessions was:

5	4	3	2	1
Very useful		Somewhat useful		Not useful

4. The organization of the Opening Sessions was:

5	4	3	2	1
Very good		Acceptable		Very poor

The following statements also seek your judgments about the Opening Sessions for the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

5. Were there questions or concerns that were not answered or addressed in the Opening Sessions? Please indicate these below. (Use reverse side for additional space.)

6. What was most helpful about the Opening Sessions?

7. Please use the space below to provide additional comments concerning the adequacy, appropriateness, usefulness, or organization of the Opening Sessions.

Section II: Discussing Proficient Performance

The following statements seek your judgments about the discussions of Proficient performance as they relate to Missouri's End-of-Course Assessments. Please circle the value on the scale under each statement that best characterizes your judgment.

8. The activities used to help operationalize Proficient performance were:

5	4	3	2	1
Very useful	Somewhat useful		Not useful	

9. By the end of the activity, my conception of Proficient performance was:

5	4	3	2	1
Very well formed	Moderately well formed		Not well formed	

The following statement also seeks your judgments about the discussions of Proficient performance as they relate to Missouri's End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

10. Please use the space below to provide additional comments concerning the activities around operationalizing Proficient performance for Missouri's End-of-Course Assessments.

Section III: Discussing Basic Performance

The following statements seek your judgments about the discussions of Basic performance as they relate to Missouri's End-of-Course Assessments. Please circle the value on the scale under each statement that best represents your judgment.

11. The activities used to help operationalize Basic performance were:

5	4	3	2	1
Very useful		Somewhat useful		Not useful

12. By the end of this activity my conception of Basic performance was:

5	4	3	2	1
Very well formed		Moderately well formed		Not well formed

The following statement also seeks your judgments about the discussions of Basic performance as they relate to Missouri's End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

13. Please use the space below to provide additional comments concerning the activities around operationalizing Basic performance for Missouri's End-of-Course Assessments.

Section IV: Discussing Advanced Performance

The following statements seek your judgments about the discussions of Advanced performance as they relate to Missouri's End-of-Course Assessments. Please circle the value on the scale under each statement that best represents your judgment.

14. The activities used to help operationalize Advanced performance were:

5	4	3	2	1
Very useful		Somewhat useful		Not useful

15. By the end of this activity my conception of Advanced performance was:

5	4	3	2	1
Very well formed		Moderately well formed		Not well formed

The following statement also seeks your judgments about the discussions of Advanced performance as they relate to Missouri's End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

16. Please use the space below to provide additional comments concerning the activities around operationalizing Advanced performance for Missouri's End-of-Course Assessments.

Section V: Item Rating Activities

The following statements seek your judgments about the item rating activities as they relate to the Missouri End-of-Course standard-setting meeting. Please circle the value on the scale under each statement that best represents your judgment.

17. Using the sample items to prepare for the actual item rating was:

5	4	3	2	1
Very helpful		Somewhat helpful		Not helpful

18. The explanation of the item data during the sample item portion of the training was:

5	4	3	2	1
Very helpful		Somewhat helpful		Not helpful

19. The Item Rating Form was:

5	4	3	2	1
Very easy to use		Somewhat easy to use		Not at all easy to use

20. The information provided prior to each round of rating was:

5	4	3	2	1
Very useful		Somewhat useful		Not useful

21. My level of understanding of the tasks I was to accomplish for each round was:

5	4	3	2	1
Very good		Acceptable		Very poor

22. The amount of time I had to complete the tasks during each round was:

5	4	3	2	1
Far too long		About right		Far too short

The following statement seeks your judgments about the item rating activities as they relate to the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

23. Please use the space below to provide additional comments concerning the instructions and explanations you received, the adequacy of the time available, your levels of understanding of the process, or any other aspects of the estimates for the multiple-choice items. (Use reverse side for additional space.)

Section VI: The Overall Missouri End-of-Course Standard-Setting Meeting

The following statements seek your judgments about the overall processes and procedures used during the Missouri End-of-Course standard-setting meeting in which you participated as a panelist and the resulting recommended standards. Please circle the value on the scale under each statement that best represents your judgment.

24. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Proficient performance.

5	4	3	2	1
To a great extent		To some extent		Not at all

25. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Basic performance.

5	4	3	2	1
To a great extent	To some extent			Not at all

26. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Advanced performance.

5	4	3	2	1
To a great extent	To some extent			Not at all

27. I believe that this standard-setting meeting has produced recommended cut scores that are defensible.

5	4	3	2	1
To a great extent	To some extent			Not at all

28. I feel that this standard-setting meeting has produced recommended cut scores that would generally be considered as reasonable.

5	4	3	2	1
To a great extent	To some extent			Not at all

The following statements seek your judgments about the overall processes and procedures used during the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

29. Please provide any comments you wish to share regarding the quality of assistance provided by the standard-setting staff.

30. Please provide any additional comments you wish to share regarding the overall meeting.

APPENDIX H: RESULTS FOR ENGLISH II

**Standard Setting for the Missouri EOC Assessment
English II**

Round 1 Ratings Summary

Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
E211	16	27	34
E231	18	24	32
E232	18	24	30
E121	17	25	33
E223	18	23	30
E233	18	25	33
E331	13	26	31
E123	14	29	33
E311	18	29	35
E221	16	22	26
E313	10	23	32
E113	24	30	33
E222	13	20	27
E332	16	22	31
Median Rating:	16.5	24.5	32.0
Average Rating:	16.36	24.93	31.43
Standard Deviation:	3.18	2.87	2.44
Lowest Rating:	10	20	26
Highest Rating:	24	30	35
Number of Items:	36	36	36
Points Possible:	39	39	39
Number of Raters:	14	14	14

