Building an Interim Assessment System:

A Workbook for School Districts

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

Building an Interim Assessment System: A Workbook for School Districts

CCSSO
Council of Chief State School Officers

Building an Interim Assessment System:

A Workbook for School Districts

Eric W. Crane
Senior Research Associate, WestEd

How to use this workbook

As someone with a stake in a school district's systems, you probably do not have all the answers around what is necessary to build an effective interim assessment system. Neither does this workbook. But it is intended to have the right questions. More precisely, this workbook contains the vision, infrastructure, and resource questions critical to a thorough, district-level self-examination of readiness for an interim assessment system. Staff at state departments of education can also use this resource when district-level staff approach the state agency seeking guidance. These questions are intended to guide you and other district stakeholders through a structured consideration of building or revisiting your district's interim assessment system. The questions are intended to be answered collaboratively, not by a district administrator in isolation. The authors of this workbook believe that a team of district stakeholders addressing these questions is the approach most likely to yield progress in creating a lasting, successful system.

The workbook is organized into the following nine sections:

- Goals and Vision
- Elements and Features
- Professional Development
- Alignment to Standards and Curriculum
- Costs and Capacity
- Test Development and Technical Quality
- Administration and Scoring
- Reporting
- Evaluation

A Definition of Interim Assessments

In a paper commissioned by the Council of Chief State School Officers' (CCSSO) collaborative on Technical Issues in Large-Scale Assessment (TILSA), Crane (2008) offers the following definition of interim assessments:

Assessments administered multiple times during a school year, usually outside of instruction, to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the student, classroom, school, or district level. The specific interim assessment designs are driven by the purposes and intended uses, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts. (p. 2)

This definition builds on earlier work by Perie, Marion, and Gong (2007). This revised definition emphasizes that interim assessment is periodic in nature, and with regard to both stakes and frequency, falls between classroom-level, low-stakes, high-frequency formative assessment and state-level, high-stakes, low-frequency summative assessment. Links to these resources, as well as updated work by Perie and her colleagues (2009), are available in Appendix A.

	Formative	Interim	Summative
Typical Use	feedback to adjust ongoing teaching and learning	monitoring student progress	student placement; school and district accountability
Frequency of Administration	continual; multiple times a day	generally 2–6 times per school year	usually once a school year
Scope of Administration	student and classroom	usually school or district (could be student, as in a response to instruction and intervention model)	usually state

Source: Crane, 2008, p. 4.

School districts are at different stages in their thinking and planning concerning interim assessment. Some are considering implementing their first-ever system. Others are revisiting an established system. The questions in this workbook address both the system and the items and tests that make up the system. Most school districts can benefit from looking at all the sections of this workbook; however, two sections are more technical than the rest. Districts just beginning the process may want to defer a serious consideration of the "Alignment to Standards and Curriculum" and "Test Development and Technical Quality" sections

until they have built a more substantial foundation to support an interim assessment system. The first section, "Goals and Vision," is in our view the most important. Discussions on an interim assessment system need to solidify and establish consensus on goals and vision before proceeding to other topics.

The Council of Chief State School Officers' (CCSSO) collaborative on Technical Issues in Large-Scale Assessment (TILSA) authored this workbook, with Eric Crane of WestEd as lead author. This TILSA group geared the workbook to a universal level. Some of your questions may not be addressed within these pages. However, we hope and expect that a self-examination guided by this workbook will prove useful to a diverse array of school districts.

GOALS AND VISION		
Question	**Relevance / Follow-up questions**	**Notes / Response**
1. What is the purpose of the interim assessment system? What are our goals for this program?	**This is the central question of this process. The importance of this question cannot be overstated.** Crane (2008) identified eight uses of interim assessments: diagnosis, prediction, preparation, placement, student evaluation, school intervention, promotion/graduation, and local accountability (see Appendix B). Which of these apply to the district, and how will they be addressed? – Diagnosis – Prediction – Preparation – Placement – Student evaluation/monitoring student progress – School intervention – Promotion/graduation – Local accountability Are the uses supported by appropriate infrastructure or technical work (e.g., if predictive validity is a use we will rely on, then technical studies must be carried out to support such use)? Note that with each additional use to which the district will put the data, additional technical work or infrastructure will need to be in place.	

