Solicitation Information
18 Nov 05

RFP # B05828

TITLE: Multi-State Development of Custom Science Assessment

Submission Deadline: 13 January 06 @ 2:00 PM (EST)

Questions concerning this solicitation may be e-mailed to the Division of Purchases at scirfp@nciea.org. no later than **9 Dec 05 at 12:00 Noon (EST)** .Questions should be submitted in a *Microsoft Word attachment*. Please reference the RFP / LOI # on all correspondence. Questions received, if any, will be posted on the Internet as an addendum to this solicitation. It is the responsibility of all interested parties to download this information.

SURETY REQUIRED: No

BOND REQUIRED: No

Jerome D. Moynihan, C.P.M., CPPO
Administrator of Purchasing Systems

Vendors must register on-line at the State Purchasing Website at www.purchasing.ri.gov

Note to Vendors:

Offers received without the entire completed three-page RIVP Generated Bidder Certification Form attached may result in disqualification.

THIS PAGE IS NOT A BIDDER CERTIFICATION FORM

Request for Proposals

(NECAP, Tri-State Science Assessment)

The Rhode Island Department of Administration, Office of Purchases, on behalf of the Rhode Island Department of Education (RIDE) is requesting Proposals from qualified respondents for the development and administration of custom science assessments to meet federal and state requirements for the states of New Hampshire, Rhode Island and Vermont in accordance with the terms of this solicitation. A single contractor will be selected for all three states, and work will begin in February 2006 through December 2010. Each of the states will award a single contract. All of the requirements of the RFP are in the attached document including the scope of work, timelines, deliverables and requirements for the technical proposal.

This cover document stipulates the specific bidding requirements for Rhode Island, including the process for submitting the proposal.

Instructions and Notifications to Bidders:

- All respondents **MUST** register online at the RIVIP's Internet website @ ***http://www.purchasing.state.ri.us***. Proposals must be in accordance with the guidelines outlined in this request and the state's general conditions of purchased which can be accessed through the website.

- A fully completed and signed *RIVIP Bidder Certification Cover Sheet – All three pages should* accompany response submitted. Failure to make a complete submission inclusive of this three-page document may **result in disqualification**.

- Should there be a need for technical assistance in registering and/or downloading any document, call the RIVIP HELP DESK @ (401) 222-2142, ext. 134. Office Hours: 8:30 AM – 4:00 PM.

- All costs associated with developing or submitting documents in response to this Request and/or in providing oral or written clarification of its content shall be borne by the respondent. The State assumes no responsibility for these costs.

- It is intended that an award pursuant to this Request will be made to a prime respondent, who will assume responsibility for all aspects of the work.

- All pricing submitted will be considered to be ***firm and fixed*** unless otherwise indicated herein.

- Submissions in response to this solicitation are considered to be irrevocable for a period of not less than sixty (60) days following the established due date and may not be withdrawn without the express written permission of the State Purchasing Agent.

- Responses misdirected to other State locations or which otherwise are not received by the State Division of Purchases by the established due date for any cause will be determined to be late and may not be considered. The office clock, for the purpose of registering the arrival of a document, is in the reception area of the Department of Administration (DOA), Division of Purchases, One Capitol Hill, Providence, Rhode Island.

- Respondents are advised that all materials submitted to the State for consideration will be considered to be public records as defined in Title 38, Chapter 2 of Rhode Island General Laws, without exception, and will be released for inspection immediately upon request once an award is made.

- During the life of this contract, the State reserves the right to solicit separately for selected initiatives within this scope of work.

- In accordance with Title 7, Chapter 1.1-99 of the Rhode Island General Laws, Foreign corporations (a corporation established other than in Rhode Island) shall have the right to transact business in the State. *This is a requirement only of the selected vendor.*

- It is intended that an award pursuant to this Request will be made to a prime contractor, who will assume responsibility for all aspects of the work. Joint venture and cooperative proposals will not be considered, but subcontracts are permitted, provided that their use is clearly indicated in the offeror's proposal, and the subcontractor(s) proposed to be used are identified in the proposal.

- The State of Rhode Island has a goal of ten per cent (10%) participation by MBE's in all State procurements. For further information, visit the web site www.rimbe.org. To speak with an M.B.E. Officer, call (401) 222-6253.

- Interested parties are instructed to peruse the Division of Purchases web site on a regular basis, as additional information relating to this solicitation may be released in the form of an addendum to this RFP / LOI

- **Equal Employment Opportunity (RIGL 28-5.1)**
  § 28-5.1-1 Declaration of policy. – (a) Equal opportunity and affirmative action toward its achievement is the policy of all units of Rhode Island state government, including all public and quasi-public agencies, commissions, boards and authorities, and in the classified, unclassified, and non-classified services of state employment. This policy applies in all areas where the state dollar is spent, in employment, public service, grants and financial assistance, and in state licensing and regulation. For further information, contact the Rhode Island Equal Employment Opportunity Office, at (401) 222-3090

The State reserves the right to accept or reject any or all options, bids, proposals, to award on the basis of cost alone, and to act in its best interest.

At any point during the review process, any proposal found to be substantially non-responsive will be dropped from further consideration.

The State may, at its sole option, elect to require presentation(s) by respondents clearly in consideration for award. Other submissions, certifications, or affirmations may be required, as appropriate.

PROPOSAL SUBMISSION

All document pages are to be **numbered in consecutive order**.

Combined **TECHNICAL/COST PROPOSAL ("original" plus FOUR (4) copies)** submissions are to be either mailed or hand-delivered in a sealed envelope marked: ***"RFP B05828– Science Assessment*** by the date & time indicated on page one of this solicitation.

RI DEPT OF ADMINISTRATION
Division of Purchases, 2nd Floor
One Capitol Hill
Providence, RI 02908-5855

NOTE: Proposals misdirected to other State locations or which are otherwise not presented in the Division of Purchases by the scheduled due date and time will be determined to be late and may not be considered. The "official" time clock is located in the Division of Purchases Reception area.

New England Common Assessment Program
Tri-State Science Assessment
Request for Proposals

INTRODUCTION

Science Assessment Contracts

Through this request for proposals (RFP), the states of New Hampshire, Rhode Island, and Vermont are seeking a single contractor for the development and administration of custom science assessments to meet NCLB requirements and related regulations. Bidders may propose to assign specific contract tasks to qualified subcontractors and vendors, but the selected contractor will retain responsibility for all contracted activities.

The contract period for the science assessments described in this RFP will begin in February 2006 and continue through December 2010. Each of the states will award an individual contract. The initial contracts will be awarded for one or two years dependent upon what is allowable in each state, with options to renew or extend the contract for multiple subsequent years. All contracts are subject to continued availability of funds. Refer to each state's specific contract provisions for additional details.

Under the contracts awarded through this RFP, science tests will be administered each spring to students in grades 4, 8, and 11. The initial operational tests will be administered in May 2008 with subsequent administrations in May 2009 and May 2010. A pilot test will be administered across the three states in May 2007. Tests will be administered in a test administration window beginning approximately May 1st and continuing for three weeks.

Overview of the Science Assessment

All tests developed for science assessment will be based on the common science Assessment Targets developed jointly by the states. The Assessment Targets and additional detailed information related to the design of the science assessment are contained in the Science Test Specifications provided as an attachment to this RFP.

The science assessment will be designed to provide individual student performance data, information useful to teachers in planning instruction, and school-wide data on the effectiveness of the school/district science program in achieving the overarching goal of science literacy for all students. These statewide, on-demand assessments will serve as one component of a comprehensive, integrated local and statewide assessment system within each of the states. None of the states support the use of a single test for high-stakes decisions on student promotion or graduation.

The assessment will include a variety of item types: One testing session, with a longer extended response task will be devoted to assessing inquiry knowledge and skills.

The assessments will yield an overall science score translated into an Achievement Level (or Performance Level), based on standard setting and establishment of cut scores. In addition to a total science score (based on approximately 61-64 possible score points), subscores will be

produced for four reporting categories that include three domains of science (Life Science, Earth/Space Science, and Physical Science) and Inquiry. It is anticipated that each NECAP science test will have at least 7 to 10 test forms.

The science assessment will be designed for administration during three testing sessions of approximately 50 minutes each and include a variety of item types: multiple-choice items (1 point), short answer items (2 points), and constructed response items (3 or 4 points). Sessions 1 and 2 will consist of a combination of multiple-choice and 4-point constructed response items measuring the three domains of science: Earth/Space, Physical Science, and Life Science. Session 3 will be devoted to assessing Inquiry knowledge and will be preceded by a 30-minute, "hands-on" activity that may or may not be a scientific investigation, depending upon the grade level. Four Broad Areas of Inquiry (described in detail in Section VI of the NECAP Science Test Specifications) will be assessed in the third testing session: Formulating Questions/Hypotheses, Planning/Critiquing Investigations, Conducting Investigations, and Developing/Evaluating Explanations.

Common/Matrix Design

The assessment will consist of common and matrix-sampled items; however, only common items will be used to determine individual student scores. Matrix-sampled items included on the science tests will serve three purposes: embedded field testing of newly developed items, equating tests across years, and generating school/district/state level domain scores. The use of matrix-sampled items should provide a sufficient breadth/depth of items to adequately cover the assessment targets in each domain of science and produce reliable domain-level subscores. The states anticipate that 7 to 10 forms per grade level will be necessary.

Proposed Timeline for Major Contract Activities

Table 1 provides an outline of the major contract activities through the initial year and subsequent years of the contract. After the award recommendation, the states will work with the contractor to establish a specific timeline for activities necessary to develop and administer the Spring 2007 Pilot Test. The contractor's response should address any concerns with the proposed timeline and include suggestions for required modifications.

Table 1
Proposed Timeline of Major Activities During the Initial Contract Year

Date	Activity
February 2006	Award recommendation and contract negotiation
March-April 2006	Begin Item development for Spring 2007 Pilot Test
Fall 2006	Limited field testing of items and inquiry tasks, if necessary
May 2007	Spring 2007 Pilot Test
May 2008	Initial Operational Administration
August 2008	Standard Setting to establish achievement level cut scores.
September 2008	Reporting of results
May 2008-2010	2nd, 3rd, and 4th annual operational test administrations
August 2008-2010	Reporting of results from 2nd, 3rd, and 4th annual operational test administrations

Project Management

The states will assemble a common assessment project management team to manage the daily operation of joint activities under the assessment contract. The management team will include the state assessment directors as well as content and support staff from each state.

The assessment director of each state serves as the contact person to the contractor on issues unique to the state (e.g., number of schools, enrollments, shipping procedures).

The states will appoint a content development team consisting of one or two Department content specialists from each state. The states' content development team will interact closely with the contractor's item development team throughout the project and will be responsible for the review, evaluation, and approval of all items included on the NECAP science tests.

Inclusiveness of Assessments

The states are committed to the principle that that common statewide assessment must be accessible to virtually all students. Therefore, the contractor's response must reflect an understanding of and commitment to this principle throughout the item development, field-testing, test form construction, administration, and reporting processes. In particular, the contractor's response must address the principles of Universal Design as articulated in materials developed by the National Center for Educational Outcomes at the University of Minnesota (NCEO) and available at http://education.umn.edu/NCEO/TopicAreas/UnivDesign/UnivDesign_topic.htm.

The states are particularly concerned about the accessibility of the common assessments to English Language Learners. To meet state and federal inclusion requirements, the common assessments will be administered to students with widely varying English proficiency – from students with nascent proficiency to those students transitioning from ESL supports. The contractor's response must discuss how the development of items and test forms will address this issue. Specifically, the response should address the contractor's current activities and plans for future research. If the contractor's response proposes research to be conducted under this contract or includes additional cost items such as the development of alternate forms of test items, test sections, or test forms a separate cost estimate must be provided.

The development of large-scale assessments accessible to virtually all students is an emerging field. For example, the states anticipate that research conducted under the Enhanced Assessment Grant will provide valuable guidance to inform the development of accessible assessments. Throughout the course of the project, the contractor must demonstrate the desire and capacity to work with the states to efficiently integrate solid research findings into the design and development of the science tests.

Policies regarding the inclusiveness of the assessments also apply to any field testing and the 2007 Pilot Test. The pilot test will include students with disabilities and English Language Learners using the accommodations they would normally use in the large-scale assessment. There may be need for over sampling of these populations to ensure a valid field test. The

contractor will address these issues in their description of the field test, and will describe how the field test will guarantee validity for the sample of students with disabilities and English Language learners.

Use of Technology

The states are committed to the use of technology to improve the efficiency and accessibility of the common assessment program. Throughout their response, the contractor should provide specific examples of how technology will be applied to support the assessment program.

The states are particularly interested contractor's proposals for the integration of technology in the design and administration of the inquiry task.

The states will not consider responses that rely solely on the use of computer-administered tests. Contractor's responses that include the use of computer-administered tests as an option for states, districts, and/or schools must include a detailed description of the methods that will be used to ensure the security of the computer-administered tests and the comparability of results from paper-and-pencil and computer-administered tests.

The states will not consider responses that rely solely on technology for the scoring of individual student responses to constructed-response items. If applicable, contractor's responses must describe how technology will be integrated into the scoring of responses to constructed-response items.

The states also acknowledge that advances in technology and the increased availability of technology may result in substantive changes to the assessment program prior to the 2010 administration. Cost and schedule adjustments related to such changes will be negotiated as needed.

NECAP Background

The science assessments described in this RFP will be a discrete component of the New England Common Assessment Program (NECAP) – a joint state assessment program administered by the departments of education in New Hampshire, Rhode Island, and Vermont. NECAP began with a common need among the states to meet the increased testing demands of No Child Left Behind (NCLB), but quickly evolved into a shared vision of high standards and quality assessment. Collaborating to develop a common assessment expands the knowledge base and resources available to the states resulting in a higher quality assessment program than each state would be able to produce on its own.

At this time, NECAP tests based on common NECAP Grade Level Expectations developed jointly by the states are administered in October to students in grades 3 through 8. Reading and mathematics test are administered to student in all six grades. Additionally, a writing test is administered to students in grades 5 and 8. Current plans call for NECAP to include the following additional components:

- high school assessments in reading, mathematics, and writing to be administered to eleventh grade students in the fall beginning in the 2007-2008 school year; and

- science assessments at grades 4, 8, and 11 (addressed in this request) to be administered beginning in the spring of the 2007-2008 school year.

NECAP is one byproduct of a loosely coupled association of the departments of education in several New England states known as the New England Compact. Instituted in 2002 by commissioners of education in Maine, New Hampshire, Rhode Island, and Vermont, the New England Compact provides a forum for the states to explore idea, build a collective knowledge base, and establish cross-state activities that benefit each state. Their mission is to improve instruction and student learning through the sharing of information and pooling of resources. The states engage in a variety of formal and informal activities designed to support its mission. States participate in joint activities to varying degrees to meet their particular needs.

A major focus of the New England Compact has been assessment, in general, and the assessment of valid and reliable assessment of English language learners and students with disabilities, in particular. The New England Compact states have been awarded two Enhanced Assessment Grant under the No Child Left Behind Act (NCLB, Title VI, Subpart 1, Section 6112: Enhanced Assessment Instruments) with a major focus of improving access to assessment for all students through the effective design and delivery of fully inclusive assessments.

COMMON RESPONSE GUIDELINES

To as great an extent as possible, contractors will be asked to prepare a single staffing plan, budget, and description of work that will be submitted to each state in response to their particular RFP.

The contractor's response must include each of the following sections described in paragraphs A-F:

A. Introduction
B. Scope of Work
C. Staffing Plan
D. Budget
E. Corporate Capability
F. Required Appendices and Cover Materials

Introduction

The contractor's response will include an introduction that briefly describes the contractor's approach for completing the tasks required for this project and demonstrates the contractor's overall understanding of the required tasks and the needs of the states. This section of the contractor's response should also introduce any alternative methods or additional tasks that the contractor plans to propose to successfully complete this project.

Scope of Work

The contractor's response will describe the contractor's specific approach and plans for accomplishing the scope of work called for in the RFP. The response must provide sufficient detail to allow the states to evaluate the proposed methods.

The contractor must respond to each task described in the scope of work as well as provide descriptions of additional tasks that the contractor determines are necessary for the successful completion of the project.

If the contractor proposes alternatives or modifications to specific tasks described in this RFP, each task must be fully described and clearly identified.

Project Staffing

The contractor's response will include a staffing plan that includes the allocation of persons and/or departments by FTE across the major tasks to be completed. Individuals in key management and test development positions as well as any individual assigned to the project 0.5 FTE or more must be named in the contractor's response. The information will be provided in on the Task Allocation Forms provided.

The contractor's response will include descriptions of experience and resumes for all individuals proposed to fill key functions within this project.

Throughout the course of the project, the states retain the right of approval of individuals assigned to key management and test development positions within this project.

Budget

The contractor's response will include a single joint budget for all tasks described in this RFP. The portion of the budget assigned to each state will be determined during contract negotiations.

The contractor's response must include a detailed narrative describing the basis for costs in each of the major task areas.

To allow comparisons across contractors, budget information must be provided on the Budget Forms provided.

Contractors who propose alternative methods or additional tasks to those specified in the RFP must submit separate budget forms detailing the costs of the alternatives proposed.

Annual budgets should be based on work completed during fiscal years beginning July 1 and ending June 30.

Corporate Capability

The contractor's response must include a description of the corporate capability of the prime contractor and all proposed subcontractors that will be performing key functions on this project. Subcontractors include individuals and organizations performing tasks directly related to educational testing/measurement such as item development or standard setting as well as individuals and performing tasks such as printing and shipping.

Each corporate capability statement must address the contractor's qualifications, background, experience, and capacity to perform the tasks required for the successful completion of this project. The response should include descriptions of previous custom large-scale assessment work and similar work performed.

The contractor's response must include a list of any large-scale assessment projects in which the organization is currently or has been involved as a prime contractor or subcontractor since 2000. The list and description should include a short description of the responsibilities and outcomes, dates engaged, and total amount of contract as well as contact information for each project (i.e., contact name, affiliation, phone number, and email address).

Required Appendices and Cover Materials

The respondent must be prepared to provide budget and personnel detail information for Rhode Island using the required forms upon the execution of the final contract.

The cover materials also provide all the detailed requirements for submitting the RFP in Rhode Island.

Evaluation Criteria

Contractor responses to this RFP will be evaluated according to the criteria contained in Table 2.

Table 2
Proposal Evaluation Criteria

Criterion	Points
Quality of Response - responsive to the RFP - demonstrated understanding of issues related to states' goals, custom large-scale assessment, high-stakes testing - technical soundness of proposed methods - innovativeness of proposed methods in terms of technical quality, efficiency, and cost effectiveness	40
Adequacy of Staffing - background and experience of proposed staff - as demonstrated in track record of meeting deadlines, delivering within budget, continuity of key project management staff, handling and solving problems – particularly in projects directly related to custom statewide large-scale assessments - sufficiency of allocated time and number of people	25
Adequacy of Physical Resources - sufficiency/availability of allocated resources - background and experience of proposed vendors (e.g., printers, overnight shipping)	10
Cost - reasonableness of cost in relation to proposed activities to meet contract requirements - reasonableness of cost in relation to states' budget.	25
Total	100

Liquidated Damages/Penalties

The final contracts negotiated under this contract will include a provision for penalties or liquidated damages due to non-performance or breach of contract. In particular, penalties or liquidated damages will be tied primarily to actions on the part of the contractor that result in the either the late delivery of test materials or reports, or the delivery of inaccurate test materials or reports. Specifics of the penalties and liquidated damages will be determined during contract negotiations. As a starting point for negotiations, the states propose a policy in which the contractor shall be penalized no more than a fixed percentage (e.g., 7.5%) of the total contracted amount in a given year. The 7.5% maximum penalty shall be prorated against the number of days in which the contractor is determined to be in non-compliance with the contract (e.g., late delivery of reports to schools).

APPLICATION PROCEDURES

Issues concerning application procedures and a review schedule are addressed in this section of the RFP. Any changes to these procedures will be posted by each state**.** In addition, contractors who submit an intent to bid by the date specified will be informed of any changes to these procedures via e-mail.

Bidders' Conference

No bidders' conference will be conducted for this RFP. Interested parties may ask questions in accordance with the terms and conditions expressed on page one of this solicitation.

Questions Concerning the RFP

The states will accept written questions via e-mail through December 9, 2005. All questions should be directed, as per the instructions on page one of this solicitation, to the following address: scirfp@nciea.org.

Questions and responses will be posted to the RI Division of Purchases website (www.purchasing.ri.gov) website no later than December 16, 2005. Contractors may submit follow-up questions via e-mail (not new questions) by December 19, 2005. The state of Rhode Island will post follow-up questions and responses by December 22, 2005.

Intent to Bid

Contractors must submit a statement of intent to bid to each state via e-mail by December 9, 2005. Statements of intent to bid on this RFP must be submitted to: Ms. Cynthia Corbridge, Rhode Island Department of Education, 255 Westminster St., Providence, RI 02903. The email address is cynthia.corbridge@ride.ri.gov

Submission Requirements

Contractors are required to submit hard copies of all materials and forms requested in this RFP. Contractors may also submit an electronic copy of materials at their discretion. Contractors may submit hard copy or electronic copies of supplemental materials and/or work samples submitted to support their proposal.

Contractors must submit original and 4 copies of all materials and forms requested in this RFP to each state.

Proposal Deadline

All proposal materials must be received by the date and time indicated on page one of this solicitation. Proposals not received by all states by this deadline will not be considered.

Proposal submission in Rhode Island must follow the requirements outlined in the cover document to this RFP.

Planned Review Schedule

The states plan to review proposals and make an award recommendation according to the following schedule. Unforeseen circumstances may result in changes to the schedule. Bidders will be informed of any schedule changes via e-mail.

The states will review proposals and identify clarifying questions for contractors by January 24, 2006. Written responses to questions must be provided by 10:30 a.m. EST January 30, 2006.

The states may choose to interview particular bidders prior to making an award recommendation. Interviews would be conducted on January 31, 2006 at a location to be determined in one of the states. Interviews will be scheduled by January 24, 2006. Contractors would be represented at the interview by the proposed project director, psychometrician, and senior staff as requested.

The states will make an award recommendation by February 10, 2006.

SCOPE OF WORK

This section of the RFP contains a description of the major tasks required of the contractor for the successful completion of this project and provides information on contract deliverables. The contractor's response must directly reference and address each of the tasks contained in this section as well as addressing the requirements discussed in the Test Specifications documents and Section 1 (Introduction) of this RFP. In addition, the contractor's response must identify any additional tasks not included in this RFP that the contractor determines are necessary for the successful completion of this project. Contractors may also wish to propose alternative or additional tasks that they feel would improve the efficiency of the project and/or quality of the materials produced for the project.

The quality of all work and materials produced by the contractor is critical to the successful completion of the Common Assessment Project. Consequently, there is no single 'quality control' task included the scope of work for this RFP. Throughout their response, the contractor must provide evidence and descriptions of the methods and procedures they use to ensure the quality of their work.

Additionally, technical documentation is a critical requirement to verify the quality of work and provide evidence for the validity of the assessments. In addition to the technical reports and publications specifically described in this RFP, the contractor is expected to provide appropriate technical documentation for tasks such as item development, test construction, scoring, etc. on an ongoing basis.

All electronic and hard copy materials developed for this project, including test items not used on operational test forms, are the sole property of the states and will not by copyrighted, resold, or reused by the contractor.

Project Management and Planning

1. Management Team

a) Project Director – The contractor will appoint a single project director who oversees the management of the project and serves as primary point of contact with the states' project director.

b) Project Manager(s) – The contractor will appoint a project manager(s) who serve as primary point of contact with states on issues unique to the state (e.g., shipping, identification of schools)

c) Management Meetings – The contractor will support regular management meetings with the states' project management team. Monthly project management meetings will be held in New England (reduced to bimonthly meetings after the initial contract year). The contractor should budget for eight one-day meetings and four two-day meetings in the initial year and six two-day meetings in subsequent years. The contractor may propose options such as video-conferencing to increase efficiency and reduce costs.

d) Conference Calls – The contractor will support regular conference calls with the states' project management team.

e) Management Reports – The contractor will provide the following reports:

(1) Weekly written project status reports

(2) Monthly Budget update reports

(3) Annual project plan and schedule (including detailed procedures and specifications)

(4) Minutes of all meetings and conference calls

2. Technical and Policy Issues

a) The contractor will attend semi-annual two-day meetings of the states' joint common assessment Technical Advisory Committee (as requested). The contractor will be represented at the meeting by the project director, project lead psychometrician, and one additional staff member as necessary.

b) The contractor will attend selected joint meetings of the state assessment directors (and deputies) upon request up to two times per year. The contractor will be represented by the project director and senior management.