Standard Setting for the Missouri EOC Assessment English II

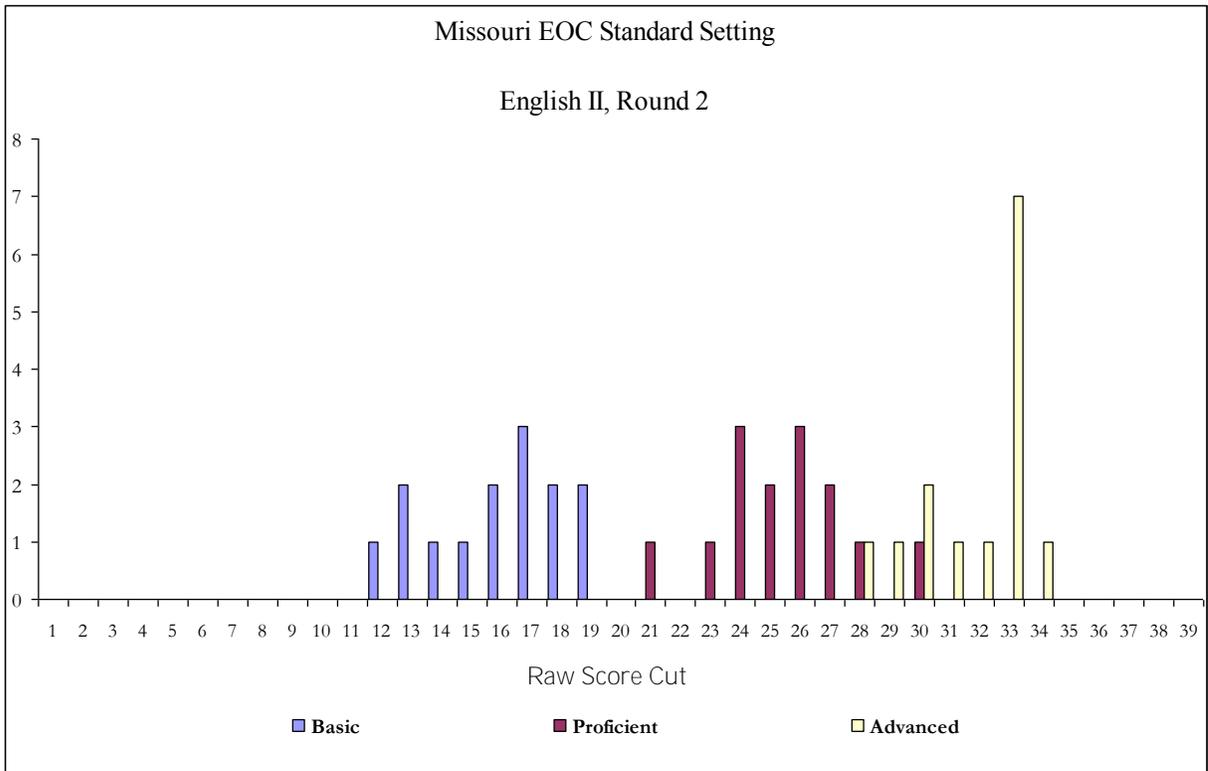
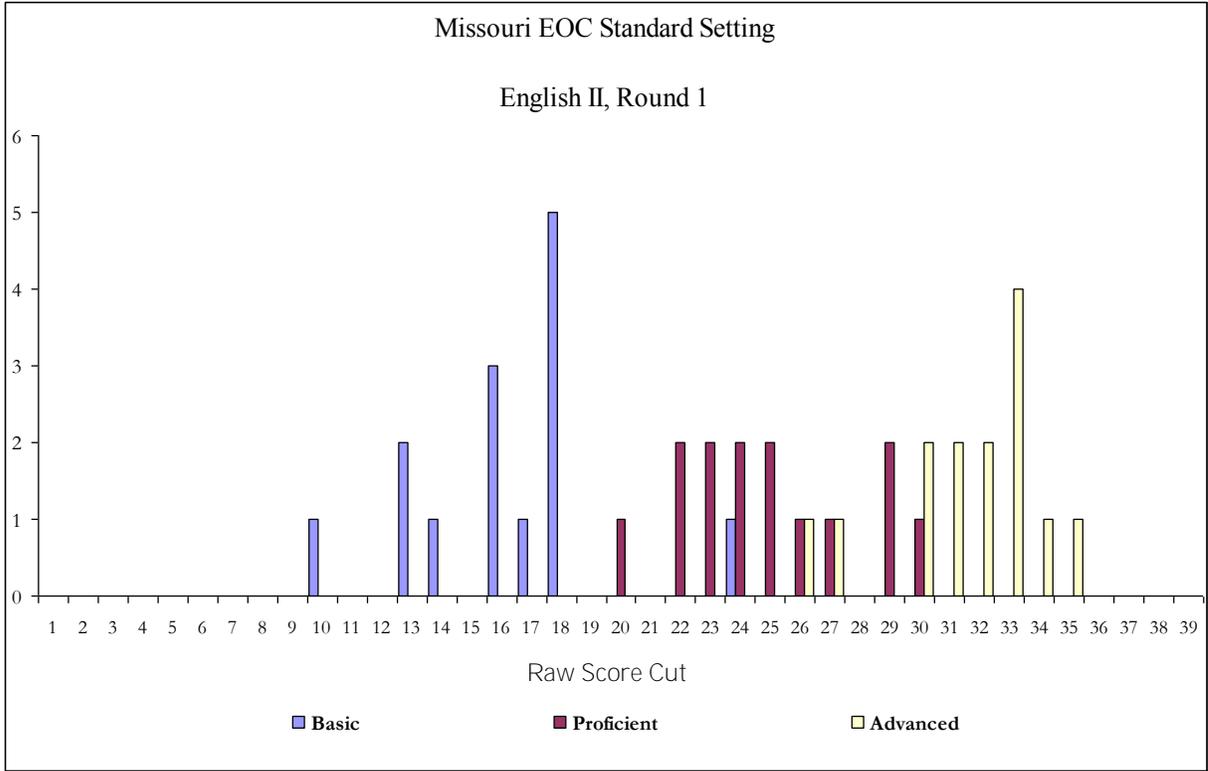
Round 2 Ratings Summary