Question	Relevance / Follow-up questions	Notes / Response
2. What action steps will be taken as a result of this assessment?	Are there curricular materials at hand that can facilitate re-teaching? Will the results be used to evaluate teachers? Can such use be justified by the assessment?	
3. How will student learning and other valued outcomes be affected as a result of using this interim assessment system?	A clear articulation of the valued outcomes—and what will change to promote these outcomes—is essential to principals' and teachers' (and other stakeholders') understanding, support, and implementation of the system. Identifying specific examples of how student learning will be affected is essential.	
4. Who will use the information gathered from this assessment? What buy-in or sense of ownership exists from various stakeholders and what will be needed to make this program successful and of value?	Realistically, there will be discomfort and resistance to a new system. The system will need champions who can explain the reasons for moving to the new system and the system's benefits. This can promote buy-in.	
5. Can we explain how the interim assessment system fits into the larger landscape of other assessments our students take?	The system should provide important information that is more standardized than the formative assessment information that teachers constantly generate. On the other hand, the interim assessment information should point to "next steps" for learning and instruction, which is not a primary purpose of the end-of-year summative assessment.	
6. What is the scope of the program? Grades? Content areas? Would phasing in the program, initially including only certain grades or content areas, make sense?	A matrix showing content coverage by grade level could help stakeholders see the potential breadth of the program.	

Question	Relevance / Follow-up questions	Notes / Response
7. Have principles of universal design been incorporated to ensure that the needs of special populations are addressed?	Look for assessment materials that developers have designed for use by all populations. If developing your own assessment materials, make sure that expertise on universal design is available.	
8. Do we want to consider developing our system internally, or do we want to purchase a vendor product?	Can we achieve better alignment than if we purchased a product? Do we have the capacity, resources, and funding to develop a product? Do we have the funding to purchase a product?	
9. Are content standards in place and is this program or assessment in support of those standards? (See also "Alignment to Standards and Curriculum" section.)	Are there district-level content standards and objectives that need to be considered along with state standards?	
10. Have we remembered to reflect the relevant questions from this document in our request for proposal (RFP)? Is the vendor willing to provide references from other school districts?	This applies when districts are purchasing a test or releasing an RFP for test or item development services such as online testing, custom test development, scoring and reporting, and item development. RFPs include performance requirements, legal statements, and many other elements that are beyond the scope of this workbook.	

Before you leave this section, verify that there is consensus around the goals and vision for the interim assessment system. Decisions about goals and vision should serve as guiding touchstones for all later work. As you answer questions in the following sections of the workbook, refer back to this "Goals and Vision" section to inform your thinking and decision making.

ELEMENTS AND FEATURES		
Question	**Relevance / Follow-up questions**	**Notes / Response**
1. Can the assessment show student growth? (See also "Reporting" section.)	Student growth can be defined in various ways. Measuring growth within a year is generally more straightforward than measuring growth across years. It may cost more to track growth across years. Growth within a year may be less expensive to track, but it may not yield the information that the district needs. Does the vendor claim that there is a vertical scale? Is there statistical information to justify such a claim?	
2. Is the assessment valid for different student populations? (See also "Test Development and Technical Quality" section.)	What technical work supports this?	
3. Does the assessment provide evidence of strengths and weaknesses at the individual student and group levels?	How much and what kinds of evidence?	
4. Does the assessment come with instructional materials and strategies? If not, does it link to instructional resources? Are the instructional materials and strategies linked to state standards or grade-level expectations?	Does the system offer any sort of next steps based on results? If instructional materials represent a separate component, what is their cost? What is their quality?	

Question	Relevance / Follow-up questions	Notes / Response
5. How do these interim assessments fit within our local and state assessment system and how do they link/connect to/support the other components in our system?	Teachers feel more positively about assessments when it is clear how they fit together in a balanced assessment system.	
6. For vendor purchases, is online or telephone technical assistance available and provided in a timely manner?	Especially when implementing a new system, support can be critical. Is the vendor willing to provide references of actual customers (other districts) who could address this question?	

PROFESSIONAL DEVELOPMENT		
Question	Relevance / Follow-up questions	Notes / Response
1. What professional development (PD) will various stakeholders want/need? How will that PD help us ensure that the data from the tests are used correctly and effectively?	What stakeholders need PD? Will the PD offered to different stakeholders be tailored to that group? If so, how?	
2. How compatible is the assessment with pre-existing or ongoing professional development provided or supported by the district?	Assessment that is coordinated with ongoing PD is more likely to be successful.	
3. Does the assessment system include professional development (PD)? If so, is the PD about how to use the product only, or does the PD include how to use the results?	Training on how to use the results is critical. Are teachers or trainers from the district expected to provide the PD?	
4. Have we planned for adequate professional development early, including prior to adoption of the system, if applicable?	Does the PD promote or reinforce buy-in? There will be discomfort and resistance to a new system. Scheduling professional development early can increase comfort with the system. In the best case, staff who had been oppositional can become advocates (Crane, 2008).	