B. Item Development

The contractor's response must address the item development requirements described in the Test Specifications as well as the specific tasks included in this section of the RFP. In particular, the contractor's response must include a description of the procedures that will be used to facilitate interactions among the contractor's item development team and the states' content specialists.

Also, the contractor's response must discuss the procedures that will be used to facilitate the work of committees consisting of educators from three states to ensure that all items are accurate in content, aligned with the assessment targets, free from bias, and accessible to the widest possible range of students. The contractor's response must address any concerns with particular requirements included in the Test Specifications.

1. Item Development Team

a) The contractor will appoint an item development team responsible for the development of items for the science tests at grades 4, 8, and 11. The team will include an overall lead developer and a developer assigned to each grade level. The item development team will also include a specialist(s) in students with disabilities and English language learners. The team will include sufficient staff to develop the required test items and represent the contractor with the states and committees.

b) Content specialists from each state Department of Education will form the states' development team. The states' content team will serve as the primary contact with the contractor's development team, will have input into the design and format of test items, and will be responsible for the review and evaluation of all test items developed. The contractor's response must discuss methods for efficiently coordinating and supporting interactions among the state content specialists and the contractor's item development team.

2. Item Review Committees

a) The contractor will support item review committees consisting of primarily of grade-appropriate teachers recruited and selected by each of the states. Additional committee members may include local curriculum coordinators, content specialists, and ESL or special education specialists. The states will determine the composition of committees. The contractor should plan on supporting the item review committees with the following specifications:

- There will be a committee of 18 members at each grade level.
- The committee will meet three times per year beginning in the spring of 2006. The contractor should plan on two meetings during the school year (early fall, early winter) and one meeting during the summer.
- The meetings will be led by the states' development team and facilitated by the contractor's development team.
- Meetings during the school year will be scheduled for two days. Summer meetings will be scheduled for three days.
- Members will be paid a stipend of $500 for participation in the summer meeting. Members' school districts will be paid a substitute reimbursement for meetings during the school year. The contractor should budget for $100 substitute reimbursement per member per day for meetings held during the school year.
- The contractors will support and arrange for lodging for committee members, states' content specialists, and the states' project management team. The contractor will also be responsible for travel expenses to attend the meeting.
- The contractor will provide breakfast and lunch each day of the meeting and be responsible for dinner expenses ($20 per day) on days which require an overnight stay.
- Meetings will be held at a hotel, conference center, or similar suitable location in one of the states. The meeting location will rotate among the states.

b) The contractor will ensure that the states' test development team has sufficient time to review and provide feedback on all materials and items prepared for the Item Review Committee meetings.

c) The contractor will schedule a half-day following the conclusion of each Item Review Committee meeting for a meeting of the contractor's and states' development teams.

d) The contractor will produce a written report documenting the meeting within two weeks of each committee meeting.

3. Bias/Sensitivity Review Committees

a) The contractor will support a bias review committee consisting of external educators selected by the states to review the content of passages, other stimuli, and test items for potential bias.

- There will be a single committee across grades 4, 8, and 11.
- The committee will contain 12 members.
- Travel, lodging, and meals for committee members and states' project management team will be arranged and paid by the contractor.
- Members will receive a stipend of $300 for meetings held during the summer.
- Assuming that six of the committee members may be teachers, the contractor should budget $100 per day for substitute reimbursement paid to the members' school district for meetings held on school days.

b) The Bias/Sensitivity Review Committee will meet 2 times per development cycle for two-day meetings. They will focus on review of stimuli proposed for the development of new field test items, review of newly developed items recommended for field test, and review of items recommended for inclusion in the operational item bank that have been flagged for DIF.

c) Bias/Sensitivity Committee meetings will be led by the states' project management team and facilitated by the contractor.

d) If feasible, the Bias/Sensitivity Review Committee meetings may be scheduled concurrently with the Item Review Committee. The contractor's response should propose a process that will help avoid the issue of having assessment items go through development and be flagged for bias/sensitivity (content review, not empirical DIF analyses) only after substantial investment in development effort.

e) The contractor will produce a written report documenting the committee meeting within two weeks of each meeting.

4. Content Review

a) The contractor must ensure that the content of all items recommended for field-testing is accurate and reflects the current state of knowledge in the appropriate field. The contractor's response must describe their methods and procedures for meeting this requirement within the item development process.

b) The contractor will produce a document outlining its incorporation of accessibility at all stages of test construction and administration, including, but not limited to, physical test design, item development, field-testing, administration, and reporting.

5. Number of Items

a) In Year 1, the contractor will develop an adequate number of items to populate forms for the Spring 2007 Pilot Test according to the requirements provided in the Test Specifications.

b) The contractor's plans for the Spring 2007 Pilot Test should yield sufficient items to build the initial operational test, practice tests, and begin development of a breach form.

c) In subsequent years, the contractor will develop an adequate number of items to support the release of 25% of the test items and support the rotation of non-released common items over a three year period.

d) The contractor will develop sufficient adequate items so that Item Review Committees and the states' Test Development team can have a choice of items to include on the operational forms of the tests. The contractor's response must include a discussion of the expected yield of items at each stage of the item development process (e.g. initial development, after item committee review, after field testing).

6. Item Bank

a) The contractor will develop and deliver to the states an item bank of all items developed for the science assessments. The item bank will include a database that provides electronic access to each item (text and graphics) as well as pertinent information for each item, including history (placement, item statistics for all administrations of the item, editing, and context).

C. Test Construction & Test Materials

1. Content of Test Forms

a) The contractor will support meetings of the states' test development team and the contractor's test development team to select items to be included on the pilot test and operational test forms.

b) The selection and ordering of items on the test forms will be based on appropriate psychometric procedures and will meet the requirements of the Test Specifications. The contractor's response must include a description of the proposed process for item selection.

c) The states will have final approval of the selection of items and test forms.

d) The contractor's response must include a discussion of an efficient procedure for cycles of item and test form review between the contractor and states' project team.

2. Test Booklets

a) The contractor should base estimates of the quantities of test booklets required on the information provided on the number of students enrolled at each grade. Estimates should include 10% overage. (Note that while the 10% figure is used as the basis for estimates throughout the RFP, very small schools will receive a fixed number of extra materials.)

b) The contractor will produce a single test booklet containing questions for sessions 1, 2, and 3. The contractor should propose whether to produce a single test booklet for sessions 1, 2, and 3 or to produce a separate test booklet for session 3 (Inquiry). The contractor's response should include a discussion of issues such as cost, security, and ease of administration.

c) The contractor will produce a separate booklet to be used by students during the hands-on portion of the session 3 Inquiry Task. The booklet will include any necessary descriptions of the task and required procedures, and also include space for students to record notes and collect data that will be used to respond to test items. This booklet will not be processed for scoring.

d) The format and layout of the test booklets will meet the requirements of a style guide agreed to by the states and contractor during the initial months of the contract. The contractor will support a one-day style guide meeting. A central component of the style guide will be the application of "universal design" principles and procedures in areas such as the design and layout of the booklet, use of graphics, and format of directions to ensure access by the broadest possible population of students. The contractor's response must address methods and procedures used to inform test booklet design.

e) The test booklets will be uniform across the three states – including the use of a common cover page identifying the assessment. The cover page of the test booklet should be printed in color to match the answer documents at the appropriate grade.

f) The contractor will produce large-print versions of test booklets and related test materials (one test form per grade level). The contractor's response should comment on the research and best practice for providing accommodations for visually impaired students, particularly the issue of multiple sizes of large-print versions.

g) The contractor will produce Braille versions of test booklets and related test materials (one test form per grade level). The contractor should budget for the production of 10 sets of Braille materials per grade.

h) As a separate cost item, the contractor should budget for the provision of a Spanish version of the tests (including ancillary materials) for Rhode Island. The contractor's response should discuss issues involved in the various approaches to accommodating English language learners (e.g., translated, side-by-side) and recommend an approach.

3. Answer Documents

a) The contractor should base estimates of the quantities of answer documents required on the information provided on the number of students enrolled at each grade. Estimates should include 10% overage.

b) All answer documents will be custom-designed for the science tests. The color of the answer documents at each grade should match the color of the corresponding test booklets

c) A single stand-alone answer document will be used for all sessions at grades 4, 8, and 11.

d) A generic answer sheet will be used across matrix-sampled test forms. A light background grid and light background lines will be printed on the constructed response areas. The contractor's response should suggest ways to help students navigate through the test and answer document to reduce errors in where students put their answers (e.g., periodic icons to link the item with a position in the answer booklet). The contractor's response should discuss methods used to collect test form information that minimize errors in processing.

e) The states wish to collect a limited amount of student demographic and program information through the answer document. The elements will be uniform across states. Some information may be provided by students during the testing and other information may be provided by the test administrator or coordinator following testing.

f) The states wish to collect student survey information at each grade level through the student answer document. The student survey will be contain common items across the states but may also contain items unique to each state. The contractor's response must discuss options for collecting customized student survey responses through a generic answer document and/or provide options for other methods of collecting survey information that would allow the linking of student-level survey responses and test results.

4. Student Labels

a) The states will provide the contractor with data files containing student identification, demographic, and program information by February 1st of each year. The contractor will use that information to produce student identification labels that will be affixed to student answer booklets during testing. The pre-printed student identification labels will reduce the amount of information schools need to provide for each student on the answer document and increase the accuracy of information collected. At a minimum, one student label per answer document will be required. The contractor's response should discuss methods that will be used to capture information from the student label and link it to the appropriate student information.

b) The contractor's response should discuss alternatives to student labels for accurately identifying student booklets. The response should also discuss options for collecting information to generate student labels after February 1st for schools with high mobility rates as well as options for schools to generate labels on-demand for newly enrolled students.

5. Ancillary Materials

The following ancillary materials will be produced for the science tests. The contractor's response should discuss the type of information included in manuals, the type and use of shipping labels and control forms, etc. Contractors should base estimates of quantities of materials on information provided on the number of students, schools, districts, etc. in each state. Estimates should include a 10% overage for all materials. Electronic versions of all ancillary materials should be available for posting on the states' websites. Any items for which only an electronic version is necessary will be clearly indicated.

a) The contractor must budget for the development of a student survey to be administered during testing to students at grades 4, 8, and 11. The survey may contain up to two pages of common questions and a single page of state-specific questions.

b) The contractor must budget for the development and administration of a staff survey to be administered to teachers, specialists, and other instructional staff at grades 4, 8, and 11. The survey may contain up to two pages of common questions and a single page of state-specific questions. The contractor's response should discuss options for the development and administration of an online survey.

c) The contractor must budget for the development and administration of a survey to be administered to the principal/test coordinator in each tested school. The survey may contain up to two pages of common questions and a single page of state-specific questions. The contractor's response should discuss options for the development and administration of an online survey.

d) The contractor will produce a Principal/Test Coordinator manual for each administration. A common manual will be produced for all grades. A single printed manual will be shipped to each district and school and copies of the manual will be distributed at the administration workshops.

e) The contractor will produce a unique Test Administrator manual for each grade level test. A single copy of the manual will be provided for every 12 students enrolled.

f) The contractor will develop a unique practice test booklet, answer document, and supporting scoring materials for each grade level. The practice test should include two sessions. Session 1 should include items from each of the three domains of science. Session 2 should provide a sample of an extended Inquiry task. Each session should include the same amount of items and mix of item types as an operational test at that grade level. Only one set of practice tests will be developed during the course of the contract – following the Pilot Test and prior to the initial operational test. Only electronic versions of the practice test are needed.

g) The contractor will produce a single-sheet reference sheet at each grade level that may contain formulas, diagrams, tables, etc.

h) The contractor will produce all forms and labels necessary for the efficient and secure shipment and receipt of materials.

i) The contractor will produce all control/processing forms necessary for the administration of the science tests.

j) The contractor will produce all forms sign-off forms necessary to ensure the security of the test materials including a form to collect principal certification of proper test administration.

k) Each year the contractor will support the states in the production of up to three reports related to science tests on issues such as test design, administration, interpretation/use of results, scoring, and validity/reliability. The intended audience for these reports will be educators or the general public. The states will determine the topics for each report. Only electronic versions of these reports are needed.

D. Administration

1. Identifying districts and schools and grade level counts

a) Each Department will provide the contractor with an updated database of districts and schools participating in the assessment by February 1st each year.

b) The contractor will propose a web-based system for districts to enter information such as enrollment by grade/school, and to confirm information provided by the state such as contact information, and grade configurations. The system should have appropriate levels for viewing and changing information, and have appropriate security. Changes should have confirmation notices sent to the responsible party in the school/district, and a summary accessible by the state.

c) The contractor will propose a system for schools to order special test materials (e.g., large-print, Braille) prior to testing.

d) Each state will provide the contractor with statewide updated projected grade level enrollments by October 1st of each year.

e) The contractor will be responsible for communications with the schools regarding verification of enrollment counts.

2. Shipping Requirements

a) The contractor will ship test materials directly to schools. The contractor will be responsible for all communications with the schools regarding shipping/receiving. The contractor's response must describe the proposed shipping method and contractor.

b) Test materials will arrive in schools in a 2-day window 7-10 working days before the first day of testing.

c) The contractor's response must include a description of procedures to deliver additional materials in a manner that does not delay test administration to schools that receive incomplete shipments or do not receive shipments.

d) Notification/Tracking: The contractor will notify schools via e-mail of all shipments. Schools will be able to track shipments online.

e) The contractor will notify the state of shipment/delivery of all materials and provide updates on the status of undelivered materials.

f) The contractor shall keep a log of complaints and issues, how they were resolved, and an indication of customer satisfaction. The log shall be viewable by the state on demand. The contractor's response should include a description of options for creating a log that include the use of technology.

g) The contractor will pay for the return shipment of testing materials from the schools. Schools will ship all secure materials directly to the contractor following testing. Schools will be able to track shipments online. The contractor's response must describe the proposed method of shipping.

h) The contractor will propose a method for accounting for the return of all secure testing materials. The contractor's response must include a description of methods and procedures used to track shipments from schools and follow-up with schools that have not returned materials.

i) The contractor will notify each of the states of the status of the return of all secure test materials. The contractor's response must include a description of the procedures used to gather information and anticipated timeline for providing the information. The contractor's response must describe the procedures that will be followed when materials are not returned.

3. Support

a) The contractor will prepare materials for and support the administration of workshops in each state approximately one month prior to administration. Department and contractor staff will conduct the workshops. At each workshop, the contractor will be responsible for on-site registration, distribution of materials, and related tasks. The contractor should budget for 5 half-day workshops per state to be held at five sites per state. Estimates of total participants should be based on two participants per school and an additional two participants per district.

b) The contractor will provide toll-free telephone support to schools during the administration period. The contractor's response should discuss options for staffing the support center, training support personnel, and duration of support. The contractor's response must also discuss procedures for ensuring that efficient service is provided during peak times as well as contingencies for providing support in the event of a breakdown in telephone service. At a minimum, the contractor's response must include providing telephone support for a period beginning one month prior to test administration and continuing two weeks following the end of the testing window.

E. Scanning/Imaging and Scoring

1. Scanning/Imaging

a) The contractor is responsible for the efficient, accurate, and reliable scanning and/or imaging of all student responses, including student responses to multiple-choice test items and questionnaire items, student responses to constructed-response items, student identification and demographic information collected through a student label, and student demographic information provided by the student and/or school principal. The contractor must provide details regarding

the accuracy and reliability of the scanning technology system. In addition, the contractor is responsible for scanning or imaging all ancillary materials, as appropriate.

b) The contractor will capture images of all student responses to constructed-response items and store those images so that they can be efficiently linked to and retrieved on the basis of student and school identification information, scores, and item information.

c) The contractor will demonstrate that programs have been prepared to accurately scan and image all test materials.

d) The contractor will ensure that the scanning database is error-free and contains valid responses in all fields.

e) The contractor will provide the states' project management with a detailed report describing any materials that could not be scanned due to damage caused by the school, contractor or other reasons.

2. Multiple-choice items

a) All multiple-choice items are machine-scored. The contractor's response must include a description of the methods used to ensure and verify that the proper key has been used to score multiple-choice items.

b) The contractor will provide each state with a report documenting irregular responses such as blank answer documents, excessive item non-response, and excessive multiple marks at the district and school levels. The states and contractor will determine levels of excessive non-response and multiple marks, and other indicators of irregular response.

3. Constructed-response items

a) Qualifications of scorers – At a minimum, all scorers must have at least two years of college credit in science courses directly related to the assessment targets addressed by the science tests. A bachelor's degree in a related field is preferred and required for at least 50% of scorers. The states encourage the use of current or retired teachers as scorers. The contractor's response must include a description of the recruiting procedures, qualifications, and experience of the scorers proposed for the science tests. The contractor's response must include a discussion of the qualifications necessary to score responses to the inquiry task.

b) The contractor's response must include a description of the scoring procedures and development of scoring materials at all stages of the scoring process including, but not limited to, the selection of anchor papers, the development of qualifying and training materials, and the development of calibration sets. The description must include an explanation of the roles of the contractor's item development staff, the contractor's scoring staff, and the states' content team.

c) The states' project management team will have final approval of all materials prepared for scoring.

d) Training/Qualification of scorers – The contractor's response will provide details on the processes and procedures used to train scorers and qualify scorers for participation in the scoring of the science tests.

e) Scoring Sites – The contractor's response must include detailed information about scoring sites that are proposed for scoring the science tests. The response must include a discussion of processes and procedures used to ensure consistency in scoring across sites and to minimize the impact of potential differences in scoring sites on results.

f) Consistency Across Years – The contractor's response will describe the procedures used to ensure, monitor, and verify the consistency of scoring across years.

g) Monitoring scorers – The contractor's response will provide details on the quality control processes used to monitor scoring rates and accuracy. The response will also provide details on processes and criteria used to identify scorers for retraining or removal and processes used to invalidate scores produced by particular scorers. This should include rates of double-scoring, selection of responses for double scoring, etc.

h) Information available from scoring – The contractor's response will describe the type and frequency of information available to be provided from the scoring process (within and across scoring sites) to the states. The states and contractor will determine the requirements for providing updated information during scoring.

i) The states' project management team shall have the right to request, "on-demand" within four hours any regular scoring report, and to do unannounced site visits to scoring centers.

j) The contractor will provide support for the states' content team or designated representatives from the states to be present during the selection of anchors, scoring qualification materials, and training materials. The contractor should budget for a two-day meeting per grade level.

k) The contractor will provide support for the states' project management team or designated representatives from the states to observe scoring. The contractor should budget for three persons to visit two days per scoring site.

l) The contractor will produce a document summarizing the scoring process for the current year that includes information described in tasks E-3-a-j, as appropriate. The states and contractor will determine the content of the document. Additional information regarding scoring will also be included in the Technical Report (G-1-j).

F. Analysis

The contractor is responsible for conducting all analyses necessary to report student, school, district, and state results from the common assessment program and to ensure that tests meet the standards of technical quality. During each year of the contract, the contractor will conduct analyses necessary to support test development, test construction, scoring, and standard-setting and validation activities. In addition, the contractor will conduct secondary analyses related to security, data interpretation, policy formation, and administrative planning.

1. Calibration and Scaling

a) The contractor will calibrate test items using an appropriate item-response theory model(s). The contractor's response must include a discussion of the benefits of the proposed IRT model, its appropriateness for the science tests, and indicate which software will be used.

b) The contractor will translate overall student scores on common items to a reporting scale developed for each grade level science test. The contractor's response must discuss alternative methods for creating a reporting score scale consistent with the reporting requirements.

c) The contractor will propose a method for developing scales to report subscore results at the domain level (physical science, earth science, life science, inquiry) at the school, district, and state levels. Subscore results will include both common and matrix-sampled operational test items. The contractor's response must include a description of the proposed method and a rationale for its use.

2. Equating

a) The contractor will design and conduct analyses required to equate the science tests from year to year at each grade level. The contractor's response must describe the proposed method for equating the tests and provide a rationale for the proposed method.

b) The contractor will design and conduct analyses required to calibrate and equate test items across matrix-sampled test forms within a single year. The contractor's response must demonstrate an understanding of the test design and describe the method proposed for accomplishing this task.

3. Item Evaluation

a) Field-test items. The contractor will produce item statistics for all field test items. The contractor's response must include a description of the item statistics that should be generated to assist in the evaluation of field test items including a discussion of the appropriate statistics for multiple-choice and constructed-response items.

b) Operational items (including common and matrix-sampled items). The contractor will produce item statistics for all operational items. The contractor's response must include a description of the item statistics that should be generated to assist in the evaluation of these items including a discussion of the appropriate statistics for multiple-choice and constructed-response items.

4. Test Construction

a) The contractor will conduct analyses to support the construction of technically sound test forms. The contractor's response must include a description of the types of analyses that will be conducted and how the results of those analyses will be disseminated and used by appropriate state and contractor staff to assist in test construction.

5. Scoring

a) In addition to the analyses conducted during scoring (section E-3) to monitor the scoring process, the contractor will conduct additional analyses after scoring to verify the accuracy of scoring. The contractor's response must include a description of the types of analyses that will be conducted and how the results of those analyses will be disseminated and used.

6. Reporting

a) The contractor will design and conduct all analyses necessary to produce student, school, district, and state results and other information included in published reports of results. Reporting specifications and requirements are described in the Test Specifications and in Section G of this RFP.

7. Additional Analyses

a) The contractor will conduct ad hoc analyses to investigate questions related to the validity and reliability of the assessments for their intended uses. The contractors should budget $25,000 per year for this task.

b) The contractor will support external validity studies commissioned by the states. The contractor should budget $50,000 per year for this task. The contractor's response must include a discussion of the types of validity studies that should be conducted to establish the validity and credibility of the science tests at various points in the program. Issues of validity in relation to accessibility and/or accommodations should be addressed in the contractor's response.

c) The contractor will support external alignment studies commissioned by the states. The contractor should budget $25,000 per year for this task.

G. Reporting

The contractor is responsible for the reporting of results of the science tests according to the schedule specified in Section 1-C of this RFP. Specific reporting dates that meet those guidelines for each administration will be agreed upon prior July 1st of the preceding year to facilitate project planning through the establishment of intermediate milestones that include, but are not limited to, a) the completion of scoring and processing, b) the development, review, and approval of reporting specifications, and report shells, c) the review and approval of equating procedures and analyses, d) the delivery, review, and approval of preliminary data files, and e) the delivery, review, and approval of sample reports. Penalties for failing to meet final reporting dates and intermediate milestones will be negotiated.

1. Reporting of Results

a) The contractor will provide the reports listed below for each grade level science test. With the exception of the Parent/Guardian Report (which will be a paper report), all reports of results will be delivered in electronic format. The contractor's response must include a detailed description of a proposed method for electronic reporting that provides easy access to results while ensuring

security and confidentiality. The electronic reporting system must enable state access to district and school reports and district access to appropriate school reports.