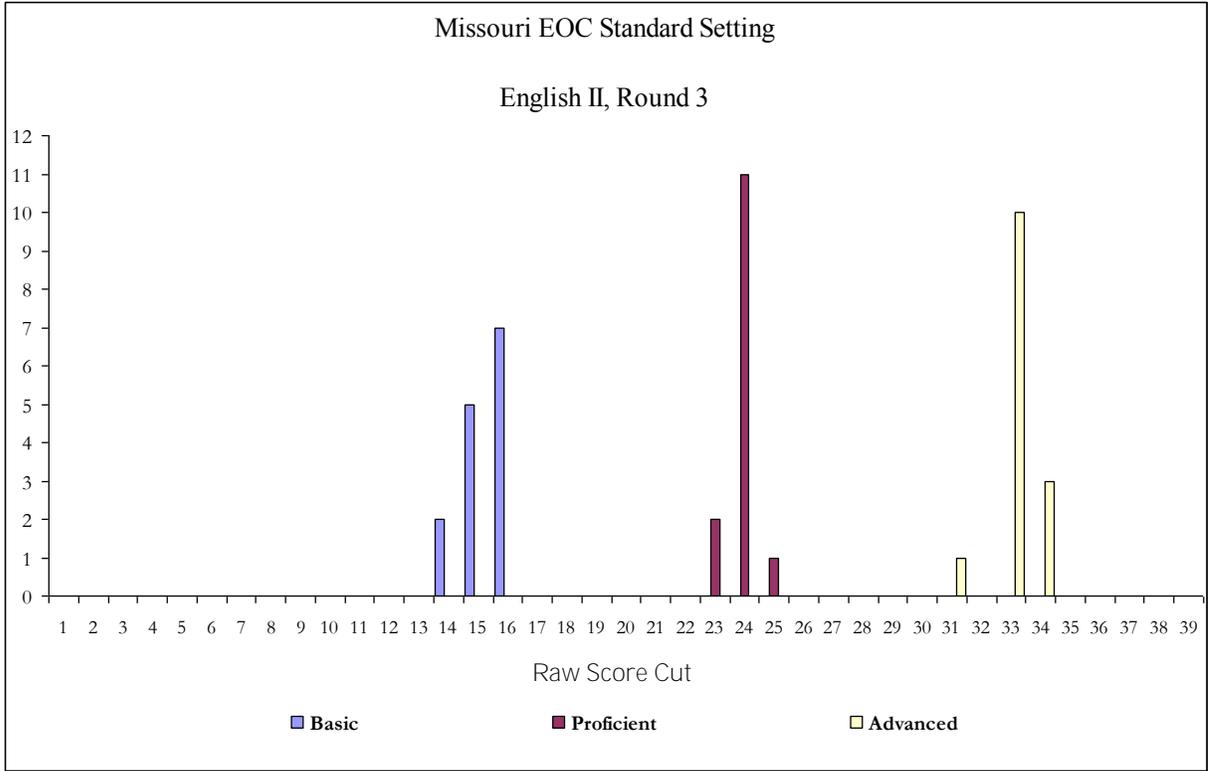
Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
E331	12	26	31
E113	16	30	33
E121	17	26	33
E123	14	28	33
E211	15	25	33
E221	17	25	30
E222	13	21	28
E223	18	24	30
E232	17	23	29
E233	19	26	33
E311	16	27	34
E313	13	24	32
E332	18	24	33
E231	19	27	33
Median Rating:	16.5	25.5	33.0
Average Rating:	16.00	25.43	31.79
Standard Deviation:	2.20	2.16	1.78
Lowest Rating:	12	21	28
Highest Rating:	19	30	34
Number of Items:	36	36	36
Points Possible:	39	39	39
Number of Raters:	14	14	14

Standard Setting for the Missouri EOC Assessment English II

Round 3 Ratings Summary

Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
E311	15	23	34
E232	15	24	33
E233	16	24	33
E222	14	23	31
E331	14	24	33
E223	16	24	33
E211	15	24	33
E121	15	24	34
E123	15	24	33
E221	16	24	33
E231	16	24	33
E113	16	24	33
E313	16	25	34
E332	16	24	33
Median Rating:	15.5	24.0	33.0
Average Rating:	15.36	23.93	33.07
Standard Deviation:	0.72	0.46	0.70
Lowest Rating:	14	23	31
Highest Rating:	16	25	34
Number of Items:	36	36	36
Points Possible:	39	39	39
Number of Raters:	14	14	14





Standard Setting for the Missouri EOC Assessment Algebra I

Round 2 Ratings Summary

Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
A122	13	21	33
A321	14	25	33
A121	12	23	32
A222	14	23	30
A322	13	21	27
A211	13	20	33
A333	16	25	32
A213	11	21	31
A312	12	23	31
A112	14	21	29
A123	16	25	31
A233	9	21	30
A311	14	27	34
A313	18	22	27
A232	11	23	35

