

# ***An Introduction to the Operational Best Practices for Statewide Large-Scale Assessment Programs (2013)***

The 2013 Edition of the *Operational Best Practices for Statewide Large-Scale Assessment Programs* (“*Operational Best Practices*”) updates and expands the original book published in June 2010. The Association of Test Publishers (ATP) and the Council of Chief State School Officers (CCSSO), the joint owners and sponsors of the *Operational Best Practices*, have prepared this Introduction to provide a description of subjects that have been added or updated since 2010.

This Introduction also contains the newly added Pre-Chapter entitled, “State Considerations for Assessment Program Design.” This section of the 2013 Edition provides a basic checklist of program design information for a state to use as it evaluates and makes changes to its existing assessment program, whether those changes are required because of changes in state law or a decision to participate as part of an assessment consortium. As such, this “Pre-Chapter” identifies and organizes most of the topic areas in the 2013 Edition. A copy of the Table of Contents is also provided, so readers can easily see the entire scope of the 2013 Edition.

As sponsors of this resource guide, ATP and CCSSO proudly invite potential users to review this free booklet. We believe you will be able to identify areas of interest and get a glimpse of exactly how the 215-page 2013 Edition can play an important role in the planning and conduct of your assessment program – whether you are a staff member of a state department, a contracting publisher, a Race to the Top Assessment Consortium, a Local Educational Agency, a Large City School District, or a school. Although the original audiences for the *Operational Best Practices* were the states and their service providers contracting entities, it has become clear that many other stakeholders can gain valuable insights into the state assessment process through this guide. We encourage individuals who are part of those groups, especially educators in districts and schools, to familiarize yourselves with the *Operational Best Practices* – the knowledge you gain about how statewide assessment programs operate can be applied to your own district and school level activities that are tied to, or complement, the statewide assessment or assessment systems used in your state. In other situations, such as for the Race to the Top Assessment Consortia, large city school districts, and local districts with their own assessments, the guidance in the 2013 Edition will serve you well in operating your own assessment programs.

Knowledge about state assessment programs is critical to enabling all stakeholders to understand how the statewide program is employing strong quality control measures and thus, that the assessment data generated from such assessments/assessment systems is being used to benefit all students. Such knowledge can also be extremely useful in bridging the trust gap that often exists between those responsible for assessing and students and their parents.

## **Background**

In June 2006, the ATP and the CCSSO embarked on a cooperative journey to build a set of voluntary, non-regulatory best practices for states and testing companies to use to strengthen

implementation and quality of statewide testing programs in the United States conducted under the No Child Left Behind Act (NCLB). That goal has been achieved.

The 2010 *Best Practices* has been widely used by States and publishers, for training and for considering changes to state testing programs. The US Government Accountability Office relied heavily on the *Best Practices* in conducting its 2012 survey dealing with state test security issues, as it reported. See “K-12 Education: States’ Test Security Policies and Procedures Varied,” GA-13-495R (May 16, 2013).

Although the 2010 Edition was comprehensive, it was focused mostly on paper-based assessments, where the sponsoring organizations possessed substantial experience to decide on and articulate what constituted best practices. Since 2010, however, many states have begun using, in whole or in part, assessments that are delivered by or taken on computer. Accordingly, the ATP and the CCSSO felt it was appropriate and necessary to expand the *Operational Best Practices* to cover those assessments, whether a statewide assessment program utilized only automated testing alone, or a hybrid assessment system comprised of both paper-based and automated assessments. The next section of this Introduction discusses what subjects are covered in the 2013 Edition.

### **What’s New in the 2013 Edition?**

The biggest change in the 2013 Edition is that it addresses “technology-based assessments” – that is, assessments delivered and/or taken by computer, including so-called “online” assessments, as well as those that are cached or remotely handled. The updated book also deals with programs using both technology-based and paper-based assessments. But because technology is moving so rapidly, the 2013 Edition does not describe the delivery or administration of technology-based assessments in terms of devices that are currently available (e.g., desk top computer or other “Net-centric” device possessing computer capability, including hand-held, tablet, cellular telephone, or other type of personal digital assistant device). By this approach, we expect the 2013 Edition will be applicable to automated systems that have not yet been implemented – or in some cases have not even been invented.

Because the reach of technology-based assessments is broad, the 2013 Edition revises many chapters from the 2010 Edition, to set forth best practices related to technology-based assessments in manner that is parallel to those practices included for paper-based assessments. Therefore, the 2013 Edition updates almost all of the existing content, from program management (need for technology expertise on both client and service provider teams), to item development and banking (technology-enhanced items), to test construction (CBT, CAT, LOFT) to test delivery, to administration, to scoring and reporting, and to data management systems.