- Parent/Guardian Report for individual students containing achievement level results, scaled score, etc. (2 paper copies per student)
- Classroom Roster providing individual student-level results at the classroom level. Classroom Roster may also include item-level results for released items. (electronic report and database)
- Classroom Summary aggregating results from the Classroom Roster. May include school, district, and state comparisons (electronic report and database)
- School Report Package containing whole school achievement level results, subgroup results as required by NCLB and subscore results as specified in the Test Specifications. May also include selected results from the student questionnaire, released items, district and state comparisons, and comparisons with previous years. (electronic report and database)
- District Report Package containing whole district achievement level results, subgroup results as required by NCLB, and subscore results as specified in the Test Specifications. May also include selected results from the student questionnaire, released items, state comparisons, and comparisons with previous years. (electronic report and database)
- State Report Package containing statewide achievement level results, subgroup results as required by NCLB and subscore results as specified in the Test Specifications. May also include selected results from the student questionnaire, released items, and comparisons with previous years (electronic report and database).
- District confidential student-level database containing information such as school identifying information, student identifying information, demographic information, raw item responses for released items, questionnaire responses, raw score totals, scaled scores, and performance level.
- Domain Subscore reports at the school, district, and state levels providing domain results based on the expanded common and matrix-sampled items.

b) The contractor will deliver the Parent/Guardian reports to the district's central office for distribution to the appropriate school.
c) The contractor's response must describe how district and school staff will be able to securely access electronic reports and data.
d) The contractor will provide each state with copies of each set of electronic reports described in section G-1-a for archiving.
e) The contractor will provide each state with a database containing the aggregated school, district, and state results provided in the electronic reports described in section G-1-a.
f) The contractor will provide each state with a confidential student-level database containing all available student-level information for students in the state such as student name and identifying information, demographic/program information, test form, raw item responses, scored item responses, accommodation information, questionnaire responses, raw score totals, scaled scores, and performance levels.
g) The contractor will provide each state with a non-confidential student-level database containing selected information from the complete student records described in section G-1-f for all students participating in the science tests across the three states.
h) The contractor will develop and produce interpretive materials for parents and schools/districts. The interpretive materials will be provided in electronic format for posting

on the states' websites. The contractor's response must include a description of the type of information to be included in such materials and methods to increase the usefulness of such materials.

i) The contractor will develop materials related to the release of a sample of common items following each test administration. The materials will include test items, item documentation mapping items to assessment targets, scoring materials, and sample student work. The materials will be provided in electronic format for posting on the states' websites. The contractor's response must include a description of the type of information to be included in such materials and methods to increase the usefulness of such materials.

j) The contractor will develop and produce an annual Technical Report that documents and provides the necessary evidence to demonstrate that the each of the common assessments and the set of common assessments as a whole serve their intended purposes, are aligned with the test blueprint, fulfill the Test Specifications (including accessibility criteria), and meet accepted professional standards for educational testing. The states and contractors will negotiate the table of contents and format for the Technical Report with input from the states' joint Technical Advisory Committee. The annual Technical Report will not replace or fulfill the requirement for ongoing technical documentation or documentation specified in other tasks. The final draft of the document will be delivered to the states no later than three months following the release of assessment results in operational test years or three months following the completion of scoring in the pilot test year. The document will be delivered in electronic format for posting to Department websites. The contractor's response must include a copy of a technical report produced for a similar state assessment program.

k) As a separate cost item, the contractor's response must include costs for providing each state, district, and school with software to enable them to analyze and produce reports from their student-level data. The contractor's response must include a detailed description of the proposed software.

l) As a separate cost item, the contractor's response must include the production of Spanish language versions of Parent/Guardian report templates and interpretive materials.

m) As a separate cost item, the contractor's response must include the electronic distribution to schools of student work from the Inquiry session of the science tests. The response must include a description of the software that will be used and its capabilities.

2. Reporting Support

a) The contractor will prepare materials for and support the administration of reporting workshops in each state within one month of the release of results. Department and contractor staff will conduct the workshops. At each workshop, the contractor will be responsible for on-site registration, distribution of materials, and related tasks. The contractor should budget for 5 half-day workshops per state to be held at five sites per state. Estimates of total participants should be based on two participants per school and an additional two participants per district.

b) The contractor will provide toll-free telephone support to districts and schools during the reporting period. The contractor's response should discuss options for staffing the support center, training support personnel, and duration of support. The contractor's response must also discuss procedures for ensuring that efficient service is provided during peak times as well as contingencies for providing support in the event of a breakdown in telephone service. At a minimum, the contractor's response must include providing telephone support for a period beginning with the shipment of results and continuing for three weeks.

c) The contractor's response must include a description of the procedures that will be used to collect, record, and investigate reports by districts and schools of discrepancies and errors in results.

3. Retrieving Student Work

At the request of a state Department, the contractor will retrieve and deliver to the appropriate Department images of student answer documents and/or actual student test materials in response to concerns about the accuracy of reported results. All requests must be made through an appropriate state Department. The contractor will bear the costs to perform this task.

a) The contractor's response must include a cost figure and timetable for retrieving and delivering images of individual student answer documents upon request of a state Department. The cost for this service will be charged upon request and should not be included in the budget.

H. Standard Setting

Student-level results from the science test will be reported according to four achievement levels indicating an overall level of science literacy. Standard setting will be conducted in the summer of 2008 following the first operational administration of the science tests.

a) The contractor will propose an appropriate standard setting methodology and procedure that meets the following goals:

- Is appropriate for science tests that cover three domains of science and inquiry and include a variety of item types.
- Supports coherence across the grade levels tested.
- Includes the direct participation of science teachers and other educators.

- Includes the validation of standard setting results with information gained from educators in the field and through the use of other available information, as appropriate.
- Is consistent with the goals and purposes of the states' science test specifications and assessment principles.

b) The contractor's response must include a comprehensive description of the proposed standard setting method that includes procedures to occur before, during, and following the formal standard setting process. The response must also include information on contractor staff that will lead and participate in standard setting activities.

c) The contractor will support all standard setting activities including, but not limited to, providing any stipends, substitute reimbursement, and covering expenses for participants in proposed meetings for the standard-setting process.

d) The contractor will produce a written report documenting all aspects of the standard setting process. The report will be delivered to the states within 30 days of the completion of standard setting.

V. Attachments

A. Number of Students per grade

B. Number of Districts and Schools

C. NECAP Science Test Specifications

D. Budget Forms

E. Task Allocation Forms

Attachment A

Number of Rhode Island Students per Grade:

Grade 4 Students	12,000
Grade 8 Students	12,000
Grade 11 Students	11,000

Attachment B

Number of Rhode Island Districts and Schools:

Number of Rhode Island Districts	40
Number of Rhode Island Schools	370

Attachment C

NEW ENGLAND COMMON ASSESSMENT PROGRAM (NECAP)
NEW HAMPSHIRE, RHODE ISLAND, & VERMONT

TEST SPECIFICATIONS FOR THE

TRI-STATE SCIENCE ASSESSMENT

November 2005

Tri-State Science Assessment Overview

The Tri-State large-scale science assessment will be designed to provide individual student performance data, information useful to teachers in planning instruction, and school-wide data on the effectiveness of the school/district science program in achieving the overarching goal of science literacy for all students. This assessment will be administered operationally for the first time in the spring (beginning in May 2008) at the end of grades 4, 8, and 11. The assessment will consist of matrix and common items; however, only common items will be used to determine individual student scores. The assessment will include multiple-choice items (1 point), short answer items (2 points), and constructed response items (3 or 4 points). One testing session, with a longer extended response task will be devoted to assessing inquiry knowledge and skills and consist of a combination of item types as described in these test specifications. The assessments will yield an overall science score translated into a Performance Level, based on standard setting and establishment of cut scores. In addition to a total science score (based on approximately 61-64 possible score points), four reporting categories will include the three domains of science (Life Science, Earth/Space Science, and Physical Science) and Inquiry. It is *anticipated* that each grade span will have to have at least 7 to 10 test forms.

The science assessment will be a power test - administered during three testing sessions of approximately 50 minutes each. Students will be given additional time to complete each part of the test, if time is needed. Due to the importance of assessing students' ability to apply their inquiry skills within the context of science content, a longer extended response (ER) task assessing inquiry will be included as the third testing session.

The three testing sessions for science can generally be described as follows. Sessions 1 and 2 will consist of a combination of multiple-choice and 4-point constructed response items measuring the three domains of science, as identified in the Tri-State Science Assessment Targets for Earth/Space, Physical Science, and Life Science. These first two sessions will each last about 50 minutes; items will be grouped by science domain. For example, Session 1 will begin with all of the Physical Science items, followed by half of the Earth/Space Science items. Session 2 will include the second half of the Earth/Space Science items followed by all of the Life Science items.

Session 3 will focus on Inquiry and include up to 30 minutes for a "hands-on/minds-on" activity that may or may not be a scientific investigation, depending upon the grade level. The hands-on/minds-on activity may include data collection, viewing a video, or exploring a case study with multiple data sets. Four Broad Areas of Inquiry (described in detail in Section VI) will be assessed in the third testing session: Formulating Questions/Hypotheses, Planning/Critiquing Investigations, Conducting Investigations, and Developing/Evaluating Explanations. Session 3 will include a combination of 2-point short-answer and 3-point constructed-response items appropriate for the extended response task. These 2- and 3-point questions will be clustered for interpretation to indicate inquiry ability across the Four Broad Areas of Inquiry (described in Section VI of test specifications).

During Sessions 1 and 2, matrix items will be included to address breadth of content, embedded field testing, and equating. Session 3 will include about 10 minutes worth of matrix-sampled generic (meaning non-task specific) inquiry items that will be used to equate inquiry performance year-to-year. Some matrix items - addressing breadth of content - will be included

in school, district, and state reporting to provide additional information relating to school/district science programs, but will not contribute to individual student scores.

Introduction to Science Test Specifications

These Science Test Specifications were developed based upon preliminary discussions with state assessment directors, the assessment target development team for science, and advice from Center for Assessment staff. Science Test Specifications have been reviewed by science content staff and modified consistent with the assessment directors' recommendations. Further revisions are subject to review by constituents of each of the partner states. These test specifications and the Tri-State Science Assessment Targets for three grade spans are scheduled for final completion by January 2006.

Alignment Considerations

The science grade-span assessments will be administered operationally for the first time in the spring of 2008 at the end of grades 4, 8, and 11, drawing upon the science assessment targets from the grade spans of K-4, 5-8, and 9-11.

To the degree possible, the assessments are to be aligned with the test specification Science Assessment Targets and consider the following:

- **The specific science content knowledge and skills** identified in each Tri-State Science Assessment Target for grades K-4, 5-8, and 9-11;
- **The intended Depth of Knowledge** (based on Webb) and the "DOK ceiling" identified for each Tri-State Science Assessment Target (as described in Section II of this document);
- **Making connections to Unifying Themes/Big Ideas of Science** (as indicated in Assessment Target coding and described in Section I and Appendix A of this document);
- **Relating "sets of items" to Statements of Enduring Knowledge** for each science domain (as indicated in Assessment Target coding and defined in Section I of this document);
- **Applying the 13 assessment constructs for 4 Broad Areas of Inquiry** and the Tri-State Planning Guide for Investigations and Tri-State Planning Guide for NON-Investigations to development of Extended Response tasks for inquiry and matrix equating items (as described in Sections VI and VII of this document);
- **The Distribution of Emphasis** identified across science domains and inquiry (as described in Section II of this document); and
- **Measurement Benchmarks for Mathematics** assessment (as described in Appendix C of this document).

Tri-State Science Test Specifications are organized into the following sections:

Test Specification Sections	Focus of Section	Pages
I. Design Features of Tri-State Science Assessment Targets	Conceptual Matrix – Unifying Themes/Big Ideas of Science and Statements of Enduring Knowledge Assessment Target Development ("Intersections") and Format (Coding, Implied DOK, Intended use of conjunctions, e.g., and i.e.)	37-43
II. Reporting Categories and Prioritization Strategies	Reporting Categories for Science Distribution of Emphasis for Reporting Categories High Emphasis Assessment Targets Prioritization Questions for Science Depth of Knowledge Levels (DOK) Science Preliminary DOK "Ceilings" for Assessment Targets	44-51
III. Item Types, Extended Response Task, and Scoring Guides	Item Types: Multiple Choice, Short Answer, Constructed Response, Extended Response Task Sample Item(s) with Target Alignment (item type, target aspects, DOK levels) Scoring Guide Models for Item Types	52-59
IV. Administration Guidelines	Testing Times, Accommodations, Use of Classroom Materials	60
V. Overview of Test Design	Overview of Test Design Three testing sessions, with additional time provided if needed Common and Matrix Items Total score points for test	61-63
VI. Extended Response Task for Assessing Inquiry	Four Broad Areas of Inquiry 13 Constructs for Assessing Inquiry Guidelines for ER Task Development with Differences across Grade Spans Tri-State Planning Guides for Investigations & NON-Investigations	64-70
VII. Sample Extended Response Tasks	Sample Grades 4, 8, and 11 Extended Response Tasks with alignment to Assessment Targets, EK Statement, and Tri-State Planning Guides for Investigations & NON-Investigations	71-88
VIII. Summary of Test Components, Rationales, and Bidder Flexibility	Review of Rationales for each Test Component with Related Bidder Requirements and/or Flexibility	89-90
Appendices	A. Unifying Themes/Big Ideas of Science B. Bibliography of References C. Mathematics Measurement Benchmarks (NECAP grades 3-8)	91-97

I. Design Features of Tri-State Science Assessment Targets

The criteria used for the development of Assessment Targets are described in this section of the test specifications. Test developers should carefully review the criteria and development process in order to fully understand how Unifying Themes/Big Ideas of Science and Statements of Enduring Knowledge for each domain of science *play a significant role* in assessment targets, test design, and item development.

A. Conceptual Matrix – Beginning with Unifying Themes/Big Ideas of Science

The conceptual matrix below served as an organizing tool for developing science assessment targets that address both the domain-specific core content of the Statements of Enduring Knowledge (found in Table 1.2) and the broader universal principles that integrate the different scientific disciplines - the Unifying Themes/Big Ideas of Science. Six Unifying Themes/Big Ideas of Science were chosen after an extensive review of the literature and are further described below in Table 1.1 (and in Appendix A).

Table 1.1:
Conceptual Matrix - Developing/Prioritizing Assessment Targets for Tri-State Science Assessment

Unifying Themes/Big Ideas of Science

(Subheadings under each Unifying Theme/Big Idea suggest but are not limited to what might be addressed,)

Scientific Inquiry	**Nature of Science**	**Systems & Energy**	**Models & Scale**	**Patterns of Change**	**Form & Function**
• Collect data • Communicate understanding & ideas • Design, conduct, & critique investigations • Represent, analyze, & interpret data • Experimental design • Observe • Predict • Question and hypothesize • Use evidence to draw conclusions • Use tools, & techniques	• Accumulation of science knowledge (evidence & reasoning, looking at work of others) • Attitudes and dispositions of science (avoiding bias, divergent ideas, healthy skepticism) • History of Science • Science/Tech/ Society • Scientific Theories	• Cycles • Energy Transfer • Equilibrium • Interactions • Interdependence • Order & Organization	• Evidence provided through… • Explanations provided through… • Relative distance • Relative sizes *Models include - experimental models, simulations, & representations used to demonstrate abstract ideas*	• Constancy and Change • Cycles • Evolutionary Change	• Natural World • Designed World

Tri-State Assessment targets are written to:

(1) **be general** enough to allow for multiple potential test items/assessment tasks with varying cognitive demands addressing each assessment target; and

(2) **have a cognitive demand (DOK) "ceiling"** generally consistent with (Webb's descriptions of) **Strategic Thinking** (Level 3) – requiring reasoning, planning, using/citing evidence, explaining/justifying thinking, drawing conclusions from data/observations, developing a logical argument for concepts, explaining phenomena in terms of concepts, or solving problems with more than one possible answer **OR Skills and Concepts** (Level 2) – classify, organize, make observations, compare data, explain relationships, describe examples/non-examples, interpret information.

B. Statements of Enduring Knowledge for 3 Domains of Science

Tri-State Science Assessment Targets are organized into three science domains and further subdivided into 10 Statements of Enduring Knowledge (EK). Each Assessment Target is linked to one Statement of Enduring Knowledge, as indicated with the target's coding (e.g., LS1 means Life Science and the first EK statement, LS2. means Life Science and the second EK statement, etc.)

Statements of Enduring Knowledge (EK)…

- Are intended to identify the fundamental knowledge/concepts for each domain of science
- Cut across grade levels, so that learning is developmental/built upon across grades (although not all aspects of the EK may be addressed at all grade levels)
- Are of comparable grain size (and generally of a larger grain size than that of any single grade's stated expectations – Grade Level Expectation/GLE or Grade Span Expectation/GSE)
- Encompass, as a set, *the essential learning for each domain of science*
- Imply topics of study (and therefore, lead to focused instruction, as identified in science standards/GSEs/benchmarks)

C. The "Intersections" Create the Assessment Targets

Assessment Targets for high school, middle school, and elementary school were developed by applying the Unifying Themes/Big Ideas of science (listed in Table 1.1) to the Statements of Enduring Knowledge for each of the science domains of Life Science, Earth and Space Science, and Physical Science (listed in Table 1.2). **Not every Unifying Theme/Big Idea has an "intersection" with every Statement of Enduring Knowledge**. Development committees used prioritization strategies and field reviews to determine which assessment targets would provide the richest opportunities for large-scale assessment purposes.

Tri-State Science Test Specification Assessment Targets are…

- Derived from and aligned with national and NH, RI, and VT's state science standards
- The "intersections" resulting from applying Unifying Themes/Big Ideas to Statements of Enduring Knowledge [e.g., What Systems & Energy concepts are essential to understanding LS1: All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, and species)?]
- Constructed with the understanding that not every Unifying Theme/Big Idea will have a meaningful "intersection" with every Statement of Enduring Knowledge
- General/broad enough to allow for multiple potential test items/assessment tasks with varying cognitive demands (DOK Levels)
- For the most part, written with an intended cognitive demand ceiling consistent with Depth of Knowledge (DOK) Levels 2 (Skills & Concepts) or 3 (Strategic Thinking) – based on the work of Norman L. Webb

Table 1.2 lists the EK Statements for each domain of science and indicates the number of Tri-State Science Assessment Targets for each grade span, by domain and by EK content cluster**.**

Statements of Enduring Knowledge are intended to focus instruction and assessment on the essential learning for each domain of science.
Each EK Statement has a code (e.g., LS1, ESS2, PS3, etc.) with appears at the beginning of each assessment target's code.

Not all aspects of the EK statement are addressed at each grade span, given that it may not be developmentally appropriate or practical (in terms of instructional time) to do so. For, example, in Earth/Space Science, the elementary school assessment focus is on ESS1 (Earth and earth materials); the middle school assessment focus is on ESS1 (Earth and earth materials) and ESS2 (solar system); and the high school assessment focus is on ESS1 (Earth and earth materials) and ESS3 (origin of the universe).

Table 1.2:
Number and Distribution of Science Assessment Targets by Grade Span (Draft July 2005).

Science Domain	Statements of Enduring Knowledge (EK)	Elem K-4	Middle 5-8	High School 9-11
Life Science	**LS 1** All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, and species).	**4**	**4**	**2**
	LS 2 Matter cycles and energy flows through an ecosystem.	**2**	**3**	**3**
	LS 3 Groups of organisms show evidence of change over time (structures, behaviors, and biochemistry).	**1**	**2**	**3**
	LS 4 Humans are similar to other species in many ways, and yet are unique among Earth's life forms.	**2**	**3**	**2**
	Life Science Totals	**9**	**12**	**10**
Earth & Space Science	**ESS 1** The Earth and earth materials as we know them today have developed over long periods of time, through continual change processes.	**6**	**5**	**4**
	ESS 2 The earth is part of a solar system, made up of distinct parts that have temporal and spatial interrelationships.	**0**	**3**	**0**
	ESS 3 The origin and evolution of galaxies and the universe demonstrate fundamental principles of physical science across vast distances and time	**0**	**0**	**4**
	Earth/Space Science Totals	**6**	**8**	**8**
Physical Science	**PS 1** All living and nonliving things are composed of matter having characteristic properties that distinguish one substance from another *(independent of size or amount of substance)*	**3**	**5**	**4**
	PS 2 Energy is necessary for change to occur in matter. Energy can be stored, transferred and transformed, but cannot be destroyed.	**3**	**2**	**3**
	PS 3 The motion of an object is affected by forces.	**2**	**1**	**3**
	Physical Science Totals	**8**	**8**	**10**
Total Tri-State Science Assessment Targets		**23**	**28**	**28**

D. Format of Science Assessment Targets

The set of assessment targets within and across a grade span and within an EK cluster have specific features that developers need to consider to assure item alignment to targets. The features include: (1) the format and meaning of the coding; (2) the implied cognitive demand; (3) the intended meaning of the use of the conjunctions "and" and "or" for item development; and (4) the intended use of "i.e." versus "e.g."

1. The format and meaning of assessment target coding

Each Assessment Target contains a code before the narrative text of the target. These codes identify the specific Statement of Enduring Knowledge, the grade span, the connections to one or more Unifying Theme/Big Idea, and finally the target number.

Table 1.3 illustrates an example: **LS1 (K-4) INQ+POC –1** means that this target addresses the first Life Science EK statement (**LS1**); the (**K-4**) grade span; is linked to Unifying Themes/Big ideas of Inquiry (**INQ**) and Patterns of Change (**POC**); and is the first assessment target listed **(1**) under the domain of Life Science.

Table 1.3 Sample Target Coding

LS1 – All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, and species)

Elementary Target	Middle School Target	High School Target
LS1 (K-4) INQ+POC –1 Sort/classify different living things using similar and different characteristics. Describe why organisms belong to each group or cite evidence about how they are alike or not alike.	LS1 (5-**8**) – INQ+ SAE- **1** Using data and observations about the biodiversity of an ecosystem make predictions or draw conclusions about how the diversity contributes to the stability of the ecosystem.	LS1 (9-**11**) INQ+SAE+FAF -**1** Use data and observation to make connections between, to explain, or to justify how specific cell organelles produce/regulate what the cell needs or what a unicellular or multi-cellular organism needs for survival (e.g., protein synthesis, DNA replication, nerve cells)

Numbering is consecutive within each domain of science for each grade span. For example, at grades K-4, Life Science targets are numbered 1 though 9 (beginning with LS1, then continuing with LS2, LS3, and LS4); Physical Science targets begin the numbering again with 1 through 8 for PS1, PS2 and PS3; and Earth/Space Science targets again begin numbering 1 through 6.

While the Statements of Enduring Knowledge are the same across all grade spans, the set of related targets within a grade span *do not address all aspects of the EK Statement*. This was done intentionally to focus instruction and assessment on the essential learning for the grade span, as well as on the developmentally appropriate concepts. For example, at the elementary grade span, LS1 will focus on organisms and external structures, while the middle school grade span will move to internal structures and include organisms and populations. Local instruction and assessment will include other foundational concepts not included for large-scale assessment at a particular grade span and opportunities for extended learning beyond assessment targets.

2. The implied cognitive demand of each assessment target

The text of the assessment target identifies specific concepts of the science domain and implies the level of cognitive demand for how students at the given grade span will interact with the content to demonstrate their understanding. Most targets indicate several ways that students will interact with the science concepts to demonstrate what they know and can do. Often, several different Depth of Knowledge levels (based on Webb) are implied.

For the most part, Tri-State Assessment Targets have been written to a cognitive demand "ceiling" consistent with (Webb's) **Level 3 Strategic Thinking** – requiring reasoning, planning, using/citing evidence, explaining/justifying thinking, drawing conclusions from data/observations, developing a logical argument for concepts, explaining phenomena in terms of concepts, or solving problems with more than one possible answer **OR Level 2 Skills and Concepts** – classifying, organizing data, making observations, comparing data, explaining relationships, describing examples/non-examples, interpreting information.

While the Depth of Knowledge ceiling cannot be exceeded for assessment, creating items that address both the ceiling and lower DOK levels is desirable. (Depth of Knowledge Levels will be elaborated on in greater detail in Section II of test specifications.)

Table 1.4 provides some samples of the implied DOK levels of some assessment targets.

Table 1.4 Sample Implied Depth of Knowledge Levels

LS1 – All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, and species)

Elementary	Middle	High School
LS1 (K-**4**) INQ+POC –**1** Sort/classify different living things using similar and different characteristics. Describe why organisms belong to each group or cite evidence about how they are alike or not alike.	LS1 (5-**8**) – INQ+ SAE- **1** Using data and observations about the biodiversity of an ecosystem make predictions or draw conclusions about how the diversity contributes to the stability of the ecosystem.	LS1 (9-**11**) INQ+SAE+FAF **-1** Use data and observation to make connections between, to explain, or to justify how specific cell organelles produce/regulate what the cell needs or what a unicellular or multi-cellular organism needs for survival (e.g., protein synthesis, DNA replication, nerve cells)

Interpreting information from "data" implies a DOK level of 2.

"Draw conclusions," using data implies a DOK level of 3.

The DOK ceiling for the target is Level 3, but other DOK levels can be assessed using multiple test items.

Explaining a concept implies a DOK level of 2.

"Justify how" implies a DOK level of 3.

The DOK ceiling for the target is Level 3, but other DOK levels can be assessed using multiple test items.

Depth of Knowledge (DOK) Levels - Implications for item development

Test developers and item writers should align items with the target's DOK ceiling, but consider how the "set of test items" provides a range of cognitive demand. *Test developers will articulate a strategy for assuring test item alignment with intended DOK ceiling levels and for providing a range of DOK within "item sets" for assessing domains of science.*

Table 1.5 illustrates the overall format of Tri-State Science Assessment Targets for Grades 4, 8 and 11 for LS1.

The Statement of Enduring Knowledge appears at the top of each page as a conceptual organizer for the set of assessment targets.

All targets related to the first Life Science EK Statement begin with "LS1."

Table 1.5 Sample Format of Tri-State Assessment Targets

LS 1
All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, & species).