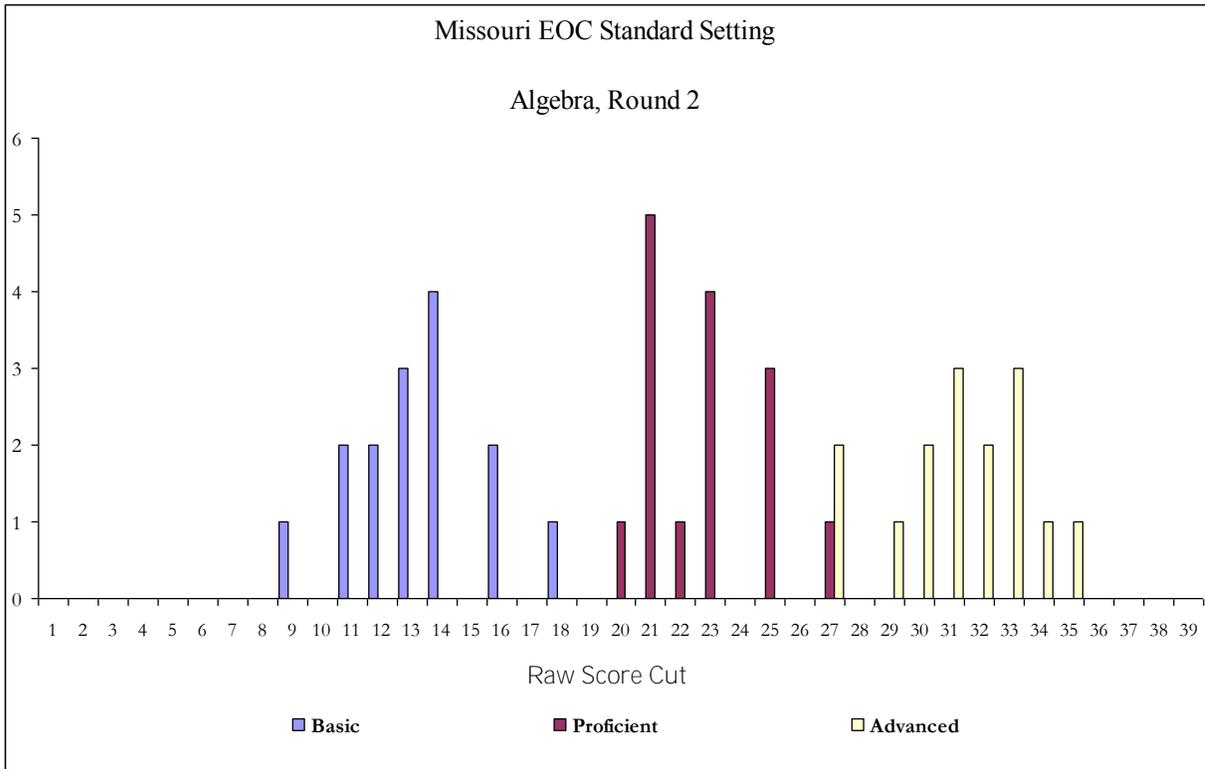
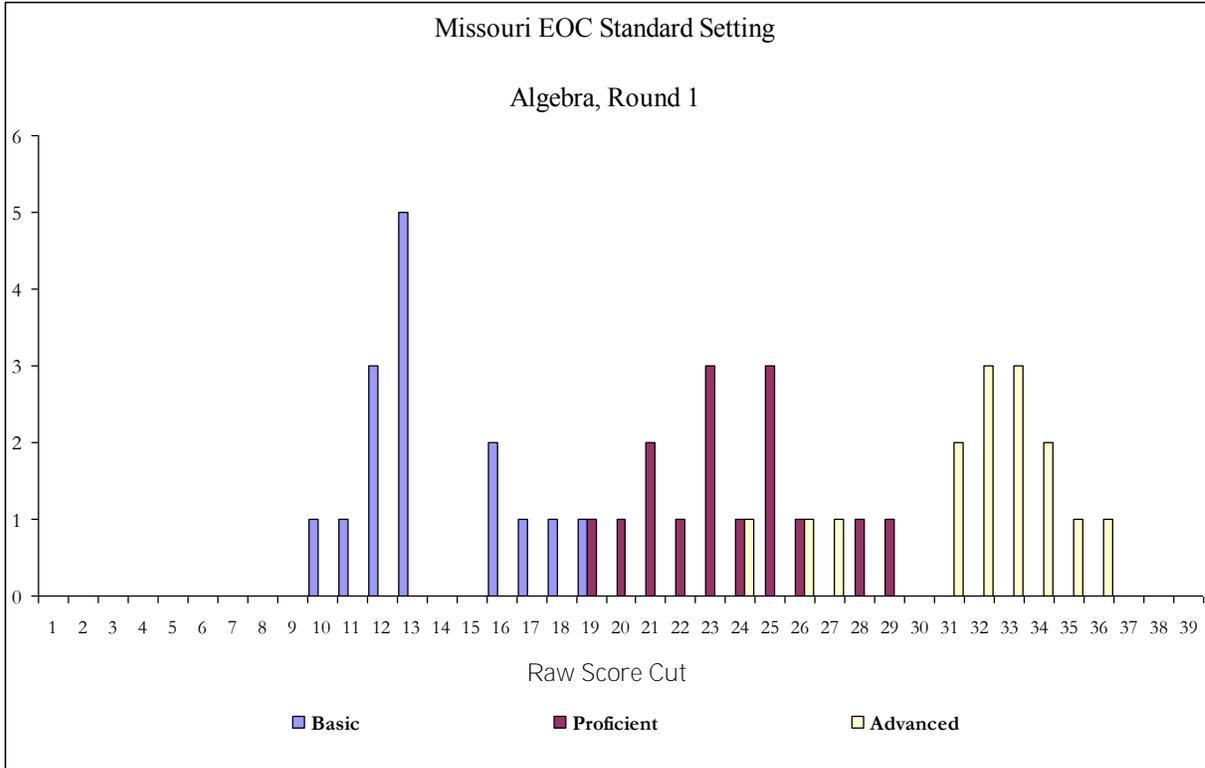
Median Rating:	13.0	23.0	31.0
Average Rating:	13.33	22.73	31.20
Standard Deviation:	2.18	1.95	2.26
Lowest Rating:	9	20	27
Highest Rating:	18	27	35
Number of Items:	36	36	36
Points Possible:	39	39	39
Number of Raters:	15	15	15

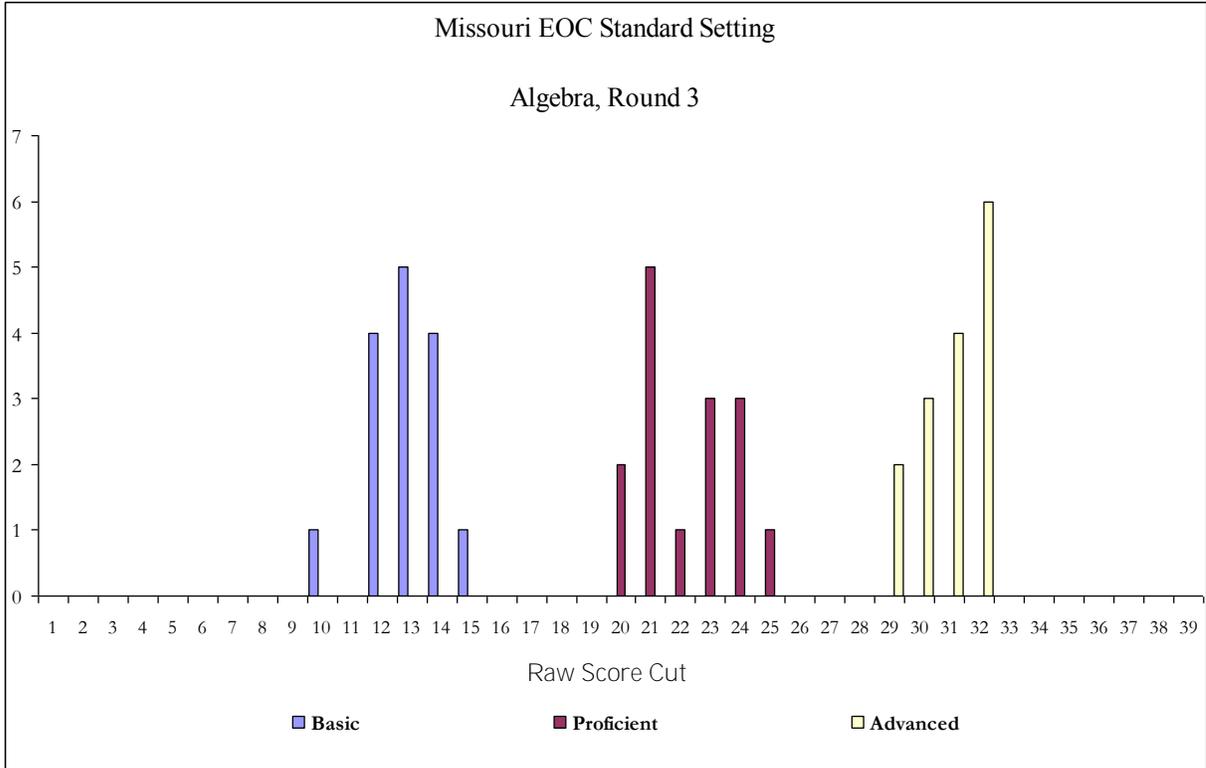
Standard Setting for the Missouri EOC Assessment Algebra I