Numerous other topics that bear a direct relationship to the use of technology-based assessments have been expanded or greatly modified. These include the following areas: new problems in program design, new practices for test security, new elements in preparing for and delivering technology-based assessments, and the need to handle test accommodations for special populations. Test security also has been enhanced to cover issues beyond a focus on student involvement; as such this expanded chapter will be useful to all stakeholders. The

challenges and complexities of accommodations have been modified to reflect the concept of accessibility for all students. Furthermore, the focus on technology means that components of technology systems (e.g., hardware, software, communications networks) and assessment programs need to be interoperable with one another (e.g., for the smooth exchange of data and content) – a new chapter on interoperability is included, as well as discussions about interoperability as it applies to item banking, data management, and score reporting.

Finally, as mentioned above, the 2013 Edition includes a pre-chapter on “State Considerations for Assessment Program Design,” that provides a basic checklist for a state to follow before it even prepares an RFP to select a service provider. Although states will be the principle user of this checklist, service providers and other stakeholders will also find this material instructive to gain an understanding about how states will go about the process of evaluating and modifying their current assessment programs when state laws change or when the assessment consortia activities come online.

### **Conclusion**

The ATP and the CCSSO recognize that the testing process is not static. We expect that the *Operational Best Practices* will continue to be reviewed on a periodic basis, updated to account for changes in legal standards (e.g., Elementary and Secondary Education Act), technology or testing methodologies, or as best practices continue to evolve in response to such technology or as states gain experience under the various Race to the Top Consortia assessments. We will strive to ensure that the content of the *Operational Best Practices* remains as current and viable as possible.

Sincerely,

Christopher Minnick  
Executive Director  
Council of Chief State School Officers

William G. Harris, Ph.D.  
Chief Executive Officer  
Association of Test Publishers

## 2013 Edition Table of Contents

Introduction to Operational Best Practices .....	5
Pre-Chapter. State Considerations for Assessment Program Design.....	17
Chapter 1. Program Management and Customer Service.....	25
Chapter 2. Item Development .....	37
Chapter 3. Item Banking .....	51
Chapter 4. Test Construction and Development.....	59
Chapter 5. Stand-Alone Field Testing .....	67
Chapter 6. Production of Test Materials .....	71
Chapter 7. Third-Party Management .....	79
Chapter 8. Security .....	81
Chapter 9. Test Administration.....	99
Chapter 10. Materials Packaging for Paper-Based Tests.....	107
Chapter 11. Distribution of Materials and Chain of Custody Management .....	111
Chapter 12. Receiving, Check-in, and Processing of Materials.....	117
Chapter 13. Scanning and Post-Scanning Editing .....	119
Chapter 14. Scoring .....	125
Chapter 15. Technology-based Assessments and Technical Support .....	135
Chapter 16. Technical Defensibility.....	145
Chapter 17. Score Reporting .....	149
Chapter 18. Data Management.....	157
Chapter 19. Assessment of Special Populations / Accessibility for all Learners .....	165
Chapter 20. Assessment Program Procurement .....	173
Chapter 21. Transition from One Service Provider to Another .....	183
Chapter 22. System and Program Interoperability .....	195
Appendix to Chapter 22.....	201

Glossary.....203

# PRE-CHAPTER. STATE CONSIDERATIONS FOR ASSESSMENT PROGRAM DESIGN

## Introduction

The assessment program design provides the foundation for development and implementation decisions to be made in a large-scale assessment program. The purpose of this chapter is to assist staff in departments of education to review best practices as they evaluate and make changes to current assessment programs, develop an entirely new assessment program, participate in assessment consortia and/or use the assets from a consortium. All decisions and planning will be subject to state procurement law, the state history, the legislative policy framework and the infrastructure or capacity of the field.

Ultimately the program goals and purposes will be the primary factors guiding the final assessment program design as decisions are made in assessment system development and implementation, including evaluations aimed at improving the interoperability of technology components and data/content exchange.