Elementary	Middle	High School
LS1 (K-**4**) - INQ+POC –**1** Sort/classify different living things using similar and different characteristics. Describe why organisms belong to each group or cite evidence about how they are alike or not alike.	LS1 (5-**8**) – INQ+ SAE- **1** Using data and observations about the biodiversity of an ecosystem make predictions or draw conclusions about how the diversity contributes to the stability of the ecosystem.	LS1 (9-**11**) INQ+SAE+FAF –**1** Use data and observation to make connections between, to explain, or to justify how specific cell organelles produce/regulate what the cell needs or what a unicellular or multi-cellular organism needs for survival (e.g., protein synthesis, DNA-replication, nerve cells)
LS1 (K-4) SAE-2 Identify the basic needs of plants and animals in order to stay alive (1.e., water, air, food, space)	LS1 (5-**8**) SAE+FAF –**2** Describe or compare how different organisms have mechanisms that work in a coordinated way to obtain energy, grow, move, respond, provide defense, enable reproduction, or maintain internal balance (e.g., cells, tissues, organs and systems).	LS1 (9-**11**) FAF+ POC **-2** Explain or justify with evidence how the alteration of the DNA sequence may produce new gene combinations that make little difference, enhance capabilities, or can be harmful to the organism (e.g., selective breeding, genetic engineering, mutations)
LS1 (K-**4**) POC –**3** Predict, sequence or compare the life stages of organisms – plants and animals (e.g., put images of life stages of an organism in order, predict the next stage in sequence, compare two organisms)	LS1 (5-**8**) POC **-3** Compare and contrast sexual reproduction with asexual reproduction.	
LS1 (K-**4**) FAF –**4** Identify and explain how the physical structures of an organism (plants or animals) allow it to survive in its habitat/environment (e.g., roots for water; nose to smell fire)	LS1 (5-**8**) FAF –**4** Explain relationships between or among the structure and function of the cells, tissues, organs, and organ systems in an organism.	

Some targets are only linked to one Unifying Theme/Big Idea.

Some targets are linked to more than one Unifying Theme/Big Idea.

Target Coding and Format - Implications for item development

Test developers and item writers should not write individual items that narrowly focus on an aspect of one target, but consider how the "set of items" for each statement of Enduring Knowledge (and multiple assessment targets) address the broader concepts suggested by EK Statements and Unifying Themes/Big Ideas. *Test developers will articulate a strategy for making intentional connections within "item sets" for assessing domains of science and the broader science concepts (e.g., items linked to a common scenario or addressing a Unifying Theme with several related targets/ items).*

3. The intended meaning for the use of conjunctions– "and" and "or" – within Science Assessment Targets

It is recognized that all aspects of the Tri-State Science Assessment Targets cannot be sampled every year for assessment. The use of specific conjunctions in the Assessment Targets is clarified for item development with the following intended meanings:

- "And" means that *to the extent possible*, elements within an assessment target connected by "and" should be included in the assessment every year. Sometimes "and" is implied though the use of multiple statements describing what students will do to demonstrate learning. (See also Appendix D for further analysis of assessment targets including "and.")
- "Or" means that items assessing those aspects of the assessment target can vary from year to year; *any or all* aspects are possible for inclusion in a given year's assessment.

Table 1.6 Sample of Use of Conjunctions "and" and "or'

LS1 – All living organisms have identifiable structures and characteristics that allow for survival (organisms, populations, and species)

Elementary	Middle	High School
LS1 (K-**4**) INQ+POC –**1** Sort/classify different living things using similar and different characteristics. Describe why organisms belong to each group OR cite evidence about how they are alike or not alike.	LS1 (5-**8**) – INQ+ SAE- **1** Using data and observations about the biodiversity of an ecosystem, make predictions OR draw conclusions about how the diversity contributes to the stability of the ecosystem.	LS1 (9-**11**) INQ+SAE+FAF -**1** Use data and observation to make connections between, to explain, OR to justify how specific cell organelles produce/regulate what the cell needs or what a unicellular or multi-cellular organism needs for survival (e.g., protein synthesis, DNA-replication, nerve cells)

Multiple statements are used in the target. Although "and" is not explicitly stated, aspects in both statements are to be assessed every year at this grade span because "and" is implied. "Or" used in the second statement means that either part can be assessed.

The use of "or" in this target means that all or some aspects of the target may be assessed on a given year. Students can *be asked* to use data and observations to: (1) make connections, (2) explain, OR (3) justify, or (4) any do combination of the three. Choices made about the assessment focus also have DOK implications – (e.g., "justify = DOK 3)

4. The intended use of "i.e." versus "e.g."

The use of "i.e." in some Assessment Targets is to limit item content to that specific list. The use of "e.g." in some Assessment Targets is to provide guidance, but not necessarily limit item content to that specific list.

In this high school target, the "e.g." list suggests the possible and appropriate focus for the test items (protein synthesis, DNA-replication, nerve cells)**, but does not limit items to this content specifically.**

Use of Conjunctions and "e.g.," - Implications for item development

Test developers and item writers will have some flexibility in determining the focus of what will be assessed when "or" or "e.g.," are used in assessment targets.

II. Proposed Reporting Categories and Prioritization for Science

A. Proposed Reporting Categories and Distribution of Emphasis

Proposed Reporting Categories
There will be an overall science score with four sub-reporting categories for the Tri-State Science Assessment. Individual students will be provided with the following feedback on their performance:

1. Overall Science Score: The overall science score will be based only on common items/score points.
2. Performance Level: The overall science score will translate into a Performance Level, based on standard setting and establishment of cut scores. There will be four Performance Levels (described more specifically in the RFP) consistent with large-scale assessment Performance Levels in other content areas (i.e., NECAP mathematics, reading, and writing).
3. Four Reporting Categories – Individual students will also have sub-scores in the three Domains of Science (Earth/Space Science, Physical Science, and Life Science) and Inquiry.

***Preliminary* Distribution of Emphasis and High-Emphasis Assessment Targets**
In grades 8 and 11, score points for common items will be evenly distributed across the four reporting categories, even though there may not be an equal number of assessment targets for each science domain. In grade 4, Inquiry and Life Science will have slightly greater emphasis (and more score points) than Earth/Space Science and Physical Science. Additionally, specific targets have been identified for *greater assessment emphasis* (meaning that CR items, ER items, or several MC items will address those targets whenever possible). These targets are called "high-emphasis" assessment targets.*

The Distribution of Emphasis for Reporting Categories is described in Table 2.1 below.

Table 2.1 *Preliminary* Science Distribution of Emphasis by Reporting Category (June 2005)
Percentages of Score Points and Number of Related Assessment Targets

Reporting Categories	Grade 4		Grade 8		Grade 11	
	% Score points	Number of targets	% Score points	Number of targets	% Score points	Number of targets
Life Science LS1, LS2, LS3, LS4	**30%**	9	**25%**	12	25%	10
Earth/Space Science ESS1, ESS2, ESS3	**20%**	6	**25%**	8	25%	8
Physical Science PS1, PS2, PS3	**20%**	8	**25%**	8	25%	10
Inquiry Combinations of one or more targets from one EK/domain	**30%**	NA	**25%**	NA	25%	NA

*A list of high-emphasis targets is included on the following page.

"High-Emphasis" Assessment Targets

The state content committee used an extensive resource review to determine which assessment targets should receive *greater instructional emphasis*, and therefore receive greater assessment emphasis - meaning devoting more test items and/or more test score points to these targets, when choices need to be made. These high-emphasis targets, which represent about half of the total assessment targets per grade span, have been identified for *potential* multiple test score points (e.g., CR items, ER items, or several MC items). Table 2.2 lists the high-emphasis targets with their preliminary DOK ceiling levels. To the degree possible, CR items, ER items, and/or at least two MC items should be aligned to these high-emphasis targets. All high emphasis targets will be assessed every year using at least one common item.

Table 2.2 – High-Emphasis Assessment Targets with *Preliminary* DOK Ceilings*

Science Domains by EK Statement	Grade 4 Targets with DOK	Grade 8 Targets with DOK	Grade 11 Targets with DOK
LS1 Survival of organisms	1 – DOK 2a, b, g 2 – DOK 1a, b	1 – DOK 3h 2 – DOK 2a	1 – DOK 3d 2 – DOK 3d
LS2 Matter and energy in ecosystems	6 – DOK 2a	5 – DOK 2a, d 6 – DOK 2a	3 – DOK 2a
LS3 Organisms change over time		8 – DOK 2a, h	8 – DOK 3a, f
LS4 Humans are similar, yet unique	8 –DOK 2a, h	11 – DOK 2a, b	
PS1 Properties and structure of matter	1 –DOK 3h	1 –DOK 2a, c, d, e, i 2 –DOK 2e, g, j 4 –DOK 2a, b	3 –DOK 2a, b 4 –DOK 3c, g, j
PS2 Energy		6 –DOK 3c, j, l, o	6 –DOK 3a, c, h
PS3 Forces and motion	7 –DOK 2a, j	8 –DOK 2a, e, g, i, j,	8 –DOK 3a, c, h 9 –DOK 2a, b 10 –DOK 2a
ESS1 Earth and earth materials	1 –DOK 2b, e, g 2 –DOK 3c, h 4 –DOK 2a, b 5 –DOK 2a, b	2 –DOK 2a 3 –DOK 2a, b 5 –DOK 3c, d, h, k	1 –DOK 3a, c, d, f, l 3 –DOK 3o
ESS2 Solar system		6 –DOK 2a, g, h, j 8 –DOK 3j, o	
ESS3 Universe and galaxies			6 –DOK 3b, c, d, l, o 7 –DOK 3o 8 –DOK 2a, b

* See pages 47-48 for a description of Depth of Knowledge (DOK) levels in science. Coding for DOK ceilings are aligned with the descriptions on page 49. For example, "2a" means Level 2 DOK and the description for "a" – Specify and explain the relationship between facts, terms, properties, or variables.

B. Prioritization

The Tri-State science content development committee employed several strategies to help prioritize the specifics identified within and across EK Statement clusters and assessment targets for large-scale assessment. This work has included a Balance of Representation Study to determine the Distribution of Emphasis for the reporting categories (Table 2.1), identification of

"high-emphasis" targets (Table 2.2), and identification of "ceilings" for Depth of Knowledge Levels for the Tri-State Science Assessment Targets (Table 2.5)

Prioritization Questions Used in Development of Tri-State Science Assessment Targets (K. Hess, 2004)

1. **Will the concept or skill *lead to a better understanding* of the unifying theme/big idea (unifying concept, essential process) as it intersects with the Statement of Enduring Knowledge of the discipline?** Is the learning described in the assessment target important for understanding the big idea?
2. **Will the concept or skill *lead to a better understanding* of the Statement of Enduring Knowledge?** Is the learning described in the assessment target essential for understanding the science concepts?
3. **Is the concept that is being assessed in the target identified by national resources as an important/essential concept for this grade span (**as documented in Resource Review)?
4. **Is the concept or skill subsumed in other indicators at that grade level?** (E.g., is the skill of knowing that multiple trials are needed subsumed in demonstrating ability to design a "fair test"?)
5. **Are concepts or skills important for success in other science domains in given grades or subsequent grades?** (E.g., is this concept about systems in Life Science important in understanding the concept of systems in Earth Science?)
6. **Is this concept or skill better assessed "locally" (at the district, school, or classroom level) – and NOT included for large-scale assessment because:**
   - There are time or materials constraints (e.g., uses tools to collect data, conducts an investigation over an extended period of time, researches and reports about a scientific discovery);
   - Too small a grain size (e.g., defining/recalling a term, property, or fact or naming the order of planets as opposed to explaining a relationship among planets);
   - Unrealistic for an on-demand assessment (e.g., design a classification system based on observed or identified similarities and differences, verify or modify prior understandings based on an analysis of new information);
   - The concept may be related to the topic of study, but not an essential learning for all students to master at this grade level - *and is likely to get less instructional emphasis* (e.g., explaining the phenomenon of spring tides and neap tides in terms of positions of the moon and sun relative to Earth). Describing what causes high and low tides may be the more important concept for all students to learn;
   - The concept or skill is assessed adequately at an earlier grade, and therefore does not need to be continued to be assessed at this grade level for large-scale assessment. Is this a foundational skill or concept that we can assume students know at this grade level?
   - Not all states have identified this as important to include in large-scale assessment (e.g., content or skills are not included in state standards at this grade level for some states).

C. Depth of Knowledge (DOK) Levels for Science

The Tri-State Science Assessment will assess Depth of Knowledge Levels 1, 2, and 3. While it may be possible to assess the extended response (ER) task for inquiry at a Depth of Knowledge of Level 4, time and materials constraints of the on-demand testing setting will likely prohibit this. Four DOK levels are included herein for illustration purposes, with general definitions for each of the Depth of Knowledge levels followed by Table 2.2, which provides further specification and sample descriptions for each of the DOK levels for science.

Descriptors of DOK Levels for Science (based on Webb, 1997 and March 2002 and TIMSS Science Assessment Framework, 2003)

Level 1 Recall and Reproduction requires recall of information, such as a fact, definition, term, or a simple procedure, as well as performing a **simple** science process or procedure. Level 1 only requires students to demonstrate a rote response, use a well-known formula, follow a set procedure (like a recipe), or perform a clearly defined series of steps. A "simple" procedure is well-defined and typically involves only **one-step**. Verbs such as "identify," "recall," "recognize," "use," "calculate," and "measure" generally represent cognitive work at the recall and reproduction level. Simple word problems that can be directly translated into and solved by a formula are considered Level 1. Verbs such as "describe" and "explain" could be classified at different DOK levels, depending on the complexity of what is to be described and explained.

A student answering a Level 1 item either knows the answer or does not: that is, the answer does not need to be "figured out" or "solved." In other words, if the knowledge necessary to answer an item automatically provides the answer to the item, then the item is at Level 1. If the knowledge necessary to answer the item does not automatically provide the answer, the item is at least at Level 2.

Level 2 Skills and Concepts includes the engagement of some mental processing beyond recalling or reproducing a response. The content knowledge or process involved is **more complex** than in level 1. Items require students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply **more than one step**. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Some action verbs, such as "explain," "describe," or "interpret," could be classified at different DOK levels, depending on the complexity of the action. For example, interpreting information from a simple graph, requiring reading information from the graph, is a Level 2. An item that requires interpretation from a complex graph, such as making decisions regarding features of the graph that need to be considered and how information from the graph can be aggregated, is at Level 3.

Level 3 Strategic Thinking requires deep knowledge using reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. The cognitive demands at Level 3 are **complex and abstract**. The complexity does not result only from the fact that there could be multiple answers, a possibility for both Levels 1 and 2, but because the multi-step task requires

more demanding reasoning. In most instances, requiring students to explain their thinking is at Level 3; requiring a very simple explanation or a word or two should be at Level 2. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Experimental designs in Level 3 typically involve more than one dependent variable. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems.

Level 4 Extended Thinking requires **high cognitive demand** and is **very complex**. Students are required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select or devise one approach among many alternatives on how the situation can be solved. Many on-demand assessment instruments will not include any assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated in such a way as to expect students to perform extended thinking. "Develop generalizations of the results obtained and the strategies used and apply them to new problem situations," is an example of a Grade 8 objective that is a Level 4. Many, but not all, performance assessments and open-ended assessment activities requiring significant thought will be at a Level 4.

Level 4 requires complex reasoning, experimental design and planning, and **probably will require an extended period of time** either for the science investigation required by an objective, or for carrying out the multiple steps of an assessment item. However, the extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2 activity. However, if the student conducts a river study that requires taking into consideration a number of variables, this would be a Level 4.

Table 2.3: Sample Descriptors for each of the DOK Levels in Science, based on Webb
(Working draft K. Hess, updated September 2005)

Level 1 Recall & Reproduction	Level 2 Skills & Concepts	Level 3 Strategic Thinking	Level 4 Extended Thinking
a. Recall or recognize a fact, term, definition, simple procedure (such as one step), or property b. Demonstrate a rote response c. Use a well-known formula d. Represent in words or diagrams a scientific concept or relationship e. Provide or recognize a standard scientific representation for simple phenomenon f. Perform a routine procedure, such as measuring length g. Perform a **simple** science process or a set procedure (like a recipe) h. Perform a clearly defined set of steps i. Identify, calculate, or measure **NOTE: If the knowledge necessary to answer an item automatically provides the answer, it is a Level 1.**	a. Specify and explain the relationship between facts, terms, properties, or variables b. Describe and explain examples and non-examples of science concepts c. Select a procedure according to specified criteria and perform it d. Formulate a routine problem given data and conditions e. Organize, represent, and compare data f. Make a decision as to how to approach the problem g. Classify, organize, or estimate h. Compare data i. Make observations j. Interpret information from a simple graph k. Collect and display data **NOTE: If the knowledge necessary to answer an item <u>does not</u> automatically provide the answer, then the item is at least a Level 2. Most actions imply more than one step.** **NOTE: Level 3 is complex and abstract. If more than one response is possible, it is at least a Level 3 and calls for use of reasoning, justification, evidence, as support for the response.**	a. Interpret information from a complex graph (such as determining features of the graph or aggregating data in the graph) b. Use reasoning, planning, and evidence c. Explain thinking (beyond a simple explanation or using only a word or two to respond) d. Justify a response e. Identify research questions and design investigations for a scientific problem f. Use concepts to solve non-routine problems/more than one possible answer g. Develop a scientific model for a complex situation h. Form conclusions from experimental or observational data i. Complete a multi-step problem that involves planning and reasoning j. Provide an explanation of a principle k. Justify a response when more than one answer is possible l. Cite evidence and develop a logical argument for concepts m. Conduct a designed investigation n. Research and explain a scientific concept o. Explain phenomena in terms of concepts	a. Select or devise approach among many alternatives to solve problem b. Based on provided data from a complex experiment that is novel to the student, deduct the fundamental relationship between several controlled variables. c. Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions d. Relate ideas *within* the content area or *among* content areas e. Develop generalizations of the results obtained and the strategies used and apply them to new problem situations **NOTE: Level 4 activities often require an extended period of time for carrying out multiple steps; however, time alone is not a distinguishing factor if skills and concepts are simply repetitive over time.**

Depth-of-Knowledge as a "Ceiling" NOT as a "Single Focus"

An important consideration of large-scale assessment design is to use the highest Depth-of-Knowledge (DOK) demand implicit in an assessment target as the "ceiling" for assessment, not the "single focus" for assessment. Table 2.4 provides three sample assessment targets with different "ceilings," meaning the highest DOK cognitive demand Level at which it could be assessed. When considering the highest DOK Level as the ceiling, not the single focus, the assessment target has the *potential* to be assessed at the Depth-of-Knowledge Level ceiling, and

up to the ceiling, depending upon the expectations of the assessment target. Table 2.4 also indicates the other potential DOK levels at which the assessment target *might be* assessed.

Table 2.4 Examples of Tri-State Science Assessment Targets and *Potential* Depth of Knowledge Levels for Assessment Purposes

Samples **Tri-State Science Assessment Targets**	**DOK Ceiling**	**Potential DOK Levels for Assessment**
Example A: LS1 (K-**4**) SAE –**2** Identify the basic needs of plants and animals in order to stay alive (i.e., water, air, food, space)	**1**	**1** (Demonstrate a rote response; Recall or recognize a fact, term, definition, or property)
Example B: PS1 (5-**8**) INQ+ SAE –**3** Collect data or use data provided to infer or predict that the total amount of mass in a closed system stays the same, regardless of how substances interact (conservation of matter)	**2**	**1** (Perform a clearly defined set of steps; Take measurements of mass at different times or under different circumstances) **2** (Construct a graph to organize, display, and compare data; Specify and explain the relationship between facts, terms, properties, or variables; Interpret information from a simple graph)
Example C: PS1 (9-**11**) MAS+ FAF – **4** Model and explain the structure of an atom or explain how an atom's electron configuration, particularly the outermost electron(s), determines how that atom can interact with other atoms	**3**	**1** (Recall or recognize a fact, term, definition, or property) **2** (Specify and explain the relationship between facts, terms, properties, or variables) **3** (Develop a scientific model for a complex situation; Provide an explanation of a principle)

Why is this distinction between DOK as a "ceiling" and not as a "single focus" important? If assessed only as the "single focus" DOK level, all assessment anchors with a Level 2 or Level 3 as their highest demand would only be assessed at those highest levels. This would potentially have two negative impacts on the assessment: 1) The assessment as a whole could be too difficult; and 2) important information about student learning along the achievement continuum would be lost. Multiple items covering a range of DOK levels can provide useful instructional information for classroom teachers.

***Preliminary* DOK Ceiling Levels for Tri-State Science Assessment Targets** have been established to guide item development. Table 2.5 provides the intended DOK ceiling levels for Tri-State Science Assessment Targets. [Depths of Knowledge Levels are also provided for the Inquiry Constructs in Section VI of the test specifications (Table 6.1).]

Table 2.5 - *Preliminary* DOK Ceilings (using Table 2.2 Descriptors) for Science Assessment Targets			
Science Domains by EK Statement	**Grade 4 Targets with DOK**	**Grade 8 Targets with DOK**	**Grade 11 Targets with DOK**
LS1 Survival of organisms	1 – DOK 2a, b, g 2 – DOK 1a, b 3 – DOK 2h, i 4 – DOK 2a	1 – DOK 3h 2 – DOK 2a 3 – DOK 2a, b, h 4 – DOK 2a, b	1 – DOK 3d 2 – DOK 3d
LS2 Matter and energy in ecosystems	5 – DOK 1a, b 6 – DOK 2a	5 – DOK 2a, d 6 – DOK 2a 7 – DOK 2a	3 – DOK 2a 4 – DOK 2a 5 – DOK 3b, k
LS3 Organisms change over time	7 –DOK 2a	8 – DOK 2a, h 9 – DOK 2b	6 – DOK 3d 7 – DOK 3j 8 – DOK 3a, f
LS4 Humans are similar, yet unique	8 –DOK 2a, h 9 –DOK 2a, b	10 – DOK 3a, b 11 – DOK 2a, b 12 – DOK 1a, d	9 – DOK 3d, h 10 – DOK 3h
PS1 Properties and structure of matter	1 –DOK 3h 2 –DOK 2a 3 –DOK 2a	1 –DOK 2a, c, d, e, i 2 –DOK 2e, g, j 3 –DOK 2a, j, k 4 –DOK 2a, b 5 –DOK 2a, g	1 –DOK 2a, c, h, i 2 –DOK 3a, b, c, d 3 –DOK 2a, b 4 –DOK 3c, g, j
PS2 Energy	4 –DOK 2a, b 5 –DOK 2a,h, i 6 –DOK 2a, c, i	6 –DOK 3c, j, l, o 7 –DOK 3c, h	5 –DOK 2a, b, c 6 –DOK 3a, c, h 7 –DOK 2a, b
PS3 Forces and motion	7 –DOK 2a, j 8 –DOK 2a, b, e, g, h, i	8 –DOK 2a, e, g, i, j,	8 –DOK 3a, c, h 9 –DOK 2a, b 10 –DOK 2a
ESS1 Earth and earth materials	1 –DOK 2b, e, g 2 –DOK 3c, h 3 –DOK 1a, i 4 –DOK 2a, b 5 –DOK 2a, b 6 –DOK 2a, b, g, h	1 –DOK 3c, d, l 2 –DOK 2a 3 –DOK 2a, b 4 –DOK 3j, o 5 –DOK 3c, d, h, k	1 –DOK 3a, c, d, f, l 2 –DOK 3l 3 –DOK 3o 4 –DOK 2a
ESS2 Solar system		6 –DOK 2a, g, h, j 7 –DOK 3c, d, k 8 –DOK 3j, o	
ESS3 Universe and galaxies			5 –DOK 3c, k, l 6 –DOK 3b, c, d, l, o 7 –DOK 3o 8 –DOK 2a, b

Use of Depth of Knowledge (DOK) "ceilings" - Implications for item development
Test developers and item writers should align items with the target's DOK ceiling, but consider how the "set of test items" provides a range of cognitive demand. *Developers will articulate a strategy for assuring alignment with intended DOK ceiling levels and for providing a range of DOK within "item sets" for assessing domains of science.*

III. Item Types, Extended Response (ER) Task, and Scoring Guides for the Science Assessment

The Tri-State Science Assessment will include four item types: multiple choice (MC), short answer (SA), 3-point constructed response (3-CR), and 4-point constructed response (4-CR). Two-, three- or four-point scoring rubrics will be used to score the short answer (SA) and constructed response (CR) items. These will be task-specific, analytical rubrics, but will follow somewhat generic formats.