Round 3 Ratings Summary

Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
A312	13	23	31
A122	12	20	31
A211	13	21	32
A232	10	20	31
A112	13	21	29
A121	12	21	32
A322	13	22	30
A313	13	23	29
A321	14	24	32
A311	14	25	32
A233	12	21	30
A333	15	24	32
A123	14	24	31
A213	12	21	32
A222	14	23	30

Median Rating:	13.0	22.0	31.0
Average Rating:	12.93	22.20	30.93
Standard Deviation:	1.18	1.56	1.06
Lowest Rating:	10	20	29
Highest Rating:	15	25	32
Number of Items:	36	36	36
Points Possible:	39	39	39
Number of Raters:	15	15	15



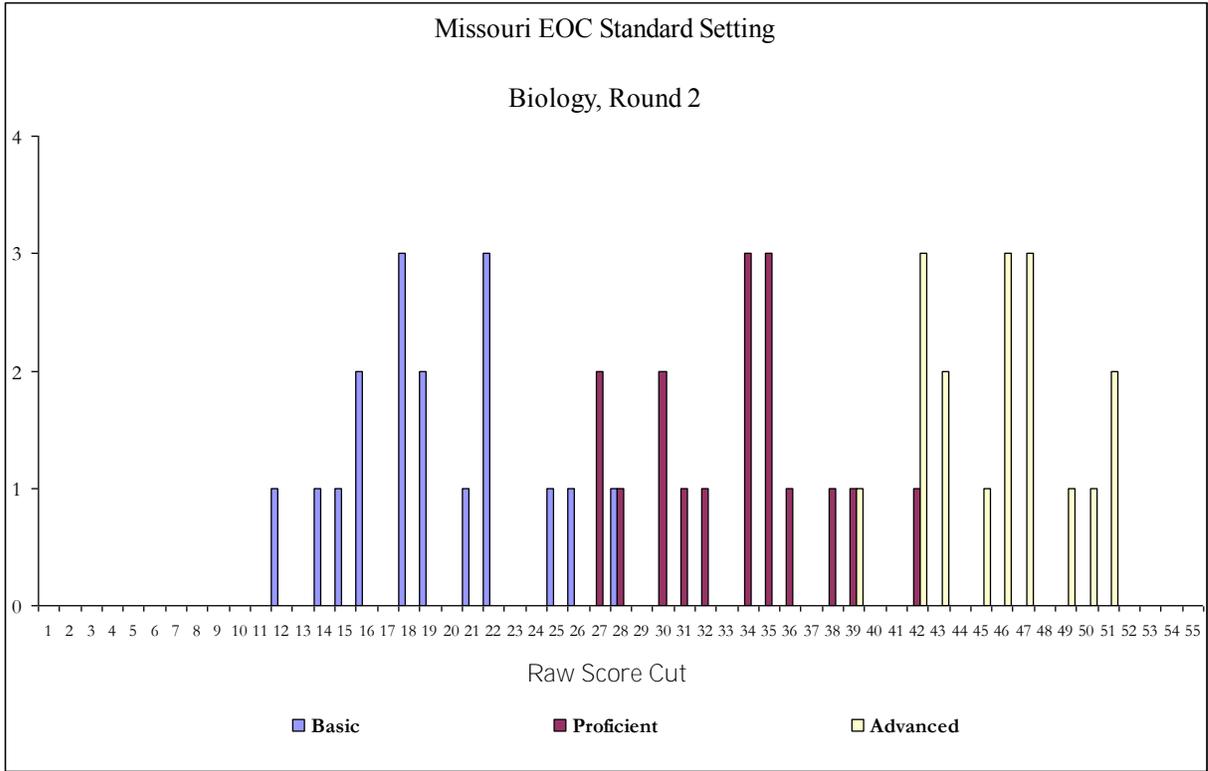
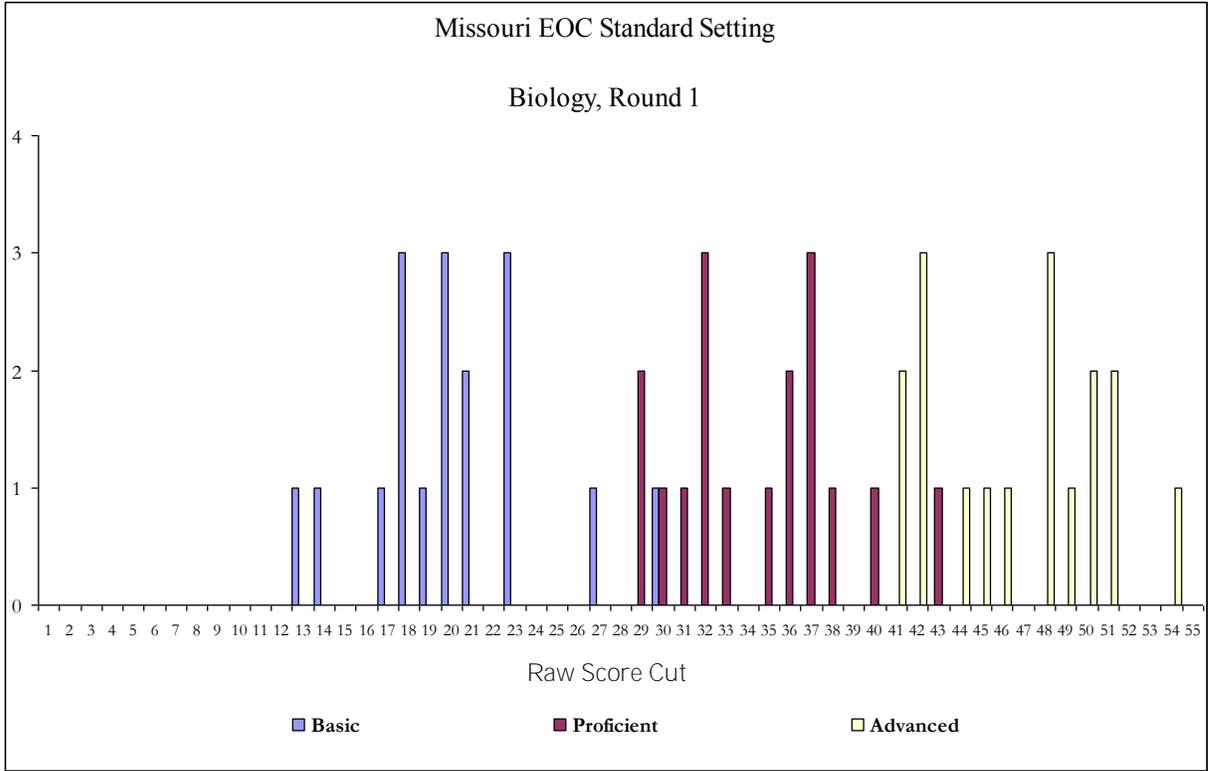