- I. **Program Design. Although the client initiating a change in the assessment program has the ultimate responsibility for the planning decisions, the client may, subject to state procurement laws, find it useful to obtain advice from existing service providers. One way to do this is through a Request for Information (RFI) or similar inquiry.**
- II. **Legislation and policy will provide the framework for the assessment program and design. A client may be responding to new legislation, transitioning to changed policy requirements, managing the impact of assessment consortium activities and products, or participating in a consortium of states for the development and implementation of an assessment program.**
  - II.a As decision-making begins, the client (e.g. state's department of education) should have a clear understanding of the state's expectations and requirements. It is important to clarify the fundamental principles and purpose because these will guide the overall assessment design.
  - II.b The client will need to consider the feasibility of execution and timeframe needed to meet the state and federal policy and legislative requirements.
  - II.c The client must understand and consider such factors as: the limitations of end users capacity, possible consortium structure, and industry best practices.
- III. **The purposes of the assessment program will guide its development and implementation. Consortia activities and requirements may also influence timing, implementation and service provider scope and involvement.**

- III.a The history of the client's assessment program design should be considered in making revisions to an existing system, developing a new assessment program, or the transitioning to a consortium system design.
- III.b Historical perspectives, past implementation, and the practicality of changes in assessment programs must be considered, and will influence implementation decisions about new assessment design, delivery, and the purpose of assessment.
- III.c The client must clearly understand the purposes of the new assessment design and the intended assessment data uses. These purposes may include but are not limited to providing data relevant to and appropriate for the following:
- State and/or federal accountability;
  - Student achievement of standards;
  - Instructional improvement;
  - School improvement;
  - Teacher effectiveness;
  - 3rd Party research; and
  - Integration with other data tools and resources.
- III.d In all cases the purposes of the assessment program under development need to be clearly articulated in the assessment program design.
- III.e The client will need a plan that ensures that both short and long-term goals and tasks of the program are realistic, outlined, and delivered with clear timelines.
- III.e(i) Short-term goals and tasks might include such things as transitioning tasks either from a previous vendor or from a previous program design, developing test specifications, item development, test construction, test administration and, in the case of technology-enhanced testing, platform construction, modification, and implementation.
- III.e(ii) Long-term goals and tasks might include such things as reporting results, implementing data communication tools, and using data in formative or summative ways.

**IV. Assessment Program Design. General design decisions will be determined by state procurement laws, policy, history, and assessment program goals. The general program design includes many components, such as content standards, testing modes, item types, psychometric models, test development and administration, scoring and reporting, and other program elements.**

- IV.a **Content standards.** Decisions about assessment program design will be based upon the client's content standards (e.g., the Common Core State Standards, standards for learning the English language), the constructs to be measured within those standards, and the achievement standards that describe levels of student performance. Additional information regarding item types is found in Chapters 2 and 17.

- IV.b **Assessment program components.** Decisions about how program components (e.g., formative, interim/benchmark, summative assessments, English Language Proficiency (ELP) assessments) integrate or coordinate need to be discussed.
- IV.c **Testing modes.** Decisions about paper-based or technology-based testing need to be considered with respect to financial, operational, information technology (IT) infrastructure impacts, diversity of student populations (e.g., range of language learning and disabilities), and assessment goals. The capacity of the state agency, districts, and schools must be considered. Client considerations may be constrained by varying factors and need to be evaluated relative to program goals. Further detail is provided throughout the book.
- IV.d **Interoperability.** Decisions about the interoperability of systems, the exchange of content, data, and results, are critical considerations. Chapter 22 addresses the concept of interoperability and the industry standards. Specific examples of the application of content interoperability are found in Chapters 2, 3, 6, 15, 17, 18, and 19.
- IV.e **Item types.** The program design and mode of testing will determine item types. Item type decisions need to be thoroughly considered because there are many different types, delivered through both paper-based and technology-based modes. Additional information regarding item types and item banking is found in Chapters 2 and 3.
- IV.f **Psychometric models.** Decisions about the measurement models including considerations such as field testing, scaling, equating, and standard setting, should be examined. All psychometric decisions must be consistent with industry standards, especially the *Joint Standards* in reference to the professional standards. Additional information regarding psychometric options is found in Chapters 4, 5, and 16.
- IV.g **Test development.** Decisions about item development processes, item reviews (e.g., content, alignment, fairness and sensitivity), and test construction are important. Universal design must be considered throughout the test development process. Additional information is found in Chapters 2, 3, 4, and 19.
- IV.h **Test administration and scheduling.** Decisions about test administration must be considered in the assessment design. The administration schedule generally dictates the timeframe for all other tasks. The client should evaluate its comprehensive schedule for test development, test administration, scoring, and reporting. Planning will need to include timelines for state and federal accountability. The schedule and training for test administration needs to be evaluated, and a plan for communicating the eventual schedule and training needs to be considered. Additional information is found in Chapters 9, 10, 11, and 12.
- IV.i **Scoring and reporting.** Decisions are primarily dependent upon the program goals, assessment purposes (e.g., summative, interim), and use of performance data. Fundamental differences exist between scoring and reporting of paper-based and technology-based assessments and should be considered in decision-making. Additional information is found in Chapters 13, 14, and 15.