The first two testing sessions will assess the three domains of science and include common and matrix multiple choice (MC) and 4-point constructed response (4-CR) items. A longer, extended response (ER) task assessing inquiry during session 3 will combine a mix of 2-point short answer (SA) and 3-point constructed response (3-CR) items.

The assessment will yield approximately 61-64 total score points: 15 points coming from each of the three science domains and the remaining score points (approximately 16-18) from the extended response (ER) task for Inquiry. Section V of these test specifications provide a more complete overview of the test design and Section VI elaborates on the ER task and constructs for assessing inquiry skills and knowledge.

Table 3.1 on the following page briefly describes each of the four item types for the science assessment and identifies the testing sessions where they will be used. Following these general descriptions are annotated samples of item types that include:

- Grade Level
- Item Type
- Testing Session
- Alignment to Assessment Target(s) – Target code with aspects of the assessment target underlined to show what is addressed in the sample item
- Supporting explanation of item design related to skills and concepts assessed and links to Unifying Themes, EK Statements, and/or other targets
- Depth of Knowledge Level (for each sample item) using descriptions for each DOK level (page 49 of test specifications).

Grade Level: 4, 8, or 11

Item Type: Multiple Choice, Short Answer, or 3- or 4-point Constructed Response

Testing Session: 1 and 2 or 3

Alignment to Assessment Target(s):
Assessment Target #code with underlining to show all or some aspects assessed in the item

Supporting information about the item design related to Unifying Theme, EK Statement, or other targets:

Depth of Knowledge: Level ---: Appropriate descriptors from table 2.3 in Section II that match the sample item.

Table 3.1 Item Types for Tri-State Science Assessment		
Item Type	**Description**	**Purpose**
1-point Multiple Choice items (MC)	The multiple-choice items will consist of four alternative answers. While distracters will be plausible and reflect common errors, they must be unambiguously wrong in the judgment of experts. The use of distracters with common misconceptions is very desirable. Each item will be scored with an answer key and be worth one score point.	**The multiple-choice items will primarily be used in Session 1 and 2** to assess understanding of specific terms, concepts, and skills related to the three domains of science. These items will typically assess student performance at Levels 1 and 2 Depth of Knowledge.
2-point Short Answer items (SA)	Short answer items will allow students to construct brief responses using text (e.g., a few phrases or a sentence), labeling diagrams, or perhaps creating a simple scientific drawing of observations, predictions, lab set-up, etc. Each item will be assessed using a 2-point, task-specific analytical rubric.	**The short answer items will only be used in Session 3** to assess understanding of inquiry or science concepts. Because of the brevity of the response, these items will typically assess student performance at Levels 1 and 2 Depth of Knowledge.
3-point Constructed Response items (3-CR)	Constructed response items will include items requiring a more full response (such as selecting data or evidence to support a concept or a stated explanation), as well as items requiring some analysis of data or explanations of scientific reasoning. Each item will be assessed using a task-specific analytical rubric and be will be worth three points.	**3-CR items will only be used in Session 3,** assessing inquiry. The purpose of these items is to require that students explain or analyze concepts while applying aspects of scientific inquiry (e.g., developing explanations, planning, conducting, and critiquing investigations). These items will typically assess student performance at Levels 2 or 3 Depth of Knowledge.
4-point Constructed Response items (4-CR)	Constructed response items will include items requiring a more full response (such as selecting details or evidence to support a concept or stated explanation), as well as items requiring some analysis or providing explanations of reasoning. Each item will be assessed using a task-specific analytical rubric and will be worth four points.	**4-CR items will only be used in Sessions 1 and 2,** assessing domains of science/science content. The purpose of these items is to require that students explain science concepts or interpret/analyze how science concepts apply to real-world phenomena or specific situations. These items will typically assess student performance at Levels 2 or 3 Depth of Knowledge.
Extended Response task for Inquiry (ER)	Extended response (ER) tasks will give students the opportunity to apply their skills and understanding of inquiry and scientific investigation within the context of science concepts. All items will require students to construct short or longer responses, providing insights into students' abilities to generate, rather than recognize scientific concepts and to make connections among science concepts and inquiry skills. *A set of related items*, using a mix of SA and 3-CR items, will comprise the ER task. Items will be aligned to 13 possible constructs described in Section VI (page 64) for assessing inquiry. The Tri-State Planning Guide for Investigations (pages 67-68) provides the model for developing "full" investigations for grades 4 and 8. The Tri-State Planning Guide for NON-Investigations (pages 69-70) provides the model for developing ER tasks for grades 8 and 11 that are *not* investigations. See Sections VI for more details and Section VII sample tasks at each grade span. Each ER task will generate approximately 16-18 total score points.	**Session 3 will focus on assessing inquiry using an ER task.** The purpose of these tasks is to assess students' ability to apply a variety of inquiry skills and knowledge to specific science concepts. As a set, these items reflect a higher overall cognitive demand (e.g., planning, conceptual understanding, and scientific reasoning) primarily assessing Levels 2 and 3 Depths of Knowledge. The performance aspect of these tasks allows students to manipulate materials, objects, or data to analyze or solve scientific problems.

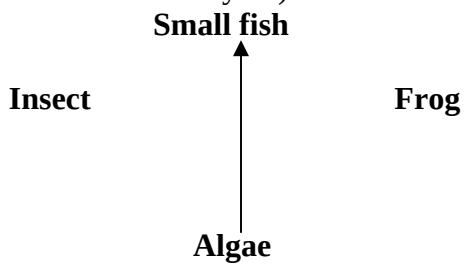
A. Sample Items Types

Four sample assessment items aligned to Tri-State Assessment targets are included (with annotations) in this section. Underlining indicates the aspect(s) of the assessment target with which the item is aligned.

3A.1 Multiple-choice items (MC) worth 1 score point will assess such things as: Recognizing a fact, term, definition, simple procedure, or property; Providing or recognizing a standard scientific representation for simple phenomenon; Representing in words or diagrams a scientific concept or relationship; Specifying or explaining the relationship between facts, terms, properties, or variables; Describing or explaining examples and non-examples of science concepts; Selecting a procedure according to specified criteria; Making observations; Classifying, organizing, or estimating; Comparing data; and Interpreting information from a simple graph.

Sample: 1. Chromosomal mutations occurring in gametes of humans can affect the appearance of offspring because… **a. many traits are usually affected** b. only one trait is usually affected c. these mutations usually speed up embryonic development d. these mutations usually result in sex-linked traits (Answer key: a) Source: Released item (Jan 2001) NYS Biology Regents	**Grade Level: 11** **Item Type:** Multiple Choice (1 point) **Testing Session:** 1 and 2 **Alignment to Assessment Target(s):** LS4 (9-**11**) NOS+INQ -**9** Use evidence to make and support conclusions about the ways that humans or other organisms are affected by environmental factors or heredity (e.g., pathogens, diseases, medical advances, pollution, mutations). **Supporting information about the item design related to Unifying Theme, EK Statement, or other targets:** This item might be used as scaffolding for a longer explanation or a comparison of effects to other organisms in other test items. It gets at a basic understanding of what mutations are and how they affect future generations of humans. It links to conceptual understandings in Target LS1 (9-11) FAF+POC –2 (alteration of the DNA sequence) **Depth of Knowledge:** Level 1 - Recall or recognize a fact, term, definition, or property

3A.2 Short Answer Items (SA) worth 2 score points will assess such things as: Stating a fact, term, definition, simple procedure, or property; Providing a standard scientific representation for simple phenomenon; Representing in words or diagrams a scientific concept or relationship; Specifying or explaining the relationship between facts, terms, properties, or variables; Describing or explaining examples and non-examples of science concepts; Selecting a procedure according to specified criteria; Making observations; Classifying, organizing, or estimating; Comparing data; and Interpreting information from a simple graph.

Sample: This picture (not included in test specifications) shows a pond ecosystem. Use this picture and what you know to answer these questions. 2. You will now finish a diagram of a food web in the pond. Draw arrows from each living thing to things that eat it. (The first arrow is drawn for you.)  (Scoring guide: 2-points – All possible connections made with no inaccuracies = 2 score points) Source: NAEP Science grade 8	**Grade Level: 8** **Item Type:** Short Answer (2 points) **Testing Session:** 3 **Alignment to Assessment Target(s):** LS2 (5-**8**) SAE-**7** Given an ecosystem, trace how matter cycles among and between organisms and the physical environment (includes water, oxygen, food web, decomposition, recycling but **not** carbon cycle or nitrogen cycle). **Supporting information about the item design related to Unifying Theme, EK Statement, or other targets:** This item might be used to assess background content knowledge about ecosystems as an introduction to an ER task addressing another related target LS2 (5-8) INQ+SAE (abiotic and biotic factors). It is a basic concept related to EK statement LS2 (cycling of matter and flow of energy). **Depth of Knowledge:** Level 1 - Represent in words or diagrams a scientific concept or relationship

3A.3 Constructed response items (3-CR) worth 3 score points will include such things as: Representing in words or diagrams a scientific concept or relationship; Specifying or explaining the relationship between facts, terms, properties, or variables; Describing or explaining examples and non-examples of science concepts; Developing or critiquing a procedure according to specified criteria; Comparing or summarizing data; Using reasoning, planning, and evidence; Explaining thinking (beyond a simple explanation or using only a word or two to respond); Interpreting/analyzing information from a complex graph; Justifying a response; Identifying a research question with explanation or based on observations/data; Using concepts to solve non-routine problems/more than one possible answer; Developing a scientific model for a complex situation; Forming conclusions from experimental or observational data; Completing a multi-step problem that involves planning and reasoning; Providing an explanation of a principle or concept; Justifying a response when more than one answer is possible; Citing evidence and developing a logical argument for concepts; Conducting a designed investigation; Explaining phenomena in terms of concepts.

Sample: This picture (not included in test specifications) shows a pond ecosystem. Use this picture and what you know to answer these questions. 3. Which living thing in the pond system breaks down dead plants and animals? _____ How is this process of breaking down dead plants and animals beneficial to the pond system? _____ _____ _____ (Scoring guide: 3-points – Bacteria is identified in part one; interrelationships within the pond system are stated with no inaccuracies AND the concept of decomposition/cycling of matter is accurately stated and applied = 3 score points) Source: Adapted from NAEP Science grade 8	**Grade Level: 8** **Item Type:** Constructed Response (3-CR) **Testing Session:** 3 **Alignment to Assessment Target(s):** LS2 (5-8) INQ+SAE – 5 Using data and observations, predict outcomes when abiotic/biotic factors are changed in an ecosystem. LS2 (5-**8**) SAE-**7** Given an ecosystem, trace how matter cycles among and between organisms and the physical environment (includes water, oxygen, food web, decomposition, recycling but **not** carbon cycle or nitrogen cycle). **Supporting information about the item design related to Unifying Theme, EK Statement, or other targets:** This item might be used to assess understanding of systems (EK statement LS2) in an ER task *prior to asking students to use data to make a prediction* about a change in factors, such an increase in algae or decrease in bacteria in the system. **Depth of Knowledge:** Level 2- Specify and explain the relationship between facts, terms, properties, or variables (e.g., cause-effect)

3A.4 Constructed response items (4-CR) worth score points will include such things as: Representing in words or diagrams a scientific concept or relationship; Specifying or explaining the relationship between facts, terms, properties, or variables; Explaining examples and non-examples of science concepts; Using reasoning, planning, and evidence to support a response; Explaining thinking (beyond a simple explanation or using only a word or two to respond); Interpreting/analyzing information from a complex graph; Justifying a response; Using concepts to solve non-routine problems/more than one possible answer; Developing a scientific model for a complex situation; Forming conclusions from experimental or observational data; Completing a multi-step problem that involves planning and reasoning; Providing an explanation of a principle or a concept; Justifying a response when more than one answer is possible; Citing evidence and develop a logical argument for concepts; Explaining phenomena in terms of concepts.

Sample:
Plants and animals depend on each other for survival in many ways.

4. Use the chart to list ways that animals depend on plants. The first one is done for you.

Ways Animals Depend on Plants
1. for food
2.
3.

Select an animal from the list below and describe how it depends on a plant for its survival.

List of animals
Bird
Insect
Fish
Frog

One example of how an animal depends on plants for its survival is _____

_____

_____

(Scoring guide: 4-points– two additional plausible ways are listed – e.g., shelter, materials for building nests, for water inside the plant, etc.- AND a specific example is accurately stated to support the concept = 4 score points)

Grade Level: 4

Item Type: Constructed Response (4-CR)
Testing Session: 1 and 2

Alignment to Assessment Target(s):
LS2 (K-**4**) SAE –**6**
Describe ways plants and animals depend on each other (e.g., shelter, nesting, food)

Supporting information about the item design related to Unifying Theme, EK Statement, or other targets: This item might be used to assess understanding of relationships between plants and animals. It is a basic understanding linked to targets LS1 (K-4) SAE-2 (basic needs of plants and animals) and LS2 (k-4) SAE-5 (energy is needed for all organisms to stay alive and grow).

Depth of Knowledge: Level 2- Specify and explain the relationship between facts, terms, properties, or variables; Describe and explain examples and non-examples of science concepts

B. Science Scoring Guide Models

The Tri-State Science Assessment is designed to provide school-wide data on science curriculum and instruction and on individual student performance according to the grade span assessment targets. Because the assessment includes a variety of item types and reporting categories, several types of scoring guides will be employed. The chart below (Table 3.2) summarizes the scoring guides to be utilized.

Table 3.2 – Scoring Guides for Science

Item Type	Scoring Guide	Score Yield
Multiple Choice	Answer Key	0 or 1 point
Short Answer	Task-Specific Rubric	0, 1, or 2 points
Constructed Response (3-CR)	Task-Specific Rubric	0, 1, 2, or 3 points
Constructed Response (4-CR)	Task-Specific Rubric	0, 1, 2, 3, or 4 points
Extended Response Task for Inquiry	The combination of several Task-Specific Rubrics for each item type (SA and 3-CR) used	Approximately 16-18 total score points

Scoring Guide Models for task-specific rubrics are included on the following pages.

Anchor responses for all SA and CR items will be selected from field-testing to further illustrate the scoring rubric descriptors. It is expected that the contractor will work with the Tri-State science content committees to review and refine items, prompts, and tasks for the extended response to be field-tested and to assist with selection of anchor papers illustrating all score points and revisions of scoring rubrics.

C. Scoring Guide Models for Item Types/Analytic Rubrics

3C.1 Analytic Rubric Model for Short Answer Items

Score 0	Student makes no attempts to respond to prompt OR Student presents only random/unrelated information.
Score 1	Student may have included some, but not all details/evidence to support the concept OR Some information is correct, but some inaccuracies are present in the response OR The information provided is correct, but the response is incomplete.
Score 2	The response clearly and completely states main concept and/or supporting details/evidence or examples as required AND Response is completely accurate.

Note: Anchor responses will be selected from field-testing to further illustrate, specify, and revise these rubric descriptors.

3C.2 Analytic Rubric Model for Constructed Response (3-CR) Items

Score 0	Student makes no attempts to respond to prompt OR Student presents only random/unrelated information
Score 1	Student may have included some, but not all required details/evidence to support the concept OR Some information is correct, but some inaccuracies are present in the response OR The information provided is correct, but the response is incomplete
Score 2	Student includes the number of details/evidence or examples required, but some details/evidence are extraneous (see anchors) or repetitious as opposed to distinctly different examples/data (see anchors) OR Student may have included sufficient, complete, and accurate details/evidence or examples to support the statement of the concept, but the concept is missing or inaccurate OR Student accurately states the concept required, without sufficient support (incomplete or insufficient evidence or examples)
Score 3	All parts of the response clearly and completely states main concept, providing sufficient supporting details/evidence or examples as required AND Response is completely accurate.

Note: Anchor responses will be selected from field-testing to further illustrate, specify, and revise these rubric descriptors.

3C.2 Analytic Rubric Model for Constructed Response (4-CR) Items

Score 0	Student makes no attempts to respond to prompt OR Student presents only random/unrelated information
Score 1	Student may have included some, but not all required details/evidence to support the concept OR Some information is correct, but some inaccuracies are present in the response OR Some information is correct, but the response is incomplete
Score 2	Student includes the number of details/evidence or examples required, but some details/evidence are extraneous (see anchors) or repetitious as opposed to distinctly different examples/data (see anchors) OR Student may have included sufficient, complete, and accurate details/evidence or examples to support the statement of the concept, but the concept is missing or inaccurate OR Student accurately states the concept required, without sufficient support (incomplete or insufficient evidence/examples) OR One part of two required parts is complete and accurate
Score 3	One part of two required parts is complete and accurate AND … some of a second part required is correct, but incomplete or inaccurate
Score 4	All parts of the response clearly and completely state main concept, providing sufficient supporting details/evidence or examples as required AND Response is completely accurate.

Note: Anchor responses will be selected from field-testing to further illustrate and revise these rubric descriptors.

IV. Administration Guidelines for Science Assessment

A. Testing time – The science assessment will be designed as a "power test" rather than as a timed test. Students will be permitted to continue working as long as they are productively engaged, as determined by the test monitor. There will be three testing sessions of approximately 50-minutes each. Constructed response items and the extended response task for inquiry will require the majority of the testing time.

   Notwithstanding the above provisions, the science assessment will also include time for instructions and test logistics.

B. Accommodations – The states of New Hampshire, Rhode Island, and Vermont will provide a set of protocols for Accommodations for statewide testing. These protocols will be included in the final test specifications.

C. Readability/Language Use – The language used in test forms should be easily understood. Readability levels - *excluding the required use of specific and necessary science vocabulary* - should be approximately one year below grade levels assessed.

D. Oral reading of prompts –Test administrators may orally read parts of items, such as the prompt for the extended response task, or provide prompts on audiotape. However, no interpretations or explanations of the prompts are allowed.

E. Use of classroom materials and preparation of the classroom environment – The Tri-State partner states will establish a policy for use of classroom materials – such as classroom wall charts (e.g., periodic table, diagrams), classroom models, additional measuring tools and technology, etc. – during administration of the science assessment. These protocols will be included in the final test specifications to ensure equity for all students.

F. Other – Except as explicitly stated or prohibited in these test specifications, or elsewhere in state policies, regulations, or laws, schools may follow local policies, procedures, and protocols for test administration.

V. Overview of the Science Assessment Design

The Tri-State Science Assessment will be given over three testing sessions as described in this section. It is estimated that there will be between 7 and 10 test forms for each grade span. Only common items will contribute to individual student scores. Some matrix items will be included in school/district reporting to provide additional information relating to school/district science programs, and other matrix items will be used for equating and field test purposes. Each year, *approximately* 25% of test items and the ER task will be released, with annotated anchor papers and scoring guides for analysis and interpretation of school performance, and for professional development and instructional purposes.

Test developers will address the science test design, using a strategy and rationale for field testing extended response tasks for inquiry (session 3). Additionally, the states are interested in exploring ways to allow for local scoring of the inquiry session (perhaps beginning after year 1 of the operational test), for the elementary and middle school levels. (Note: Since local scoring of the Extended Response Task could impact the use of matrix-sampled generic equating items, it may be necessary to embed matrix items for equating inquiry within sessions 1 and 2.)

Table 5.1 shows an overview of the three testing sessions and the approximate number of score points that each session will generate. Sessions 1, 2, and 3 are then elaborated upon, with descriptions of common and matrix items.

Table 5.1 Overview of Test Design:
Testing Sessions and Individual Student Score Points Generated

Testing Session	Focus	Individual Score Points
1 and 2	**Science Domains –** Earth/Space Science Physical Science Life Science	**15/science domain** **= 45**
3	**Extended Response –** Inquiry + some science content	**16-18 total points** (1 ER task)

Sessions 1 and 2 (approximately 50 minutes each)

Sessions 1 and 2 will consist of a combination of multiple-choice (MC) and 4-point constructed response (CR) items measuring the three domains of science, as identified in the Tri-State Science Assessment Targets for Earth/Space, Physical Science, and Life Science. Items in sessions 1 and 2 will be grouped by science domain. Session 1 will begin with all of the Earth/Space Science items, followed by half of the Physical Science items. Session 2 will begin with the second half of the Physical Science items followed by all of the Life Science items.

To the degree possible, multiple items assessing the same assessment target should be developed keeping in mind connections to the Unifying Themes/Big Ideas identified for that target (e.g., Form & Function, Systems & Energy, etc.); and sets of items assessing multiple targets linked to the same Statement of Enduring Knowledge should attempt to make connections to the concepts in the EK Statement. (E.g., the PS3 EK Statement is: The motion of an object is affected by forces. The scenario(s) used for the set of PS3 items should collectively provide reinforcement of the EK statement.)

Use of item types and distributions for Sessions 1 and 2
There will be 12 common items producing 15 score points within each of the three science domains - 11 multiple choice (MC) and 1 constructed response (4-CR) item per science domain. Each student will complete a total of 6 CR items (worth 4 points) across the first two sessions - 3 common items and 3 matrix items. This distribution was determined to be necessary to preserve grouping items within each science domain. The proposed distribution of items is:

Session 1: Phys Science: 8MC + 1CR, 9MC + 1CR; Earth/Space Science 8MC + 1CR
Session 2: Earth/Space Science: 9MC + 1CR; Life Science 8MC + 1CR, 9MC + 1CR.

This design allows for 11 common and 6 matrix MC items per domain. The 6 MC matrix item slots are available for each purpose as follows: 2 MC for equating, 2 MC for content breadth (school/district reporting), and 2 MC items embedded for field test.

Session 3 (approximately 50 minutes)

Session 3 will focus on Inquiry and typically begin with a "hands-on/minds-on" activity (taking up to 30 minutes) that may or may not be a scientific investigation, depending on the grade level. The states recognize that there are time and materials constraints for conducting high-quality investigations in an on-demand setting; therefore, not all grade levels will conduct "full" investigations each year. Four Broad Areas of Inquiry (described in greater detail in Section VI) will be assessed:

- **Formulating Questions/Hypotheses**
- **Planning/Critiquing Investigations**
- **Conducting Investigations**
- **Developing/Evaluating Explanations.**

Session 3 will include a combination of 2-point short-answer (SA) and 3-point constructed-response (3-CR) items appropriate to the extended response (ER) task. Session 3 will also include about 10 minutes worth of matrix-sampled generic (meaning non-task specific) inquiry items that will be used to equate the inquiry performance year-to-year. Thirteen inquiry constructs (see page 65) will be used to design both the matrix equating items and common items for the Extended Response Task. *Developers are invited to propose alternative strategies for equating the ER tasks.*

Extended Response Performance Tasks for Session 3
At each grade span, students will be asked to apply their inquiry knowledge and skills to a longer, extended response performance task. Students will manipulate materials, objects, images, and data to solve scientific problems. Each grade span will have a slightly different focus for the ER task: grade 4 will *always* conduct a hands-on investigation; grade 8 will conduct a scientific investigation some years, and other years use data to make predictions, design or critique investigations, or develop explanations; high school *will not* conduct full investigations, but will focus on use of data and evidence to develop or critique explanations and investigations. A variety of approaches may be used to set the context for the ER tasks, especially at middle and high school levels. (See Sections VI and VII for general guidelines.) *Developers are encouraged to provide novel approaches to presenting engaging ER tasks, such as through the use of video, scenarios, simulations, case studies, etc.*

Table 5.2 provides a summary of matrix and common items for each session.

Table 5.2 Summary of Item Types, Estimated Times, and Sessions

Item Types	Estimated Time/Item	Session and Focus	Number of common Items per Student	Number of matrix Items per Student	Total items/student
MC (1 point)	1 minute	**1 and 2** Domains	11/domain = 33	6/domain [2 – equating 2 – field test 2 – breadth] = 18	51
Constructed response (4-CR)	8-minutes	**1 and 2** Domains	1/domain =3	1/domain =3	6
Short Answer (2 points)	1-2 minutes	**3** Inquiry constructs + some content	Approx 1-4; varies by grade span and investigation or non-investigation	?	Approx 1-4; varies by grade span and investigation or non-investigation
Constructed response (3-CR)	3-5 minutes	**3** Inquiry constructs + some content	Approx 5-7; varies by grade span and investigation or non-investigation	?	Approx 5-7; varies by grade span and investigation or non-investigation
Matrix items for equating Inquiry	10 minutes	**3** Inquiry constructs	NA	? item & item types still TBD for equating inquiry	?