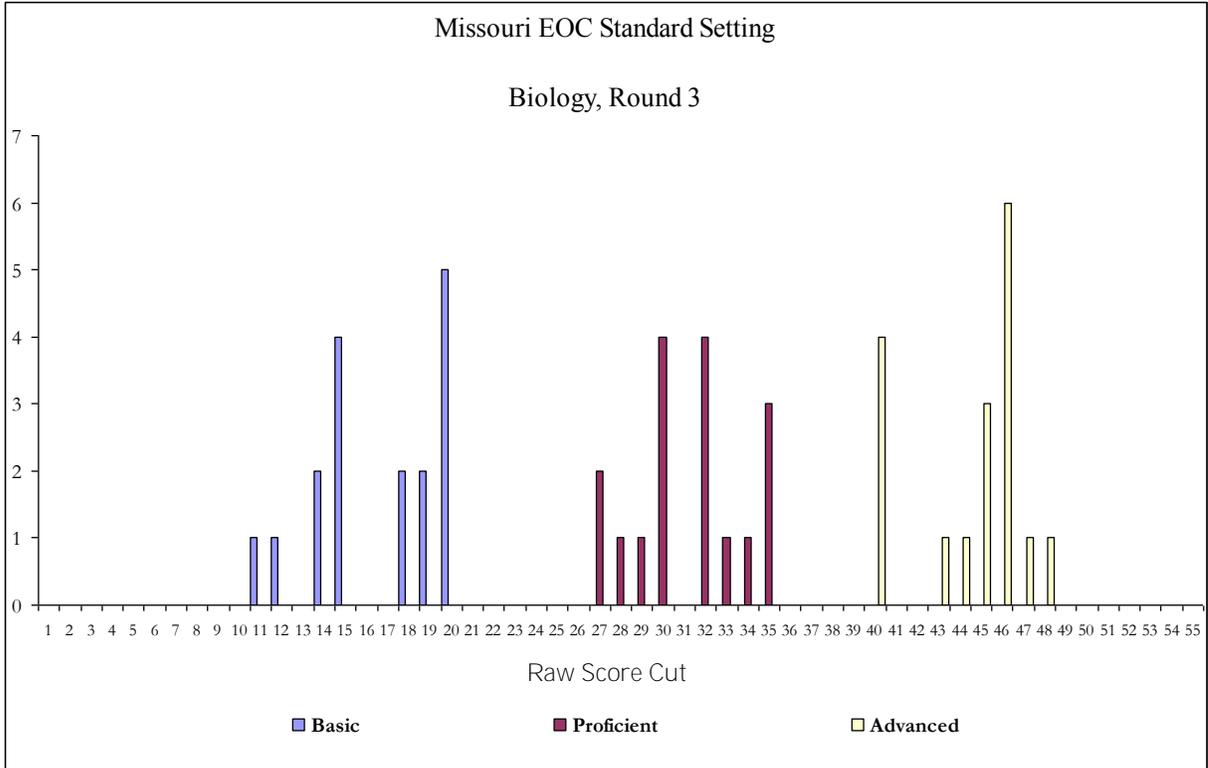
Standard Setting for the Missouri EOC Assessment Biology

Round 3 Ratings Summary

Rater	Individual Rater Cut Scores		
	Basic	Proficient	Advanced
B223	18	35	46
B232	19	32	46
B322	20	35	46
B111	20	34	48
B213	20	32	45
B332	15	30	46
B211	11	28	40
B212	15	30	40
B321	14	27	40
B131	20	35	45
B312	18	33	47
B231	19	32	46
B112	12	27	46
B311	20	32	45
B233	15	30	40
B113	15	30	44
B122	14	29	43

Median Rating:	18.0	32.0	45.0
Average Rating:	16.76	31.24	44.29
Standard Deviation:	2.96	2.58	2.61
Lowest Rating:	11	27	40
Highest Rating:	20	35	48
Number of Items:	46	46	46
Points Possible:	55	55	55
Number of Raters:	17	17	17





APPENDIX K: RESULTS OF PARTICIPANT EVALUATION

Missouri End-of-Course Standard Setting

EVALUATION FORM

This form contains six sections, five of which ask for feedback on specific aspects of this standard-setting meeting. The last section asks for general reactions to the standard-setting meeting. Please fill out each of these sections as completely as possible in order to provide information that will help in the improvement of similar meetings in the future. Your identification number is used for analysis purposes only. Your responses to these questions will be held in strict confidence and will be analyzed in conjunction with those of the other judges who participated in this meeting.

Judge's I.D. (optional) _____

Section I: Opening Training Sessions

The following statements seek your judgments about the Opening Sessions for the Missouri End-of-Course standard-setting meeting. Please circle the value on the scale under each statement that best characterizes your judgment.

- The Opening Sessions provided adequate background information about the Missouri End-of-Course Assessments.

	5	4	3	2	1
	Completely		Somewhat		Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	10	67%	5	36%	6	35%
4	3	20%	8	57%	11	65%
3	2	13%	1	7%	0	0%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.5	.74	4.3	.61	4.4	.49

*Percentages may not sum to 100% due to rounding.