**V. Technology. Decisions about the assessment program design, including accessibility for diverse student populations, will drive the technology requirements needed to support the program. The successful adoption of technology-based assessments requires evaluation of current and emerging technology for operational capabilities as well as consideration of system interoperability.**

V.a **Infrastructure.** The client will need to consider the readiness of existing local and state technology infrastructure. The client will need to be aware of and communicate the minimum technology requirements to successfully support the defined program design goals.

V.b **Methodology.** The client will need to consider establishing a process for implementing appropriate methodology, for evaluating current local site readiness, and for planning for emerging technology requirements.

V.c **Systems interoperability.** The client will need to consider the goals and benefits of interoperability, including the use of available industry standards, at the initial point of program design. Putting in place interoperable solutions may benefit the assessment program by enabling enhancements or simply providing more efficient transitions from one service provider's solution (either hardware, software, or both) to another service provider's solution (see Chapter 22).

V.d **Data integration and resources.** The client will need to consider its goals for integrated data warehousing, management, and reporting as part of technology planning to support the program. Integration efforts along with existing and new systems development will be needed. These efforts should all be considered, as part of the client's overall assessment program, while supporting systems and infrastructure are developed, interoperable solutions are evaluated, and data management objectives set (see Chapters 18 and 22).

V.e **Sustainability and interoperability.** The client will need to consider future plans to sustain and manage technology platforms, systems and infrastructure components, including how to best achieve interoperability. Overall program goals, projected budgets, and implementation timelines will influence and guide client requirements (see Chapter 22).

V.f **Training.** The client will need to consider establishing a plan for training, including such elements as format, participants, and logistics (see Chapters 1, 8, 9, 14, and as referenced in other chapters). Training also should include technical and administrative training on relevant test administration processes for all stakeholders.

**VI. Security. Increased awareness of security issues, increased public scrutiny of testing practices, and the value of strong security practices in the proper uses of test results make it important to evaluate a state's security policies and practices. A comprehensive security plan should cover such topics as security policies, legal support, identification of threats and vulnerabilities, and how to respond to security incidents. Best practices on**

these topics and others are found in Chapters 8, 9, and 15, and are noted in other chapters.

- VII. Accessibility and Accommodations. Accessibility and accommodations design considerations must be guided by state and federal legislative requirements. Accessibility needs should be informed by principles of universal design. Allowable accommodations need to be clearly defined and communicated, and then evaluated on a periodic basis. Additional information is available in Chapter 19.**
- VIII. Data Management Decisions. The client will need to consider decisions regarding data management. Additional information is found in Chapter 18.**
  - VIII.a The steps in the collection and management of data will vary depending upon infrastructure and capacity. Additionally, reporting requirements may impact how data are managed and used.
  - VIII.b Rules, procedures and a communication plan need to be established to assure data quality, security, and management.
  - VIII.c Decisions for enabling data interoperability will provide more flexibility in managing data throughout the assessment program (see Chapter 22).
- IX. Roles and Responsibilities. Decisions about the roles and responsibilities of the client and service providers will be established and documented. Additional information is found in Chapters 1 and 7.**
  - IX.a The assignment of roles and responsibilities may vary between the client and the service provider(s) as documented in the applicable contract(s). For example, the client and the service provider(s) may share responsibilities for the development of the assessment program, professional development, materials preparation, and/or customer service.
  - IX.b If consortia arrangements and responsibilities are included in the planning discussions, the roles of all parties need to be defined, agreed upon, and communicated to all parties.
- X. Budget. Decisions about budget must be considered and determined by the client.**
  - X.a The client should prioritize budgetary considerations such as resources, capabilities, and training.
  - X.b Procedures for managing changes in scope must be articulated, documented, and applied.
- XI. Procurement. After gathering and evaluating design, development, and implementation requirements, the client will, in accord with procurement laws, issue a Request for Proposal (RFP) or similar document (see Chapter 20). The RFP should contain sufficient**

**information to allow all qualified service providers to understand the scope and specifications of the program being sought.**