Section VI - Tri-State Schema for Assessing Inquiry - Extended Response (ER) Task

The third testing session at each grade span will be devoted to assessing inquiry skills and knowledge. Session 3 will include an Extended Response (ER) Task, and will take up to 30 minutes for an activity and approximately 45 minutes to complete test items. This *set of items* – which *will* include a mix of short answer (SA) items and constructed response (CR) items – will be developed by applying combinations of 13 constructs related to 4 Broad Areas of Inquiry. (More detail about the 13 separate constructs and four Broad Inquiry Areas are on page 65.) **ER task items should not be interdependent, but can scaffold topics to get at higher order thinking.**

While the majority of the ER task's items will focus on assessing inquiry knowledge and skills, a small number of items may be included to assess related science concepts identified in other science assessment targets at the grade span.

The chart below describes a general overview of how each grade span's ER task will address the four broad areas of Inquiry.

Session 3: Approximately 45 minutes of testing, with 30 minutes for a hands-on/minds-on task*		
Grade 4	**Grade 8**	**Grade 11**
Grade 4 *will always have a hands-on investigation*, from question or prediction though collecting and interpreting data. The ER task will be aligned to the Tri-State Planning Guide for Investigations (included in these test specifications on pages 67-68).	The ER task will include (some) constructs *sampling all 4 Broad Areas* of Inquiry, however not necessarily sequentially. The ER task *may include a hands-on science investigation some years, but not every year*. If an investigation is developed, it should be closely aligned with the Tri-State Planning Guide for Investigations.	The ER task *will not necessarily include constructs from all 4 Broad Areas* of Inquiry and there will be no full investigation at this grade span. *The 4th broad area, Developing and Evaluating Explanations must be emphasized at this grade span.* ER tasks should be closely aligned with the Tri-State Planning Guide for NON-Investigations.

*Session 3 will include a 30-minute hands-on/minds-on task (e.g., viewing a video, conducting an investigation, making observations/collecting data, manipulating materials or data, reviewing data, etc.).

For the purpose of equating inquiry performance across years, approximately 10 minutes of testing time will be used for an additional number of matrix-sampled generic inquiry items (not specific to the ER task). These matrix items will not be used in determining individual student scores.

Understanding the Background for Development of Inquiry Constructs

The states began by identifying several essential questions for assessing inquiry skills and knowledge that were generic to all three grade spans. To answer these questions and thus determine the test design for the third testing session, 4 Broad Areas of Inquiry were identified. Then using readings and national science resources, each Broad Area was further defined with separate constructs for assessing inquiry. General guidelines, appropriate for each grade span, determine how a combination of items – a set of items linked to one Statement of Enduring Knowledge – will comprise the ER Task.

Essential Questions for Assessing Inquiry:

- What does it mean to know and do inquiry? (**Four Broad Areas of Inquiry** for instruction and assessment of inquiry were identified.)
- What constitutes evidence of knowing and doing inquiry? (Assessment items/tasks will be developed using combinations of the **13 constructs** below.)
- How can evidence be elicited from students? (**Grade-appropriate differences** demonstrate how these constructs will be assessed at different grade spans.)
- What valid inferences can we make from the evidence? (The Extended Response task will **link a set of items** –- to a Statement of Enduring Knowledge, INQ assessment Target(s), and Inquiry constructs so that reasonable inferences can be made about a student's and school's inquiry and content knowledge.)

Table 6.1: Tri-State Schema for Assessing Scientific Inquiry (with DOK levels for constructs)

Broad Areas of Inquiry to be Assessed	Formulating Questions & Hypothesizing	Planning and Critiquing of Investigations	Conducting Investigations	Developing and Evaluating Explanations
Constructs for each Broad Area of Inquiry (including intended DOK Ceiling Levels, based on Webb Depth of Knowledge Levels for Science – see also Section II) ***Inquiry Constructs answer the question: What is it about the broad area of Inquiry that we want students to know and be able to do?***	**1.** Analyze information from observations, research, or experimental data for the purpose of formulating a question, hypothesis, or prediction: **(DOK 3)** **1a.** Appropriate for answering with scientific investigation **1b.** For answering using scientific knowledge **2.** Construct coherent argument in support of a question, hypothesis, prediction **(DOK 2 or 3 depending on complexity of argument)** **3.** Make and describe observations in order to ask questions, hypothesize, make predictions related to topic **(DOK 2)**	**4.** Identify information/evidence that needs to be collected in order to answer the question, hypothesis, prediction **(DOK 2 – routine; DOK 3 non-routine/ more than one dependant variable)** **5.** Develop an organized and logical approach to investigating the question, including controlling variables **(DOK 2 – routine; DOK 3 non-routine)** **6.** Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation **(DOK 2)**	**7.** Follow procedures for collecting and recording qualitative or quantitative data, using equipment or measurement devices accurately **(DOK 1 – use tools; routine procedure; DOK 2 – follow multi-step procedures; make observations)** **8.** Use accepted methods for organizing, representing, and manipulating data **(DOK 2 – compare data; display data)** **9.** Collect sufficient data to study question, hypothesis, or relationships **(DOK 2 – part of following procedures)** **10.** Summarize results based on data **(DOK 2)**	**11.** Analyze data, including determining if data are relevant, artifact, irrelevant, or anomalous **(DOK 2 – specify relationships between facts; ordering, classifying data)** **12.** Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis **(DOK 3)** **13.** Communicate how scientific knowledge applies to explain results, propose further investigations, or construct and analyze alternative explanations **(DOK 3)**

General Schema for Assessing Inquiry: Extended Response (ER) Tasks

- **For all Extended Response (ER) inquiry tasks, developers will use a combination of constructs from the Four Broad Areas of Inquiry**. ER tasks for grades K-4 will be aligned with the Planning Guide for Investigations. ER tasks for grades 5-8 *must include* constructs from all 4 areas and may include an investigation (also closely aligned with Planning Guide for Investigations). ER tasks for grades 9-11 may or may not include constructs from all 4 Broad Areas of Inquiry, but *must emphasize* Developing and Evaluating Explanations, using constructs from the 4th column and the Planning Guide for NON-Investigations.
- **ER task focus**: Assessment targets designated as INQ (inquiry) will be selected as the focus for the ER inquiry task at a given grade span and address the related Statement of Enduring Knowledge within the *set* of ER items.
- **Statement of Enduring Knowledge**: Test developers are encouraged to use additional/multiple targets from the same Statement of Enduring Knowledge to develop ER tasks. To the degree possible, prompts for the ER task must support/focus on the content related to the Statement of Enduring Knowledge.
- **Differences across grade spans:** Grade 4 will always include an investigation. Grade 8 *may* include a hands-on investigation or an extended task using data collection and interpretation. Grade 11 will focus on Developing Explanations.
- **Depth of Knowledge**: Each assessment target has an intended Depth of Knowledge level as its "DOK ceiling" (highest level that may be assessed). The ER inquiry task should assess constructs at and below the intended Depth of Knowledge Level ceilings (see Section II for Webb's Depth of Knowledge Levels for Science).
  Depth of Knowledge NOTES: Assessing combinations of constructs will likely yield a higher DOK ceiling than assessing constructs separately. *If constructs are assessed in relation to each other (e.g., Formulating questions AND Planning Investigation 1+ 4+5+6; or all constructs for Conducting an Investigation 7+8+9+10) =* ***the ceiling DOK of Level 3.*** *If constructs for all 4 Broad Areas - developing a research question, designing and conducting an investigation, drawing conclusions, and communicating results are all included, this would be considered a* ***DOK ceiling of Level 4.*** *Level 4 will not be assessed due to time constraints.*
- **Assessment targets** designated as INQ (inquiry) may also be assessed in testing sessions 1 and 2, using items not related to the ER inquiry task.
- **Tri-State Inquiry Targets:** Any assessment target designated as an "inquiry target" (INQ in target code) may be selected as the focus of an ER task. "High Emphasis" targets (also listed on page 45) are also highlighted in Table 6.2.

Table 6.2: Tri-State Inquiry Assessment Targets Listed by Grade Span (Highlight = High Emphasis)

Elementary School	Middle School	High School
LS1-1 classify living things	LS1-1 biodiversity	LS1-1 cell organelles
PS1-1 classify objects	LS2-5 abiotic/biotic	LS2-3 ecosystem/energy flow
PS2-6 heat moving	LS4-10 human body	LS3-7 sexual reproduction
PS3-7 force/motion	LS4-11 genetics	LS3-8 cite evidence/ evolution
PS3-8 magnets	PS1-1 mass, volume, density	LS4 –9 humans – environmental factors
ESS1 – 1 classify earth materials	PS1-2 classify substances	PS1-1 chemical/phys properties
ESS1 – 2 water + earth materials	PS1-3 conservation of mass	PS2-6 energy flow/chemical. change
ESS1 – 4 wind, water, ice	PS2-7 heat transfer	PS3-8 force/motion
	PS3-8 force/motion	ESS1-1 geologic data
	ESS1-1 geologic evidence – plates	ESS1-4 dating methods
	ESS1-5 rock history	

Tri-State Planning Guide *for Investigations*

Two planning Guides are provided in these test specifications to assure that across grade levels, all inquiry constructs will consistently be assessed. The planning guide on page 68 was created to assist test developers in aligning hands-on investigations for all grade 4 Extended Response Tasks and for grade 8 if a hands-on investigation is included as the ER task. Generally, this is how the ER Investigation tasks will be organized; however, the exact order of items and inclusion of content-related items may be changed for a better flow of ideas.

Introduction - Prompt with Content-Related Question

[Item 1] After a brief introduction to the topic, a general content question is asked. This item may link to the EK statement in general, an aspect of the INQ target chosen as the focus of the task, or to another content-related assessment target for the grade span. This is a 2-point ***Short Answer*** (SA) – background knowledge/content question.

Formulating Questions and Hypothesizing (Inquiry Construct #1)

[Item 2] A scenario is provided to set the context for a specific problem or situation. An investigation question is posed and prediction or testable question is asked for. This is a 2-point ***Short Answer*** (SA) – Prediction/Hypothesis with short explanation.

Planning Investigations (Inquiry Construct #6)

[Item 3] A procedure is provided for the investigation. Materials are listed and a diagram of the investigation set-up may be included. This is a 2-point ***Short Answer*** (SA) —About nature of this investigation: provide reasoning for appropriateness of materials, tools, scale used, or procedures e.g., (information to be collected, variables, fair test, etc.)

Conducting Investigations (Inquiry Constructs #7, #8, #10)

Two items address conducting investigations.

[Item 4] A ***Constructed Response*** (3-CR) -- Data Collection and Data Representation—Student follows procedures provided to complete investigation and presents data/observations in labeled drawing(s), or chart/graph. A blank graph/chart template is provided for grade 4.

[Item 5] A 2-point ***Short Answer*** (SA)—Summarizes results based on data.

Developing and Evaluating Explanations (Inquiry Constructs #11, #12, #13)

Two items will be included. The "a" options are for grade 4 and the "b" options are for grade 8. Grade 8 assesses additional parts of constructs not included for grade 4.

[Item 6a] (Grade 4 only)—***Constructed Response*** — Determine connection between Prediction and data. Evidence is provided to justify explanation/conclusions.

[Item 6b] (Grade 8 only) — ***Constructed Response*** —Analysis of data—relevant, irrelevant, artifact given the concept, with example to support and Determine connection between Prediction and data. Evidence is provided to justify conclusions.

[Item 7a] (Grade 4 only) —***Constructed Response*** —What is a new question about this content that could be investigated further? Apply to related concept within Statement of Enduring Knowledge.

[Item 7b] (Grade 8 only) -- ***Constructed Response*** —propose new investigation with general procedures that extend application to related concept within Statement of Enduring Knowledge or suggest an alternative to conclusions drawn.

The template on the next page summarizes this information graphically.

Tri-State Planning Guide *for Investigations* (Extended Response)							
Item #	**General Description**	**Inquiry Construct OR Content**	**Broad Area of Inquiry OR Content**	**Item type**	**Grade Level**		
1	Introduction Prompt with Content Question	**Content:** Background Knowledge – facts, definitions, concepts	**Content:** Link to EK Statement, INQ target or another content-related target	SA	4	8	-
2	State prediction or question (gr 4)/ Hypothesis, prediction, or question (gr 8) w/explanation	**1.** Analyze information for the purpose of formulating a question, prediction or hypothesis	**Inquiry:** Formulating Questions and Hypothesizing	SA	4	8	-
3	Provide reasoning for tools, materials procedures, scale	**6.** Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation	**Inquiry:** Planning/Critiquing Investigations	SA	4	8	-
4	Collect and Display Data Procedures & set-up are provided (Template provided for graph or chart – grade 4)	**7.** Follow procedures for collecting and recording qualitative and quantitative data, using equipment and measurement devices accurately **8.** Use accepted methods for organizing, representing, and manipulating data	**Inquiry:** Conducting Investigations	3-CR	4	8	-
5	Summarize Results (no analysis)	**10.** Summarize results based on data	**Inquiry:** Conducting Investigations	SA	4	8	-
6a	Determine connections btw prediction & data, with evidence to justify explanation /conclusions	**12.** Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis	**Inquiry:** Developing and Evaluating Explanations	3-CR	4	-	-
6b	Analyze data for relevance; Determine connections btw prediction & data, with evidence to justify conclusions	**11.** Analyze data, including determining if data are relevant, artifact, irrelevant, or anomalous **12.** Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis	**Inquiry:** Developing and Evaluating Explanations	3-CR	-	8	-
7a	New question that could be investigated; Apply to related concept within EK Statement	**13.** Communicate how scientific knowledge applies to explain results, propose further investigations	**Inquiry:** Developing and Evaluating Explanations **with Content:** Link target concepts to EK statement	3-CR (scoring note – 2 points for inquiry and 1 point for content)	4	-	-
7b	Propose new investigation; Apply to related concept within EK Statement	**13.** Communicate how scientific knowledge applies to explain results, propose further investigations, or construct and analyze alternative explanations	**Inquiry:** Developing and Evaluating Explanations **with Content:** Link target concepts to EK statement	3-CR (scoring note – 2 points for inquiry and 1 point for content)	-	8	-
	Total Points Possible			17			

Tri-State Planning Guide *for NON- Investigations*

Two planning Guides are provided in these test specifications to assure that across grade levels, all inquiry constructs will consistently be assessed. The planning guide on page 70 was created to assist test developers in aligning NON-investigations for all grade 11 Extended Response Tasks and for grade 8 if a NON-investigation is included as the ER task. Generally, this is how the ER NON-Investigation tasks will be organized; however, the exact order of items and inclusion of content-related items may be changed for a better flow of ideas.

Introduction - Prompt with Content-Related Question

[Item 1] After an introduction to the topic/scenario/case study, a general content question is asked. This item may link to the EK statement in general, an aspect of the INQ target chosen as the focus of the task, or to another content-related assessment target for the grade span. This is either a 2-point ***Short Answer*** (SA) or ***Constructed Response*** (3-CR) background knowledge/content question.

Conducting Investigations (Inquiry Constructs #8 and #9); *with emphasis on* Developing and Evaluating Explanations (Inquiry Constructs #11 and #12)

[Item 2] Students will be asked to analyze information/data and determine connections between hypothesis/ prediction and data, providing evidence to justify their explanation/response. All constructs might not be addressed if the context doesn't lend itself to all aspects. The analysis of data might come later in the sequence of items, but is essential to the ER task. This is a ***Constructed Response*** (3-CR) — Determine connection between Prediction and data. Evidence is provided to justify explanation/conclusions.

Formulating Questions and Hypothesizing (Inquiry Construct #1 and #2 or #3)

[Items 3a or 3b] Students use data analysis to either formulate a research question/prediction/hypothesis OR construct an argument in support of a question/ prediction/hypothesis. This is a ***Constructed Response*** (3-CR) —only one of the options (3a or 3b) is used. Both 3a and 3b require support for explanations.

Planning/Critiquing Investigations (Inquiry Construct #4, #5, #6); Conducting Investigations (Inquiry Constructs #7 and #8)

[Items 4a or 4b] Students either design a proposed investigation OR critique an investigation provided. This is a ***Constructed Response*** (3-CR) —only one of the options (4a or 4b) is used. For example, option 4b might be used with 3b.

Conducting Investigations (Inquiry Constructs #10); *with emphasis on* Developing and Evaluating Explanations (Inquiry Constructs #12 and #13)

[Item 5] Students summarize results in order to either construct an alternative explanation or analyze an explanation. This item may come earlier in the sequence of items. This is a ***Constructed Response*** (3-CR)—using evidence to support response/explanation.

[Item 6] A content-related question allows for connections to other targets or EK Statement concepts. This is a ***Short Answer*** (SA) or ***Constructed Response*** (3-CR).

The template on the next page summarizes this information graphically.

Tri-State Planning Guide for NON-Investigations (Extended Response)							
Item #	**General Description**	**Inquiry Construct OR Content**	**Broad Area of Inquiry OR Content**	**Item type**	**Grade Level**		
1	Introduction Prompt with Content Question	**Content:** Background Knowledge – facts, definitions, concepts	**Content:** Link to EK Statement, INQ target or another content-related target	SA or 3-CR	-	8	11
2	Analyze information/data for relevance, etc.; Determine connections btw hypothesis/ predic-tion & data; give evidence to justify explanation/response	**8.** Use accepted methods for organizing, representing, and manipulating data **9.** (Is there) sufficient data to study question, hypothesis, or relationships **11.** Analyze data, including determining if data are relevant, artifact, irrelevant, or anomalous **12.** Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis	**Inquiry:** Conducting Investigations **Inquiry:** Developing and Evaluating Explanations	3-CR	-	8	11
3a	Formulate research question or hypothesis w/explanation using observations or information	**1.** Analyze information for the purpose of formulating a question, prediction /hypothesis **3.** Make and describe observations in order to ask questions, hypothesize, make predictions related to topic	**Inquiry:** Formulating Questions and Hypothesizing	3-CR	-	8	11
3b	Construct coherent argument in support of a question, hypothesis, prediction w/explanation using information provided	**1.** Analyze information for the purpose of formulating a question, prediction /hypothesis **2.** Construct coherent argument in support of a question, hypothesis, prediction	**Inquiry:** Formulating Questions and Hypothesizing	3-CR	-	8	11
4a	Design proposed investigation	**4.** Identify information/evidence that needs to be collected in order to answer the question, hypothesis, prediction **5.** Develop an organized and logical approach to investigating the question, including controlling variables **6.** Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation	**Inquiry:** Planning/Critiquing Investigations	3-CR	-	8	11
4b	Critique investigation design, including sufficiency and appropriateness of data presented	**6.** Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation **7.** Follow procedures for collecting and recording qualitative and quantitative data, using equipment and measurement devices accurately **8.** Use accepted methods for organizing, representing, and manipulating data	**Inquiry:** Planning/Critiquing Investigations **Inquiry:** Conducting Investigations	3-CR	-	8	11
5	Construct or analyze alternative explanation	**10.** Summarize results based on data **12.** Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis **13.** Communicate how scientific knowledge applies to explain results, propose further investigations, or construct and analyze alternative explanations	**Inquiry:** Conducting Investigations **Inquiry:** Developing and Evaluating Explanations	3-CR	-	8	11
6	Content Question applied to related EK concept	**Content:** Use related EK statement concept application or related concept	**Content:** Link to EK Statement, INQ target or another content-related target	SA or 3-CR	-	8	11
Total Points Possible				16-18			

Section VII – Applying the Tri-State Schema for Assessing Inquiry

Section VII of these test specifications includes three annotated ER task examples that apply the Tri-Sate Planning Guides in their design.

About the Sample ER task for Grade 4 – A Sample Investigation ER Task: "Up the Hill"

- This ER task is aligned to INQ Assessment Target PS3 (K-**4**)-INQ+SAE –**7.**

- All content knowledge questions relate to forces and motion concepts (PS3).

- **Important Note about paired and individual work:** While this grade 4 example includes a part for pairs of students to work together to collect data before working individually to analyze data and draw conclusions, *it is not intended* that all investigations for grades 4 and 8 must have this "cooperative" component. To the degree that it is possible to assess individual skills and knowledge with pairs working on some aspect of data collection or exploration of materials, the states would like it to be included in the ER task, as it models sound instructional practice. The states invite test developers to suggest ideas for joint and individual work (e.g., splitting the testing session into two parts, with the first part used to collect data to bring to part 2 where questions are answered individually; or collecting the first trial data together with the second trial data collected separately and used to support responses).

- Comments about how this ER task aligns to the Tri-Sate Planning Guide for Investigations:

  There are 7 items for this sample ER task; some items have two parts, such as making and explaining a prediction in question #2. An introductory scenario is provided as context for the investigation task, along with a diagram of the set-up and a template for the data table. The order of items follows the Guide's order of items. They correspond to the Planning Guide as follows: item 1 (content with explanations) –item 2 (prediction with explanation) –item 3 (fair test) –item 4 (record data) –item 5 (summarize results) – item 6a (use evidence to support/refute prediction) –item 7a (new question to test). All constructs related to Planning Guide items 1-2-3-4-5-6a-7a are assessed with the items in this task.

About the Sample ER task for Grade 8 – A Sample NON-Investigation ER Task: "Data Interpretation and Prediction"

- This ER task is aligned to INQ Assessment Target PS2 (5-**8**) INQ+SAE+POC – **7.**

- It includes additional content knowledge for Assessment Target ESS1 (5-**8**) SAE+ POC – **4,** which is from a different Statement of Enduring Knowledge, but related to heat and heat transfer concepts.

- Comments about how this ER task aligns to the Tri-Sate Planning Guide for Non-Investigations:

  There are 6 items for this sample ER task; a few have two parts, such as graphing and analyzing the data. Generally the order of items follows the Planning Guide, however not exactly. This task's items use the following order (as they correspond to the Guide) item 1 (1/content)-item 2 (4a/control variables)- item 3 (2/graph & analyze data) – item 4

(3b/predict and support) –item 5 (5/summarize data) – item 6 (6/related content). The change in order helped to facilitate a better flow of ideas using this data set. Almost all constructs related to the Planning Guide items 1-2-3b-4a-5-6 are assessed with the items in this sample task.

About the Sample ER task for Grade 11 – A Sample NON-Investigation ER Task: **"What's Causing the Changes to this Lake Huron Feeder Stream?"**

- This ER task is aligned to INQ Assessment Target LS2 (9-**11**) INQ+SAE –**3**.
- It includes additional content knowledge for Assessment Target LS2 (9-**11**) POC+ SAE –**4.**
- Comments about how this ER task aligns to the Tri-Sate Planning Guide for Non-Investigations:

  A case study with scenario and data set are provided as context for this task. There are 6 items for this sample ER task; students provide background content knowledge, graph data (with multiple variables), analyze and interpret data, make predictions, design an investigation to test the hypothesis, and speculate about how to reverse trends in data.

  The order of items follows the Planning Guide format. This task's items use the following order (as they correspond to the Guide) item 1 (content with explanations) – item 2 (graph & analyze data) -3a (formulate prediction)-4a (design investigation to test hypothesis) –5 (support a conclusion) –6 (extend thinking with related content knowledge.

  Almost all constructs related to the Planning Guide items are assessed with the items in this sample task.

A. Developing Extended Response (ER) Tasks as Hands-On Investigations

The purpose of this Section of the science test specifications is to provide a sample of what grade 4 investigations might look like, using the Tri-State Planning Guide for Investigations as a model for ER task design.

Sample Grade 4 Investigation - Force and Motion Assessment

In this investigation, students are asked to find out how the force needed to move objects up a ramp (hill) change when the ramp gets steeper.

Materials: adjustable ramps and ladders for changing heights of ramps, rulers for measuring heights, washers for adding weight, string, paper clips for attaching washers to strings, and small cars.

Grade Level: 4
Extended Response Task: Up the Hill – Force & Motion Assessment
Item Types: SAs and 3-CRs

Testing Session: 3

Alignment to INQ Assessment Target(s):
FOCUS: PS3 (K-**4**)-INQ+SAE –**7**
Use data to predict how a change in force (greater/less) might affect the position, direction of motion, or speed of an object (e.g., ramps and balls).

Depth of Knowledge: Levels 1, 2, and 3

Source: Adapted from Barre Town School District, "Up the Hill"

Sample Elementary School ER Task: "Up the Hill"

Part 1: In part 1, you will answer some questions about what you have learned regarding forces and motion. Read the questions carefully and write down your answers.

1a) A heavy truck and a light car are stopped on a road. Explain what is needed to get the vehicles to move.

_____

_____

1b) Explain the relationship between the weight of each of the vehicles, the truck and the car, and what is needed to get each vehicle to move.

_____

_____

_____

Part 2: In part 2, you will work (with a partner) to investigate the effects of force on a moving car.