2. The topics covered in the Opening Sessions were appropriate to providing a context for my role in this meeting.

5	4	3	2	1
Completely	Somewhat		Not at all	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (15)	%
5	9	60%	8	57%	7	41%
4	4	27%	4	29%	10	59%
3	2	13%	2	14%	0	0%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.5	.74	4.2	.76	4.4	.51

*Percentages may not sum to 100% due to rounding.

3. The content of the Opening Sessions was:

5	4	3	2	1
Very useful	Somewhat useful		Not useful	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	8	53%	2	14%	3	18%
4	4	27%	11	79%	11	65%
3	3	20%	1	7%	3	18%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.3	.82	4.1	.47	4.0	.61

*Percentages may not sum to 100% due to rounding.

4. The organization of the Opening Sessions was:

5	4	3	2	1
Very good		Acceptable		Very poor

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	8	53%	3	21%	2	12%
4	4	27%	9	64%	9	53%
3	1	7%	1	7%	6	35%
2	2	13%	1	7%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.2	1.08	4.0	.78	3.8	.66

*Percentages may not sum to 100% due to rounding.

The following statements also seek your judgments about the Opening Sessions for the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

5. Were there questions or concerns that were not answered or addressed in the Opening Sessions? Please indicate these below. (Use reverse side for additional space.)

6. What was most helpful about the Opening Sessions?

7. Please use the space below to provide additional comments concerning the adequacy, appropriateness, usefulness, or organization of the Opening Sessions.

Section II: Discussing Proficient Performance

The following statements seek your judgments about the discussions of Proficient performance as they relate to Missouri’s End-of-Course Assessments. Please circle the value on the scale under each statement that best characterizes your judgment.

8. The activities used to help operationalize Proficient performance were:

5	4	3	2	1
Very useful		Somewhat useful		Not useful

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	8	53%	5	36%	4	24%
4	4	27%	7	50%	9	53%
3	3	20%	2	14%	4	24%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.3	.82	4.2	.70	4.0	.71

*Percentages may not sum to 100% due to rounding.

9. By the end of the activity, my conception of Proficient performance was:

5	4	3	2	1
Very well formed	Moderately well formed		Not well formed	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (15)	%
5	9	60%	8	57%	7	41%
4	5	33%	3	21%	8	47%
3	1	7%	3	21%	2	12%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.5	.64	4.4	.84	4.3	.69

*Percentages may not sum to 100% due to rounding.

The following statement also seeks your judgments about the discussions of Proficient performance as they relate to Missouri’s End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

10. Please use the space below to provide additional comments concerning the activities around operationalizing Proficient performance for Missouri’s End-of-Course Assessments.

Section III: Discussing Basic Performance

The following statements seek your judgments about the discussions of Basic performance as they relate to Missouri’s End-of-Course Assessments. Please circle the value on the scale under each statement that best represents your judgment.

11. The activities used to help operationalize Basic performance were:

5	4	3	2	1
Very useful	Somewhat useful			Not useful

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	7	47%	5	36%	2	12%
4	6	40%	5	36%	10	59%
3	2	13%	4	29%	5	29%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.3	.72	4.1	.83	3.8	.64

*Percentages may not sum to 100% due to rounding.

12. By the end of this activity my conception of Basic performance was:

5	4	3	2	1
Very well formed	Moderately well formed			Not well formed

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	7	47%	6	43%	5	29%
4	7	47%	6	43%	9	53%
3	1	7%	2	14%	3	18%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.4	.63	4.3	.73	4.1	.70

*Percentages may not sum to 100% due to rounding.

The following statement also seeks your judgments about the discussions of Basic performance as they relate to Missouri’s End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

13. Please use the space below to provide additional comments concerning the activities around operationalizing Basic performance for Missouri’s End-of-Course Assessments.

Section IV: Discussing Advanced Performance

The following statements seek your judgments about the discussions of Advanced performance as they relate to Missouri’s End-of-Course Assessments. Please circle the value on the scale under each statement that best represents your judgment.

14. The activities used to help operationalize Advanced performance were:

5	4	3	2	1
Very useful	Somewhat useful		Not useful	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (16)**	%
5	6	40%	6	43%	3	19%
4	6	40%	7	50%	9	56%
3	3	20%	1	7%	4	25%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.2	.77	4.4	.63	3.9	.68

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

15. By the end of this activity my conception of Advanced performance was:

5	4	3	2	1
Very well formed	Moderately well formed		Not well formed	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (16)**	%
5	8	53%	7	50%	4	25%
4	5	33%	7	50%	10	63%
3	2	13%	0	0%	2	13%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.4	.74	4.5	.52	4.1	.62

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

The following statement also seeks your judgments about the discussions of Advanced performance as they relate to Missouri’s End-of-Course Assessments. Please write your responses to each prompt on the lines provided.

16. Please use the space below to provide additional comments concerning the activities around operationalizing Advanced performance for Missouri’s End-of-Course Assessments.

Section V: Item Rating Activities

The following statements seek your judgments about the item rating activities as they relate to the Missouri End-of-Course standard-setting meeting. Please circle the value on the scale under each statement that best represents your judgment.

17. Using the sample items to prepare for the actual item rating was:

5	4	3	2	1
Very helpful	Somewhat helpful			Not helpful

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	9	60%	6	43%	9	53%
4	4	27%	5	36%	4	24%
3	1	7%	1	7%	4	24%
2	1	7%	1	7%	0	0%
1	0	0%	1	7%	0	0%
Mean, <i>SD</i>	4.4	.91	4.0	1.24	4.3	.85

*Percentages may not sum to 100% due to rounding.

18. The explanation of the item data during the sample item portion of the training was:

5	4	3	2	1
Very helpful	Somewhat helpful		Not helpful	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	6	40%	8	57%	8	47%
4	7	47%	5	36%	3	18%
3	2	13%	1	7%	6	35%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.3	.70	4.5	.65	4.1	.93

*Percentages may not sum to 100% due to rounding.

19. The Item Rating Form was:

5	4	3	2	1
Very easy to use	Somewhat easy to use		Not at all easy to use	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	8	53%	9	64%	9	53%
4	6	40%	5	36%	7	41%
3	0	0%	0	14%	1	6%
2	1	7%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.4	.83	4.6	.50	4.5	.62

*Percentages may not sum to 100% due to rounding.

20. The information provided prior to each round of rating was:

5	4	3	2	1
Very useful	Somewhat useful		Not useful	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	10	67%	8	57%	8	47%
4	4	27%	5	36%	9	53%
3	1	7%	1	7%	0	0%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.6	.63	4.5	.65	4.5	.51

*Percentages may not sum to 100% due to rounding.

21. My level of understanding of the tasks I was to accomplish for each round was:

5	4	3	2	1
Very good	Acceptable		Very poor	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	14	93%	7	50%	13	76%
4	0	0%	6	43%	2	12%
3	1	7%	1	7%	2	12%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.9	.52	4.4	.65	4.6	.70

*Percentages may not sum to 100% due to rounding.

22. The amount of time I had to complete the tasks during each round was:

5	4	3	2	1
Far too long	About right		Far too short	

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (14)	%	Count (17)	%
5	2	13%	0	57%	3	18%
4	1	7%	5	36%	4	24%
3	12	80%	9	64%	10	59%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, <i>SD</i>	3.3	.72	3.4	.50	3.6	.80

*Percentages may not sum to 100% due to rounding.

The following statement seeks your judgments about the item rating activities as they relate to the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

23. Please use the space below to provide additional comments concerning the instructions and explanations you received, the adequacy of the time available, your levels of understanding of the process, or any other aspects of the estimates for the multiple-choice items. (Use reverse side for additional space.)

Section VI: The Overall Missouri End-of-Course Standard-Setting Meeting

The following statements seek your judgments about the overall processes and procedures used during the Missouri End-of-Course standard-setting meeting in which you participated as a panelist and the resulting recommended standards. Please circle the value on the scale under each statement that best represents your judgment.

24. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Proficient performance.

5	4	3	2	1
To a great extent	To some extent			Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (13)**	%	Count (17)	%
5	11	73%	10	77%	7	41%
4	3	20%	3	23%	9	53%
3	1	7%	0	14%	1	6%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, <i>SD</i>	4.7	.62	4.8	.44	4.4	.61

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

25. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Basic performance.

5 4 3 2 1

To a great extent To some extent Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (13)**	%	Count (17)	%
5	10	67%	9	69%	7	41%
4	4	27%	3	23%	10	59%
3	1	7%	1	8%	0	0%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.6	.63	4.6	.65	4.4	.51

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

26. I feel that this standard-setting meeting provided me an opportunity to use my best judgment in selecting and revising estimates for a recommended standard of Advanced performance.

5 4 3 2 1

To a great extent To some extent Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (13)**	%	Count (17)	%
5	11	73%	9	69%	7	41%
4	2	13%	4	31%	9	53%
3	2	13%	0	14%	1	6%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.6	.74	4.7	.48	4.4	.61

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

27. I believe that this standard-setting meeting has produced recommended cut scores that are defensible.

5	4	3	2	1
To a great extent	To some extent			Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (13)**	%	Count (17)	%
5	9	60%	10	77%	10	59%
4	5	33%	1	8%	6	35%
3	1	7%	2	15%	1	6%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.5	.64	4.6	.77	4.5	.62

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

28. I feel that this standard-setting meeting has produced recommended cut scores that would generally be considered as reasonable.

5	4	3	2	1
To a great extent	To some extent			Not at all

	Algebra I*		English II*		Biology*	
	Count (15)	%	Count (13)**	%	Count (17)	%
5	10	67%	8	62%	10	59%
4	4	27%	3	23%	6	35%
3	1	7%	2	15%	1	6%
2	0	0%	0	0%	0	0%
1	0	0%	0	0%	0	0%
Mean, SD	4.6	.63	4.5	.78	4.5	.62

*Percentages may not sum to 100% due to rounding.

**One panelist did not respond to this question.

The following statements seek your judgments about the overall processes and procedures used during the Missouri End-of-Course standard-setting meeting. Please write your responses to each prompt on the lines provided.

29. Please provide any comments you wish to share regarding the quality of assistance provided by the standard-setting staff.

30. Please provide any additional comments you wish to share regarding the overall meeting.

APPENDIX L: DATE RANGES FOR HAND SCORING OF PERFORMANCE EVENT/WRITING PROMPT ITEMS

Appendix Table 6.1: Hand Scoring Dates for the Spring 2008 Field Test

	English II	Algebra I	Biology
Team Leader Training	May 29–30, 2008	May 29, 2008	May 29, 2008
Scorer Training	June 9, 2008	June 2, 2008	June 2, 2008
Scoring Window	June 9–25, 2008	June 2–24, 2008	June 2–24, 2008

Training schedules for the Spring 2008 field test varied because they were scheduled based on the availability of the RIFs and the completion of the preparation of the original training materials.

Appendix Table 6.2: Hand Scoring Dates for the Fall 2008 Operational Test

	English II	Algebra I	Biology
Team Leader Training	February 2, 2009	February 2, 2009	February 2, 2009
Scorer Training	February 3, 2009	February 3, 2009	February 6, 2009
Scoring Window	February 3–17, 2009	February 3–13, 2009	February 3–13, 2009

ARC used this scoring process as an opportunity to identify potential team leaders for the Spring 2009 operational scoring. In addition to scoring the Fall 2008 booklets, these candidates learned to conduct training, use reports, handle personnel issues, and oversee other administrative duties for which they would be responsible.

Appendix Table 6.3: Hand Scoring Dates for the Summer 2009 Operational Test

	English II	Algebra I	Biology
Team Leader Training	April 21–23, 2009	April 15, 2009	April 22–23, 2009
Scorer Training	April 27, 2009	April 28, 2009	April 28, 2009
Scoring Window	April 28–June 5, 2009	April 28–June 8, 2009	April 28–June 4, 2009

Appendix Table 6.4: Hand Scoring Dates for the Fall 2009 Operational Test

	English II	Algebra I	Biology
Team Leader Training	April 21–23, 2009	April 15, 2009	April 22–23, 2009
Scorer Training	April 27, 2009	April 28, 2009	April 28, 2009
Scoring Window	April 28–June 5, 2009	April 28–June 8, 2009	April 28–June 4, 2009

Appendix Table 6.5: Hand Scoring Dates for the Spring 2010 Operational Test

	English II	Algebra I	Biology
Team Leader Training	April 21–23, 2009	April 15, 2009	April 22–23, 2009
Scorer Training	April 27, 2009	April 28, 2009	April 28, 2009
Scoring Window	April 28–June 5, 2009	April 28–June 8, 2009	April 28–June 4, 2009