"Up the Hill"

Alex and Mia were in their family's car one day, driving to a favorite summer campground. When the car started to travel up a hill, they noticed that the engine seemed to get louder. Mia thought about the force and motion ideas that her class investigated before school closed for the summer. She wondered if the engine got louder because the car needed more force to go up the hill. When she shared her ideas with Alex, he said, "I think you have a good prediction Mia, but we can't test it because experimenting with a car engine would be too dangerous."

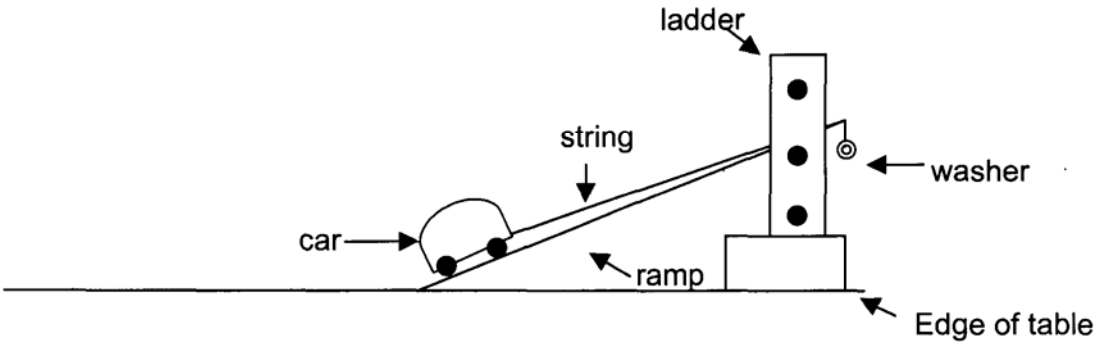
Mia agreed, but she added, "We could experiment with a toy car as a model and see if the amount of force needed to move the car up a ramp **changes** when you make the ramp (hill) steeper. Since Alex and Mia will be camping for the next couple of weeks, they are depending on you to complete the investigation for them.

Your investigation question is: **Does the force needed to move a toy car up hill change when the hill gets steeper?**

Look at the equipment that is in front of you. The drawing below shows how you will set up the equipment to investigate the force needed to move the vehicle up the ramp. You will change the ***amount of force*** *by changing the number of washers attached to the string. More washers on the string will provide more force to pull the car up the ramp.*

You will use these materials to try out some ideas that you have about using force to move vehicles up the hill.

Materials for Ramp Investigations



2) Predict how the amount of **force** changes if you need to move the vehicle up a steeper ramp (or hill)? **Explain** your prediction.

_____

_____

Part 3: You will now conduct an experiment to test your prediction. Use the materials in front of you and the drawing to set up your investigation.

Read over the steps that you will follow:
Your investigation will test the force needed for three different ramp heights.

1. Measure the first height of the ramp and record it on your recording sheet.
2. Add force (washers) to the end of the string until your car begins to move. Under Trial 1, record the number of washers (force) needed to move the car. Then, remove all washers.
3. Keep the height of the ramp the same and repeat the same steps. Under Trial 2, record the number of washers (force) needed to move the car. Then, remove all washers.
4. Now change the height of the ramp and repeat the same steps. Under Trial 1, record the number of washers (force) needed to move the car at this ramp height. Then, remove all washers.
5. Continue until you have tested 3 different ramp heights.

3) What will you be doing to make sure that your experiment is a fair test?

_____

_____

4) Record data from your experiment here.

Recording Sheet for Uphill Vehicle Experiment

Height of the Ramp	**Amount of Force Needed** Trial 1	**Amount of Force Needed** Trial 2

Part 4 (Individual Work): Now you will be asked to use the data that you collected when you investigated the relationship between the steepness of the hill and the amount of force needed to move a vehicle up the hill. Use your data from your investigation to answer the following questions.

5) Look at the information in your chart and summarize your results. What did you discover about the force needed to move a vehicle uphill when you make the hill steeper? **Use data from your chart to help you explain.**

_____

_____

6) Look back at your prediction (in question 2). Explain how the data either supported or did not support your prediction.

_____

_____

_____

7) How can you use what you learned about moving an object up a hill to learn more about force and motion? Write a new question you can test with an investigation about the amount of force needed to move an object in another way? Use what you learned from your experiment to **explain the reason for your answer.**

_____

_____

_____

_____

B. Developing Extended Response (ER) Tasks as NON-Investigations

The purpose of this Section of the science test specifications is to provide a sample of what grade 8 NON-investigations might look like, using the Tri-State Planning Guide for NON-Investigations as a model for ER task design.

Sample Grade 8 Investigation – "Data and Interpretation and Prediction"

In this extended response task, students are asked to graph data, interpret findings, and critique the experimental design and conclusions. Students are also asked to apply results to make predictions about real-world applications and explain phenomena related to heat transfer and differential heating.

Grade Level: 8
Extended Response Task: "Data and Interpretation and Prediction" **Assessment**
Item Types: SAs and 3-CRs

Testing Session: 3

Alignment to INQ Assessment Target(s):
FOCUS: PS2 (5-**8**) INQ+SAE+POC – **7**
Use data to draw conclusions about how heat can be transferred (convection, conduction, or radiation)
RELATED CONTENT: ESS1 (5-**8**) SAE+ POC –**4**
Explain the role of differential heating or convection in ocean currents, winds, weather and weather patterns, atmosphere, or climate.

Depth of Knowledge: Levels 1, 2, and 3

Source: Adapted from *Science Scope*, January 2005 (page 37), "Data and Interpretation and Prediction"

Sample Middle School ER Task
"Data Interpretation and Prediction"

In this task, you will answer questions to show what you have learned about heat, heat transfer, and the interpretation of data. Read the questions carefully and answer them completely, using data or examples to support your responses.

Part 1:
You have learned that there are three methods of heat transfer: conduction, convection, and radiation.

1. If your teacher instructs you to heat a sample of soil using a heat lamp, which method of heat transfer will you be using? _____

Use what you know about energy to show or describe how the heat lamp will heat the soil sample. You may use descriptions and diagrams in your response.

_____

_____

Item type: SA (2 points)

Background Knowledge about Content
PS2 (5-**8**) INQ+SAE+POC – **7**
Use data to draw conclusions about how heat can be transferred (convection, conduction, or radiation)

DOK 1 - Recall or recognize a fact, term, definition, simple procedure (such as one step), or property Represent in words or diagrams a scientific concept or relationship
DOK 2 - Specify and explain the relationship between facts, terms, properties, or variables

Questions 2 and 3 refer to an experiment your teacher asks you to perform. You are to compare the heating rate of soil to the heating rate of water.
To do this, you are given the following materials:

2 heat lamps	1 sample of soil
2 bins	1 sample of water
2 thermometers	1 timer

2. There are many experimental variables that must be controlled in order to perform this experiment accurately. Name three of these variables and explain how and why you would control each one.

_____

_____

_____

Item type: 3-CR

Planning & Critiquing Investigations
 # 4. Identify information/evidence that needs to be collected in order to answer the question, hypothesis, and prediction
#5. Develop an organized and logical approach to investigating the question, including controlling variables
 # 6. Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation

DOK 2 – Make a decision as to how to approach the problem

In this experiment, you are instructed to heat a sample of soil and a sample of water with heat lamps and then measure the temperature of each sample several times.

Your prediction was that the temperature of the water would rise at a faster rate than the rate that the temperature of the soil would rise.

Suppose that the experiment yielded the results shown in the table below.

Time (min)	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Soil Temp (°C)	20	21	22.5	24	26	27.5	29.5	30.5	32
Water Temp (°C)	20	21.5	23	23.5	24	25.5	26	27.5	28.5

2. Use the grid to graph the data from the table. Then analyze the results to determine whether the data supports or refutes your prediction. Be sure to label each axis clearly.

Use evidence to explain whether the data supported or refuted your prediction.

_____

_____

Part 2

Item type: 3-CR

Conducting Investigations
 # 8. Use accepted methods for organizing, representing, and manipulating data
Developing Explanations
 # 11. Analyze data, including determining if data are relevant, artifact, irrelevant, or anomalous

4. At a beach that has white sand, you measure the temperature of the sand and the temperature of the seawater at 9:00 a.m. You find that both have a measure of 16°C. If it is clear and sunny all morning, what do the data from the classroom experiment predict about the temperature of the white sand compared to the seawater at noon?

_____

_____

Explain your answer.

_____

_____

Item type: 3-CR

Formulating Questions & Hypothesizing
 # 1. Analyze information for the purpose of formulating a question, prediction /hypothesis
 # 2. Construct coherent argument in support of a question, hypothesis, and prediction

DOK 2 or 3 depending on complexity of argument
DOK 2 – Specify and explain the relationship between facts, terms, properties, or variables
DOK 3 – Form conclusions from experimental or observational data; justify a response; Explain thinking (beyond a simple explanation or using only a word or two to respond)

5. Summarize the results from the classroom data. Then explain why a prediction based on this data *might be* wrong.

_____

_____

_____

Item type: 3-CR

Developing and Evaluating Explanations
Inquiry Construct:
 # 10. Summarize results based on data
#12. Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis

DOK 2 - Interpret information from a simple graph; Specify and explain the relationship between facts, terms, properties, or variables
DOK 3 – Form conclusions from experimental or observational data; justify a response; Explain thinking (beyond a simple explanation or using only a word or two to respond)

6. Use what you know and what you have learned from this data analysis to explain how the differential heating of water can affect the climate near a large body of water.

_____

_____

Item type: 3-CR

Content Alignment: ESS1 (5-**8**) SAE+ POC –**4**
Explain the role of differential heating or convection in ocean currents, winds, weather and weather patterns, atmosphere, or climate.

DOK Level 3: Provide an explanation of a principle; Explain phenomena in terms of concepts

C. Developing Extended Response (ER) Tasks as NON-Investigations

The purpose of this Section of the science test specifications is to provide a sample of what grade 11 NON-investigations might look like, using the Tri-State Planning Guide for NON-Investigations as a model for ER task design.

Sample Grade 11 Investigation: "What's Causing the Changes to this Lake Huron Feeder Stream?"

In this extended response task, students are asked to use a case study of a stream that feeds Lake Huron to graph and analyze data, make interpretations and predictions, and make connections to real-world phenomena related to causes of environmental disturbances and their effects on ecosystems.

Grade Level: 11
Extended Response Task: "What's Causing the Changes to this Lake Huron Feeder Stream?" **Assessment**
Item Types: SAs and 3-CRs

Testing Session: 3

Alignment to INQ Assessment Target(s):
FOCUS: LS2 (9-**11**) INQ+SAE -**3**
Using data from a specific ecosystem, explain relationships or make predictions about how environmental disturbance (human impact or natural events) affects the flow of energy or cycling of matter in an ecosystem

RELATED CONTENT: LS2 (9-**11**) POC+ SAE –**4**
Trace the cycling of matter (e.g., carbon cycle) and the flow of energy in a living system from its source through its transformation in cellular, biochemical processes (e.g., photosynthesis, cellular respiration, fermentation)

Depth of Knowledge: Levels 1, 2, and 3

Source: Adapted from *Science IV: Essential Interactions*, (Chapter 4, pages 112- 151: Sustaining Aquatic Habitats) Centre Pointe Learning, Inc., 2001, and *Science Exemplars*, Preview Kit (pages 1-14), "What's Causing the Changes in pH?" www.exemplars.com.

Sample High School ER Task:
"What's Causing the Changes to this Lake Huron Feeder Stream?"

In this task, you will answer questions to show what you have learned about ecosystems, the flow of energy and cycling of matter in ecosystems, and the analysis and interpretation of data. Read the questions carefully and answer them completely, using data or examples to support your responses.

Part 1:
You have learned that there are many biotic and abiotic factors that affect the flow of energy and cycling of matter in ecosystems.

1. Use what you know about the flow of energy in an ecosystem to explain the specific concepts and processes that allow water plants to survive and thrive in freshwater ecosystems. Be sure to start with the energy source and trace it through its major transformations, including photosynthesis and cellular respiration. You may use diagrams with explanations in your response.

_____

_____

_____

Item type: CR (3 points)

Background Knowledge about Content
LS2 (9-**11**) POC+ SAE –**4**
Trace the cycling of matter (e.g., carbon cycle) and the flow of energy in a living system from its source through its transformation in cellular, biochemical processes (e.g., photosynthesis, cellular respiration, fermentation)

DOK 1 - Recall or recognize a fact, term, definition, simple procedure (such as one step), or property Represent in words or diagrams a scientific concept or relationship
DOK 2 - Specify and explain the relationship between facts, terms, properties, or variables

Case Study of a Lake Huron Feeder Stream*

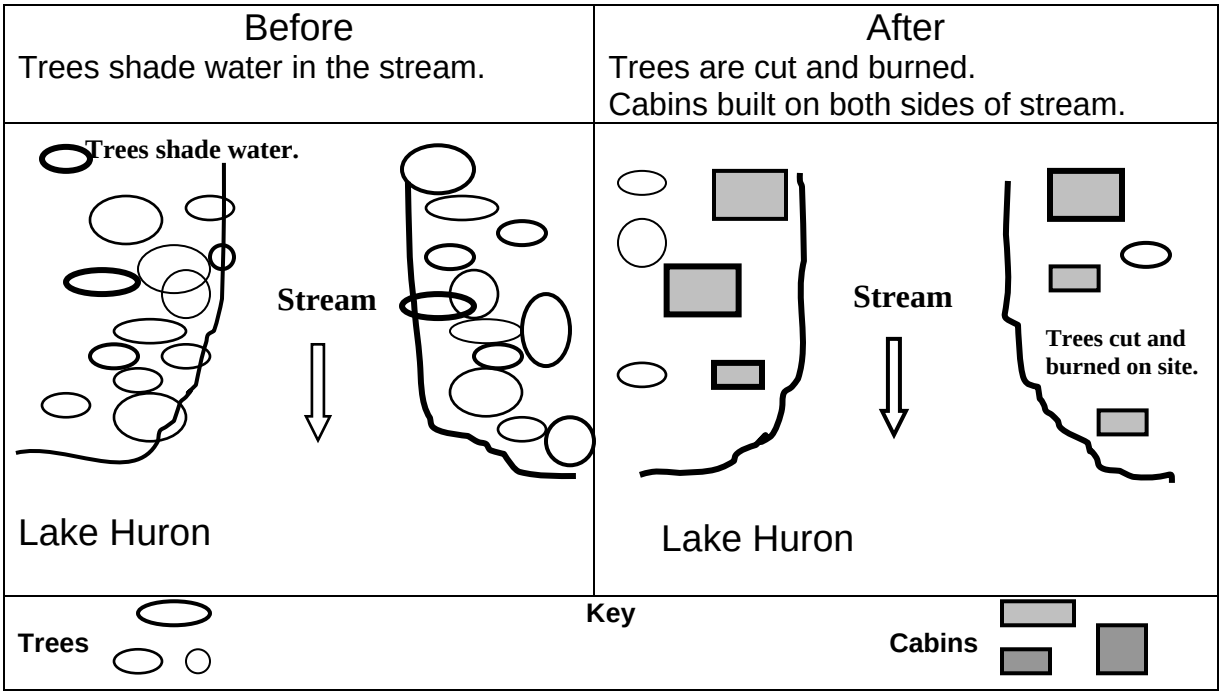
The Great Lakes are the largest surface freshwater system on Earth, covering an area of approximately 244,000 square km. These five lakes provide drinking water and water for industrial use, transportation, fishing, and recreational activities. Lake Huron is one the Great Lakes; it borders the state of Michigan and the province of Ontario, Canada.

At one time, a stream flowing into Lake Huron was a prime location for fishing and boating. This made it a perfect location for development, as well. A developer cleared most of the trees to make room to build cabins on both sides of the stream. Many of these trees that once provided shade for the water in the stream, were stacked in piles and burned.

Plant growth in the stream soon increased, which made boating increasingly difficult. The developer attempted to address the problem by spraying herbicides on the plants. Many plants died, but soon fishing in the stream was also negatively affected.

The diagram below shows before and after views of the stream.

* Source: *Science IV: Essential Interactions,* (pages 126-128- Sustaining Aquatic Habitats) Centre Pointe Learning, Inc., 2001



Numerous tests were conducted in order to collect data before and after the developer began to remove trees along the stream banks in order to build cabins. The data from the results of those tests are presented in the table below.

Table 1: Test results before and after deforestation along the Lake Huron feeder stream

(Data Source: *Science IV: Essential Interactions,* (page 128 - Sustaining Aquatic Habitats) Centre Pointe Learning, Inc., 2001

Data collection Tests	Before removal of trees	1 year after tree removal	5 years after tree removal
Mean temperature (°C) in June	12	15	17
Mean water depth	1.5	1.2	1.1
Mean dissolved oxygen (ppm) in June	7.8	5.5	5.1
Phosphates (ppm)	0.2	3.6	2.5
Nitrates (ppm)	0.1	1.2	0.6
Inorganic solids (g/L)	0.1	1.6	1.3

2a. Use the grid below to construct a graph (bar or line) of the data from the table to show changes in mean dissolved oxygen, phosphate, and nitrate levels. You will then analyze the data in order to explain how and why these levels changed. Be sure to label each axis and each data set clearly.

2b. Researchers did not find evidence of fertilizer use of other non-point source pollution during this time that might have contributed to the changes in nitrate and phosphorous levels. Use the data from your graph and data from Table 1 to describe how the nitrate and phosphorous levels in the stream changed over this period of time. Provide an explanation about what probably caused these changes to occur after the trees were cut down.

3. Using the data provided, predict what effect these changes might have on plants and animals in the stream ecosystem during the next five years.

Item type: 3-CR

Formulating Questions & Hypothesizing
 # 1. Analyze information for the purpose of formulating a question, prediction /hypothesis
 # 2. Construct coherent argument in support of a question, hypothesis, and prediction

DOK 2 or 3 depending on complexity of argument
DOK 2 – Specify and explain the relationship between facts, terms, properties, or variables
DOK 3 – Form conclusions from experimental or observational data; justify a response; explain thinking (beyond a simple explanation or using only a word or two to respond)

4. Design an investigation to test your prediction that includes: your hypothesis; data that needs to be collected; tools, methods, and procedures for controlling variables and collecting sufficient data; and a rationale for this experimental design. You may use diagrams with explanations in your response.

Item type: 3-CR

Planning & Critiquing Investigations
#4. Identify information/evidence that needs to be collected in order to answer the question, hypothesis, and prediction
#5. Develop an organized and logical approach to investigating the question, including controlling variables
#6. Provide reasoning for appropriateness of materials, tools, procedures, and scale used in the investigation

DOK 2 – Make a decision as to how to approach the problem
DOK 3 –Explain thinking (beyond a simple explanation or using only a word or two to respond); Identify research questions and design investigations for a scientific problem

5. Using the data provided, explain how changes in water depth, temperature, and dissolved oxygen relate to one another.

_____

_____

_____

Item type: 3-CR
Developing Explanations
10 Summarize results based on data
12 Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis
13 Communicate how scientific knowledge applies to explain results, propose further investigations, or construct and analyze alternative explanations

DOK 2: Make a decision as to how to approach the problem; Specify and explain the relationship between facts, terms, properties, or variables
DOK 3: Explain thinking (beyond a simple explanation or using only a word or two to respond); Form conclusions from experimental or observational data; justify a response;

6. Speculate about what things might reverse these changes in the stream and how they might impact the ecosystem.

_____

_____

_____

Item type: 3-CR
Developing Explanations
10 Summarize results based on data
12 Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis
#13 Communicate how scientific knowledge applies to explain results, propose further investigations, or construct and analyze alternative explanations

DOK 2: Make a decision as to how to approach the problem; Specify and explain the relationship between facts, terms, properties, or variables
DOK 3: Explain thinking (beyond a simple explanation or using only a word or two to respond); Form conclusions from experimental or observational data; justify a response.

Item type: 3-CR

Content Alignment: LS2 (9-**11**) INQ+SAE **-3**
Using data from a specific ecosystem, explain relationships or make predictions about how environmental disturbance (human impact or natural events) affects the flow of energy or cycling of matter in an ecosystem.

DOK Level 3: Provide an explanation of a principle; Explain phenomena in terms of concepts

VIII. Summary of Rationales for Test Components and Bidder Flexibility

Below is a summary of the components of the test specifications for the Tri-State Science assessment and bidder requirements and flexibility related to each component.

Table 7.1: Test Components, Rationales, and Bidder Requirements and Flexibility

Components of Science Test Specifications	Rationale	Bidder Flexibility
Reporting Categories by Science Domains and Inquiry (See Sections I and II and Tri-State Assessment Targets)	An overall score in science is supported by additional reporting categories for each domain of science and in inquiry. Information derived from extended response items will provide valuable instructional/curricular feedback.	• Bidders should provide methods and formats for reporting individual, school, district, and state score reports based on common items. Does the contractor have ideas about how to interpret score reports in relation to EK Statements within science domains, for example? • Bidders may propose strategies for using additional matrix items for school/district and state reports so that a broader view of science programs and/or analysis of achievement of science literacy might be provided.
Distribution of Emphasis across domains of science and Inquiry, based on Balance of Representation (See Section II)	The Distribution of Emphasis was established to provide a distribution of items across a form that would result in stable forms year-to-year AND reflect recommended assessment emphasis in relationship to assessing the 3 domains of science and inquiry.	• Some aspects of each Assessment Target should be assessed every year. • The partner states would not expect that the number of items/score points across individual Assessment Targets be rigidly adhered to from year to year, but does expect that the relative emphasis be attended to in sampling science domains and inquiry. • Contractors should provide strategies that consider the maintenance of Distribution of Emphasis for both item types and score points across science domains and inquiry.
Depth of Knowledge Levels and "ceilings" (See Section II)	Including items that assess the cognitive range of assessment targets is designed to potentially assure that: 1) The assessment as a whole will not be too difficult; 2) Important information about student learning along the achievement continuum will not be lost.	• The general alignment protocol for this aspect is that an assessment target should ***not*** be assessed above its "ceiling," and to the extent possible at the "ceiling" and at least one level below the "ceiling." (See pages 50-51) • Bidders should include strategies to address the distributions across the DOK levels and/or indicate any concerns or limitations that may arise from this specification. • Bidders should be prepared to work with the Tri-State content committee to solidify the Depth of Knowledge descriptors for science, Target "ceilings," and appropriate item examples for each DOK level.

Components of Science Test Specifications (continued)	Rationale (continued)	Bidder Flexibility (continued)
Intent of use of Conjunctions "and" & "or" and use of "i.e." and "e.g." in assessment targets. (See Section I of Test Specifications and Tri-State Science Assessment Targets document)	The "ands" and "ors" were established to guide test developer's sampling of assessment targets, as it reflects the prioritization that occurred throughout the assessment target development process. The use of "e.g.," is to suggest possible appropriate item focus; whereas the use of "i.e." *limits the content sampling* to those specific examples.	• Every assessment target should be assessed by at least 1 item (common items or matrix items for breadth) every year. • Assessment target aspects may be combined within the same item. (For example, an item might ask for stating a concept or explanation and use of evidence to support the concept/explanation, in the same CR item.) • To the extent possible, aspects of assessment targets connected with "and" should be assessed every year. (An analysis of Tri-State Science targets with "and" is available upon request.) • Elements of assessment targets connected with "or" can be rotated across years or used as additional matrix items for greater breadth of content assessed. • Targets with "e.g." have some flexibility in item content applications.
Item Types and Scoring Guides ➢ Multiple Choice (1 point) ➢ Short answer (2 points) ➢ Constructed Response (3 points) ➢ Constructed Response (4 points) ➢ Extended Response Task for Inquiry (16 – 18 points assessed with a mix of SA and 3-CR items) (See Sections III, V, and VI)	To provide a variety of item formats that include opportunities to demonstrate understanding of science concepts, apply skills and knowledge of inquiry, and apply reasoning and analysis skills in specific science contexts/real-world applications.	• All items used for individual scores are common items. • Sessions 1 and 2 will utilize MC and 4-CR items. Session 3 will utilize SA and 3-CR items. • Potential bidders should provide a strategy (e.g., using field test responses) for adapting/ modifying analytic rubric scoring models to assess each of these item types. Each SA and CR item will include an item-specific analytic scoring guide that includes a detailed, task-specific description of required performance at each score point. • Bidders should provide actual field test data, a strategy, and rationale for selecting anchor papers for scoring SA and CR items. • Bidders should provide a method for involving state content committee in the item review process.
Extended Response Task for Inquiry (See Sections VI and VII)	To provide opportunities for students to apply skills and knowledge of inquiry, reasoning, and analysis skills in specific science contexts/real-world applications, and to encourage more use of hands-on investigation as a teaching methodology. Thirteen inquiry constructs and 2 Planning Guides have been created to guide test and item development.	• Bidders should provide a method for assuring alignment of ER tasks with the Tri-State Planning Guide for Investigations, NON-Investigations, and other specific guidelines provided. (See pages 69-70) • Bidders may provide an alternative strategy for equating inquiry assessment year-to-year. • Bidders may want to suggest a strategy for local scoring of the ER task after year one of the operational test. • Bidders are invited to suggest ways that both paired and individual aspects are possible.

Appendix A:
Rationale for Addressing Unifying Themes/Big Ideas of Science in the Tri-State Science Assessment (K. Hess, working draft 2005)

The conceptual approach of the Tri-State Science Assessment is supported by recommendations of the National Research Council (NRC) for designing science assessments to satisfy the No Child Left Behind Act. The NRC, located in Washington, is a division of the congressional chartered National Academies, which provides research for the government, scientists, engineers, and the public. A recent article, "NRC Weighs in on States' Science Assessments" (www.edweek.org, published: July 13, 2005) describes a committee of testing experts and university researchers convened by the research council to produce the report, 'Systems for State Science Assessment.' The authors of the report say "the tests should be built around 'organizing principles' or "big ideas" of science, such as evolution and molecular theory, to give students a stronger sense of how different aspects of the discipline connect."

Organizing curriculum and assessment around the unifying themes, or big ideas, of science is not a new idea. References to big ideas of science appear in many national science standards documents written in the 1990s. The American Association for the Advancement of Science describes it this way: "Some powerful ideas often used by mathematicians, scientists, and engineers are not the intellectual property of any one field or discipline. Indeed, notions of system, scale, change and constancy, and models have important applications in business and finance, education, law, government and politics, and other domains, as well as in mathematics, science, and technology. These common themes are really ways of thinking rather than theories or discoveries" (AAAS, 1993, page 261).

Unifying Themes represent the key organizing concepts that pervade science education, crossing traditional science domain boundaries and making up the inquiry tools that scientists use to better investigate and understand phenomena (*NAEP Science Framework*, 1994). Statements of Enduring Knowledge represent the fundamental knowledge of the domains of science. "They are called 'enduring' because they contain essential ideas that students need to internalize and retain in order to achieve science literacy" (*Grade Expectations for VT's Framework of Standards and learning Opportunities*, 2004). The Tri-State Science Assessment Targets integrate six Unifying Themes/Big Ideas of Science with Statements of Enduring Knowledge for the domains of Life Science, Earth/Space Science, and Physical Science. The Unifying Themes/Big Ideas listed below are further described on the pages that follow.

- **Scientific Inquiry**
- **Nature of Science**
- **Systems and Energy**
- **Models and Scale**
- **Patterns of Change**
- **Form and Function**

"Scientific Inquiry (INQ) is more complex than popular conceptions would have it. It is, for instance, a more subtle and demanding process than the naive idea of "making a great many careful observations and then organizing them." It is far more flexible than the rigid sequence of steps commonly depicted in textbooks as "the scientific method." It is much more than just "doing experiments," and it is not confined to laboratories. More imagination and inventiveness are involved in scientific inquiry than many people realize, yet sooner or later strict logic and empirical evidence must have their day." (AAAS, 1993)

The Inquiry focus in the Tri-State Assessment will be on the ability to question, hypothesize, predict, design and critique investigations, conduct investigations, use science tools and techniques, collect, organize, and interpret/analyze data, use evidence to draw conclusions, develop explanations, and communicate understanding.

Nature of Science (NOS) "Generalizations about how the scientific enterprise operates would be empty without concrete examples. Consider, for example, the proposition that new ideas are shaped by the context in which they are conceived; are often rejected by the scientific establishment; sometimes spring from unexpected findings; and usually grow slowly, through contributions from many different investigators. Without historical examples, such generalizations would be no more than slogans, however well they might be remembered." (AAAS, 1993)

The Nature of Science focus in the Tri-State Assessment will be on the use of tools and technology, how fundamental theories change when applying new evidence and reasoning, how scientists build on the work of others, and attitudes and dispositions of science (e.g., avoiding bias, divergent ideas, healthy skepticism).

Systems and Energy (SAE) "One of the essential components of higher-order thinking is the ability to think about a whole in terms of its parts and, alternatively, about parts in terms of how they relate to one another and to the whole… If these can be specified quantitatively, a computer simulation of the system might be run to study its theoretical behavior, and so provide a way to define problems and investigate complex phenomena" (AAAS, 1993). The concept of energy, which cuts across all fields of the natural sciences and technology, is often used to analyze how systems function. As described in *Benchmarks for Science Literacy*, "Most of what goes on in the universe--from exploding stars and biological growth to the operation of machines and the motion of people--involves some form of energy being transformed into another." (AAAS, 1993)

The Systems and Energy focus in the Tri-State Assessment is on order & organization, interactions, interdependence, equilibrium, energy transfer, and cycles.

Models and Scale (MAS) "Models can be physical, mathematical, or conceptual. They are very effective tools for learning about the things they are intended to resemble. Physical models (such as model rockets) are the most obvious to children. Whether models are physical, mathematical, or conceptual, their usefulness as an instructional device lies in suggesting how things either do work or might work. The more sophisticated concept has to do with the effect of changes in scale. Specifically, the way things work may change with scale." University of Texas at Austin. Mission to Mars: Project Based Learning [On-line] Available: http://www.edb.utexas.edu/missiontomars/unify.html).

The Models and Scale focus in the Tri-State Assessment is on evidence and explanations through models, proportions, magnitude, relationships, and relativity.

Patterns of Change (POC) "Much of science and mathematics has to do with understanding how change occurs in nature and in social and technological systems, and much of technology has to do with creating and controlling change. Constancy, often in the midst of change, is also the subject of intense study in science…. Somewhat different aspects of constancy are described by the terms stability, conservation, equilibrium, steady state, and symmetry. These various ideas are interrelated in some subtle ways." (AAAS, 1993)

The Patterns of Change focus in the Tri-State Assessment is on cycles, constancy and change, and evolutionary change.

Form and Function (FAF) "Form and function are complementary aspects of objects, organisms, and systems in the natural and designed world. The form or shape of an object or system is frequently related to use, operation, or function. Function frequently relies on form. Understanding of form and function applies to different levels of organization. Students should be able to explain function by referring to form and explain form by referring to function." (NRC, 1996)

The Form and Function focus in the Tri-State Assessment is on understanding form and function in the natural world. While form and function in the designed world may be included as it relates to other

targets that address technology, f*orm and function as it relates to engineering design will be assessed locally.*

Appendix B: References

The following sources were used in the development of the Tri-State Assessment Targets, Broad Areas of Inquiry, and the Tri-Sate Science Test Specifications.

American Association for the Advancement of Science/Project 2061. (2001). *Atlas of Science Literacy,* AAAS and National Science Teachers Association, Washington, D.C.

American Association for the Advancement of Science. (1993). *Benchmarks for Scientific Literacy*. New York: Oxford University Press.

Champagne, A. and Minstral, J. (2005). unpublished paper (inquiry constructs).

Driver, Squires, Rushworth & Wood-Robinson (2002). *Making Sense of Secondary Science: Research into Children's Ideas*. London and New York: RouteledgeFalmer.

Duschl, R., and D. Gitomer. (1997). "Strategies and challenges to changing the focus of assessment and instruction in science classrooms." *Educational Assessment* 4 (1): 37-73.

Krueger, A. and Sutton, J. (2001). *EDThoughts: What We Know about Science Teaching and Learning*. Aurora, CO: Mid-continent Research for Education and Learning.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher,* 23(2), 13-23.

NAEP Science Consensus Project. (1994). *NAEP Science Framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington, D.C.: National Assessment Governing Board, U.S. Department of Education.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. Washington, DC: National Academy Press.

New Hampshire Department of Education. (1995) *New Hampshire K-12 Science Curriculum Framework.* Concord, NH: Author. [On-line]. Available: http://www.ed.state.nh.us/CurriculumFrameworks/k-124.htm

New Hampshire Department of Education. (1995) *K-6 and 7-10 Science Addendums for New Hampshire K-12 Science Curriculum Framework*. Concord, NH: Author. [On-line]. Available: http://www.ed.state.nh.us/Assessment/7-10sciaddum.htm

New Standards Project (1998). *Performance Standards: Volume 1: Elementary School.* National Center on Education and the Economy and the University of Pittsburgh.

New Standards Project (1997). *Performance Standards: Volume 2: Middle School*. National Center on Education and the Economy and the University of Pittsburgh.

New Standards Project (1998). *Performance Standards: Volume 3: High School.* National Center on Education and the Economy and the University of Pittsburgh.

National Science Teachers' Association. (2000). NSTA Pathways to the Science Standards, Elementary Edition, *Second Edition. Arlington, VA: Author.*

National Science Teachers' Association. (2000). NSTA Pathways to the Science Standards, Middle School Edition, *Second Edition. Arlington, VA: Author.*

National Science Teachers' Association. (2000). *NSTA Pathways to the Science Standards, High School Edition*, Second Edition. Arlington, VA: Author.

Mid-continent Research for Education and Learning. (2000). *McREL Science Standards,* Aurora, CO: Author.

Tasmania Department of Education, *Thinking - Inquiry Standards K-12*. [On-line]. Available: http://www.education.tas.gov.au/ocll/publications/outcomes/inquiry.htm

U.S. Department of Education. (2003). *The Third International Mathematics and Science Study (TIMSS) Science Assessment Framework*. Washington, D.C.: Author.

Rhode Island Department of Elementary and Secondary Education (1995). *The Rhode Island Science Framework*. Providence, RI: Author.

Vermont Department of Education (2000). *Vermont's Framework of Standards and Learning Opportunities*. Montpelier, VT: Author.

Vermont Department of Education (2004). *Grade Expectations for Vermont's Framework of Standards and Learning Opportunities*. Montpelier, VT: Author.

Webb, N. (March 2002). "Depth of Knowledge Levels for Four Content Areas." (unpublished paper).

Webb, N. (1997). Research Monograph Number 6, *Criteria for Alignment of Expectations and Assessments on Mathematics and Science Education*. Washington, D.C.: CCSSO.

Webb, N. (August 1999). Research Monograph No. 18: "Alignment of Science and Mathematics Standards and Assessments In Four States." Washington, D.C.: CCSSO.

Appendix C: Measurement Benchmarks (for NECAP Mathematics Assessment grades 3-8)

The following is a list of the measurement benchmarks and equivalences that ***can be used*** in mathematics problems across the content strands at each grade level to address the expectations in M (G&M)–X–7 for the New England Common Assessment Program[1].

GLE M (G&M)–X–7: **Uses units of measures appropriately and consistently, and makes conversions within systems when solving problems** across the content strands.

The type of measure (e.g., length, time, etc.), the unit (e.g., inches, feet, etc.), the degree of accuracy where appropriate (e.g., ½ inch); and equivalences (e.g., 12 inches in a foot) are identified for grades 2 – 8. In addition to measurement benchmarks identified below students will be expected to use the appropriate units when solving problems involving area, volume, surface area, conversions, and rates (e.g., miles per hour, price per pound, pounds per square inch) on the NECAP Assessment[1].

Measures	**Grade 2**	**Grade 3**	**Grade 4**
Length	**Unit (accuracy):** Inch (to whole inch); Foot (to whole inch); Centimeter (to whole centimeter); Meter (to whole centimeter) **Equivalencies:** 12 inches in 1 foot; 100 centimeters in 1 meter	**Unit (accuracy):** Inch (to 1/2 inch); Foot (to whole inch); Centimeter (to whole centimeter); Meter (to whole centimeter) **Equivalencies:** 12 inches in 1 foot; 100 centimeters in 1 meter	**Unit (accuracy):** Inch (to 1/4 inch); Foot; Centimeter (to 0.5 centimeter); Meter (to 0.5 centimeter); Yard; Mile (use in scale questions); Kilometer (use in scale questions) **Equivalencies:** 12 inches in 1 foot; 100 centimeters in 1 meter; 3 feet in 1 yard; 36 inches in 1 yard
Time	**Unit (accuracy):** Hour (to 15 minute interval) **Equivalencies:** 60 minutes in 1 hour	**Unit (accuracy):** Hour (to 5 minute interval); Day; Year **Equivalencies:** 24 hours in 1 day; 7 days in 1 week; 365 days in 1 year	**Unit (accuracy):** Hour (to 5 minute interval); Day; Year **Equivalencies:** 24 hours in 1 day; 7 days in 1 week; 365 days in 1 year; 60 seconds in 1 minute; 60 minutes in 1 hour
Temperature	**Unit (accuracy):** Degree (to 1 degree)	**Unit (accuracy):** C^o and F^o (to 1 degree)	**Unit (accuracy):** C^o and F^o (to 1 degree)
Capacity		**Units (accuracy):** Quart (to whole quart)	**Unit (accuracy):** Quart (to whole quart)
Mass		**Unit (accuracy):** Kilogram (to whole kilogram); Gram (to whole gram)	**Unit (accuracy):** Kilogram (to whole kilogram); Gram (to whole gram)
Weight		**Unit (accuracy):** Pound (to whole pound)	**Unit (accuracy):** Pound (to whole pound)

[1] Contractors will be asked to devise a system to measure the degree to which students use units of measures and make conversions consistently and appropriately when applicable to problems across content strands.

Measures	Grade 5	Grades 6 – 8
Length	**Units (accuracy):** Inch (to 1/8 inch); Foot; Centimeter (to 0.5 centimeter); Meter (to 0.5 centimeter); Yard; Mile (use in scale questions); Kilometer (use in scale questions) **Equivalencies:** 12 inches in 1 foot; 100 centimeters in 1 meter; 3 feet in 1 yard; 36 inches in 1 yard; 10 millimeters in 1 centimeter	**Units (accuracy):** Inch (to 1/16 inch); Foot; Centimeter (to 1/10 centimeter); Meter (to 1/100 meter); Yard; Mile (use in scale and rate questions); Kilometer (use in scale and rate questions) **Equivalencies:** 12 inches in 1 foot; 100 centimeters in 1 meter; 3 feet in 1 yard; 36 inches in 1 yard; 10 millimeters in 1 centimeter; 1000 millimeters in 1 meter
Time	**Unit (accuracy):** Hour (to 1 minute); Day; Year **Equivalencies:** 24 hours in 1 day; 7 days in 1 week; 365 days in 1 year; 60 seconds in 1 minute; 60 minutes in 1 hour	**Unit (accuracy):** Hour (to 1 minute); Day; Year **Equivalencies:** 24 hours in 1 day; 7 days in 1 week; 365 days in 1 year; 60 seconds in 1 minute; 60 minutes in 1 hour
Temperature	**Unit (accuracy):** Cº and Fº (to 1 degree)	**Unit (accuracy):** Cº and Fº (to 1 degree)
Capacity	**Unit (accuracy):** Quart (to 1 ounce); Gallon; Pint **Equivalencies:** **32 ounces in 1 quart; 4 quarts in 1 gallon; 2 pints in 1 quart**	**Unit (accuracy):** Quarts (to 1 ounce); Gallon; Pint; Liter **Equivalencies:** 32 ounces in 1 quart; 4 quarts in 1 gallon; 2 pints in 1 quart; 1000 milliliters in 1 liter
Mass	**Unit (accuracy):** Kilogram; Gram (to whole gram)	**Unit (accuracy):** Kilogram; Gram (to 1/10 gram)
Weight	**Unit (accuracy):** Pound (to 1 ounce) **Equivalencies:** 16 ounces in 1 pound	**Unit (accuracy):** Pound (to 1 ounce) **Equivalencies:** 16 ounces in 1 pound
Angles and Rotation	**Unit (accuracy):** Degree (to 2 degrees)	**Unit (accuracy):** Degree (to 2 degrees) **Equivalencies:** 360º in 1 circle; 90º in 1 right angle

ATTACHMENT D

Budget Forms: Line Item Budget

- All lines must be completed: Enter $0 if appropriate.
- Travel costs should be included in line items, where appropriate.
- Enter subtotals for higher level categories on corresponding lines.(e.g. labor, other direct costs, shipping and receiving, documents and databases)
- Budgets for identified as separate cost items and alternative or additional tasks proposed by the contractor must be provided on separate forms.
- The contractor's response must include a detailed narrative explaining and supporting the costs provided on this form.
- Note that FY06 and FY11 represent partial years. FY06 includes February 2006 through June 2006. FY11 includes July 2010 through December 2010.

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
I. Labor (subtotal)								
A. All except scorers of constructed-response items	**All**							
B. Scorers	**E3**							
II. Other Direct Costs (subtotal)								
A. Office Operations (telephone, misc. postage, photocopying, supplies, misc. shipping, etc.)	All							
B. Shipping and Receiving (subtotal)								
Test Administration	D1-D2							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
Reporting of Results	G1							
C. Documents and Databases (subtotal)								
Test Materials (subtotal)								
Development Team	B,C,E,G							
Item Review Committees	B2							
Bias Committee	B3							
Other item-related costs	B4-B6							
Test Booklets	C1,C2 a-e							
Answer Booklets	C3							
Student Labels	C4							
Ancillary Materials	C5a-j							
Large Print Materials	C2f							
Braille Materials	C2g							
Reports of Results (subtotal)								
Parent/Guardian Report	G1							
Classroom Roster	G1							
Classroom Summary	G1							
School Report Package	G1							
District Report Package	G1							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
State Report Package	G1							
Domain Subscore Reports	G1							
Results Databases	G1							
Interpretive Materials	G1							
Released Item Materials	G1							
Technical Report	G1							
Additional Documents and Publications (subtotal)								
Management Reports	A1							
Annual Ancillary Reports	C5k							
Scoring Documentation	E3l	--						
Additional Analyses	F7							
Ongoing Documentation	All							
Standard Setting Document	Hd	--		--	--	--	--	--
D. Meetings/Conferences/ Workshops (subtotal)								
Management Meetings	A1							
Development Team Meetings	B,C,E,G							
Item Review Committee	B2							
Technical Advisory Committee	A2a							
NE Compact Board meetings	A2b							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
Item Selection Meetings	C1							
Bias Committee Meetings	B3							
Standard-Setting Meetings	H	--		--	--	--	--	--
Test Administration Workshops	D3a							
Test Reporting Workshops	G2a							
E. Support (subtotal)								
Administration Support	D3b							
Reporting Support	G2b							
III. Overhead and Fees (subtotal)								
A. Overhead								
B. Fees								
IV. Total (Labor + ODCs + Overhead and Fees)								
Optional Cost Items								

Budget Forms: Major Task Area Budget Summary
Fiscal Year _____
Directions: Complete a separate form for each fiscal year

	I. Labor			II. Other Direct Costs	Total
	All employees (regular and temporary) except scorers	Scorers	Subcontractors		
A. Project Management and Planning					
B. Item Development					
C. Test Construction					
D. Administration					
E. Scanning/Imaging and Scoring					
F. Analysis					
G. Reporting					
H. Standard Setting					
Overhead					
Fees					
Total					

ATTACHMENT E

Task Allocation Form

Directions: Please complete a separate form for each fiscal year of the project. Include the name of each staff person assigned to the project for 0.5 FTE or more in a given year. Include the FTE by department name or job description for major tasks completed across other staff. Column totals should reflect an estimate of the total FTE assigned to each task. Additional pages may be submitted, as necessary, and should be clearly labeled (e.g., page x of y). The final page must contain a grand total across pages. Corresponding forms must be submitted and clearly labeled for separate cost items specified in the RFP or additional/alternative tasks proposed by the contractor.

Fiscal Year:	**Staff Allocations by Task Area**								
Name of Department/Staff	A Project Mgmt	B Item Dev.	C Test Construction	D Administration	E Scanning / Scoring	F Analysis	G Reporting	H Standard Setting	TOTAL
Total (page subtotal)									

Attachment A

Number of Rhode Island Students per Grade:

Grade 4 Students	12,000
Grade 8 Students	12,000
Grade 11 Students	11,000

Attachment B

Number of Rhode Island Districts and Schools:

Number of Rhode Island Districts	40
Number of Rhode Island Schools	370

Attachment C

NEW ENGLAND COMMON ASSESSMENT PROGRAM (NECAP)
NEW HAMPSHIRE, RHODE ISLAND, & VERMONT

TEST SPECIFICATIONS FOR THE

TRI-STATE SCIENCE ASSESSMENT

November 2005

ATTACHMENT D

Budget Forms: Line Item Budget

- All lines must be completed: Enter $0 if appropriate.
- Travel costs should be included in line items, where appropriate.
- Enter subtotals for higher level categories on corresponding lines.(e.g. labor, other direct costs, shipping and receiving, documents and databases)
- Budgets for identified as separate cost items and alternative or additional tasks proposed by the contractor must be provided on separate forms.
- The contractor's response must include a detailed narrative explaining and supporting the costs provided on this form.
- Note that FY06 and FY11 represent partial years. FY06 includes February 2006 through June 2006. FY11 includes July 2010 through December 2010.

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
I. Labor (subtotal)								
A. All except scorers of constructed-response items	**All**							
B. Scorers	**E3**							
II. Other Direct Costs (subtotal)								
A. Office Operations (telephone, misc. postage, photocopying, supplies, misc. shipping, etc.)	All							
B. Shipping and Receiving (subtotal)								
Test Administration	D1-D2							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
Reporting of Results	G1							
C. Documents and Databases (subtotal)								
Test Materials (subtotal)								
Development Team	B,C,E,G							
Item Review Committees	B2							
Bias Committee	B3							
Other item-related costs	B4-B6							
Test Booklets	C1,C2 a-e							
Answer Booklets	C3							
Student Labels	C4							
Ancillary Materials	C5a-j							
Large Print Materials	C2f							
Braille Materials	C2g							
Reports of Results (subtotal)								
Parent/Guardian Report	G1							
Classroom Roster	G1							
Classroom Summary	G1							
School Report Package	G1							
District Report Package	G1							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
State Report Package	G1							
Domain Subscore Reports	G1							
Results Databases	G1							
Interpretive Materials	G1							
Released Item Materials	G1							
Technical Report	G1							
Additional Documents and Publications (subtotal)								
Management Reports	A1							
Annual Ancillary Reports	C5k							
Scoring Documentation	E31	--						
Additional Analyses	F7							
Ongoing Documentation	All							
Standard Setting Document	Hd	--		--	--	--	--	--
D. Meetings/Conferences/ Workshops (subtotal)								
Management Meetings	A1							
Development Team Meetings	B,C,E,G							
Item Review Committee	B2							
Technical Advisory Committee	A2a							
NE Compact Board meetings	A2b							

Line Item	Scope of Work Task	FY06	FY07	FY08	FY09	FY10	FY11	Total
Item Selection Meetings	C1							
Bias Committee Meetings	B3							
Standard-Setting Meetings	H	--		--	--	--	--	--
Test Administration Workshops	D3a							
Test Reporting Workshops	G2a							
E. Support (subtotal)								
Administration Support	D3b							
Reporting Support	G2b							
III. Overhead and Fees (subtotal)								
A. Overhead								
B. Fees								
IV. Total (Labor + ODCs + Overhead and Fees)								
Optional Cost Items								

Budget Forms: Major Task Area Budget Summary
Fiscal Year _____
Directions: Complete a separate form for each fiscal year

	I. Labor			II. Other Direct Costs	Total
	All employees (regular and temporary) except scorers	Scorers	Subcontractors		
A. Project Management and Planning					
B. Item Development					
C. Test Construction					
D. Administration					
E. Scanning/Imaging and Scoring					
F. Analysis					
G. Reporting					
H. Standard Setting					
Overhead					
Fees					
Total					

ATTACHMENT E

Task Allocation Form

Directions: Please complete a separate form for each fiscal year of the project. Include the name of each staff person assigned to the project for 0.5 FTE or more in a given year. Include the FTE by department name or job description for major tasks completed across other staff. Column totals should reflect an estimate of the total FTE assigned to each task. Additional pages may be submitted, as necessary, and should be clearly labeled (e.g., page x of y). The final page must contain a grand total across pages. Corresponding forms must be submitted and clearly labeled for separate cost items specified in the RFP or additional/alternative tasks proposed by the contractor.

Fiscal Year:	**Staff Allocations by Task Area**								
Name of Department/Staff	A Project Mgmt	B Item Dev.	C Test Construction	D Administration	E Scanning / Scoring	F Analysis	G Reporting	H Standard Setting	TOTAL
Total (page subtotal)									