ALIGNMENT TO STANDARDS AND CURRICULUM

Question	Relevance / Follow-up questions	Notes / Response
1. How aligned is the assessment to state academic content standards? Do we want the assessment to be aligned to any other standards in addition?	If the alignment to state content standards is not as good as desired, how should results and inferences be discussed differently? Can the vendor indicate to what standard the items *are* aligned? What happens and what are the consequences if the assessment is less than fully aligned?	
2. Has the vendor conducted an alignment process? Do we want to sponsor or conduct our own alignment study?	Review the vendor's materials to verify claims. Ask for additional details if materials are vague. How was the alignment conducted? Will the vendor be responsive to findings from an item review by the district's teachers?	
3. If there is an alignment study, are the following elements published: the number and percentage of items aligned to the standards, the number and percentage of standards for which there are no items, and the percentage of items that do not align to the standards?	Besides topics, is the cognitive complexity or depth of knowledge published?	
4. Do the assessments reflect a logical scope and sequence in curriculum?	Have the state and local curriculum standards been examined to ensure that what is being assessed is being taught? Is there a mechanism in place for teacher committees to study the assessment results and make curriculum recommendations or modifications?	
5. How do we want the assessments to "link" or connect to our existing textbooks and curriculum materials/resources?	Some purchased interim assessments may not match the timing of items in the curriculum.	

COSTS AND CAPACITY

Question	Relevance / Follow-up questions	Notes / Response
1. What is the cost of the system? Is the cost in line with the value of the system and with its goals?	Are additional resources needed to get full benefit from the system? Cost includes opportunity cost of other things the district could have done, instructional time, resources, staff time, and computer time.	
2. Can we afford it?	It is sometimes better to envision the ideal before focusing on reality. It may help with the visioning process. If keeping costs down is paramount, are there possibilities for collaboration with other districts?	
3. Do we have the staff time available for training in a new system?	Are the likely benefits so high that we will find the time, even if it appears that we do not have it available?	
4. Do we have the technical expertise to manage certain parts of the system ourselves?	The vendor may not be the only option for a consultant to help manage the system.	
5. Is this an online system hosted by the vendor?	Does the district need servers? Scanners? How well equipped are school sites? It is important to include information technology staff in the planning and discussion.	

This section, "Test Development and Technical Quality," contains more technical information than other sections of the workbook. Not all school districts will have in-house expertise regarding the technical issues of assessment development. We encourage school districts to work with qualified psychometricians or other advisors if the districts need additional technical expertise.

TEST DEVELOPMENT AND TECHNICAL QUALITY

Question	Relevance / Follow-up questions	Notes / Response
1. What type(s) of items do we want in the tests?	Does the assessment include tasks or performances? This is especially important if the state assessments emphasize these formats and the interim assessments are meant to monitor readiness for those exams.	
2. What access will various stakeholders have to the items? Will items be released?	If items are secure, this will impact remediation. Consider publishing release and exposure policy on the district's web site.	
3. How much involvement do we want to have in test development?	Do we want/need tests that are pre-made, or do we want a bank of items so that we can build our own assessments? Does the district have the necessary technology and professional expertise to support teacher-developed assessments? Does the system allow for item development by teachers?	
4. How much staff time do we have to conduct the tasks required by the administration? (See also "Administration and Scoring" section.)	Would other administration or scoring models require less (or more) time?	
5. Is the interim assessment predictive of other important measures or outcomes?	What kinds of data would back up claims of predictive validity for interim assessments? Be a critical consumer of vendor examples, asking questions about the evidence.	
6. How relevant is the information that was in the documentation of technical quality? What level of technical rigor do we require?	Sometimes high-quality testimonials and quasi-experimental studies are more relevant than a gold-standard research design (such as that required by the federal What Works Clearinghouse).	

Questions	Relevance / Follow-up questions	Notes / Response
7. How rigorous is the technical foundation of the assessment?	There is considerable range of district practice here. A gold-standard level of demonstrated technical quality will be appropriate for some districts. Other districts may be comfortable without the same level of demonstrated technical quality. Examples include: – detailed study design – links to instruction – thorough statistical and psychometric information – evidence that each item was edited for spelling, grammar, usage conventions, and cuing and item writing principles (adapted from Siskind and Potter, 2006; a more complete listing of Siskind and Potter's elements of a rigorous technical foundation is in Appendix C)	

ADMINISTRATION AND SCORING

Question	Relevance / Follow-up questions	Notes / Response
1. How often will we administer the tests? How are the limitations of our calendar and our equipment matched to these issues?	Will retesting be allowed with the same items? If yes, then item exposure may be a problem: exposure may be a problem if accountability is a purpose, but retesting may be advisable instructionally, if preparation is a purpose.	
2. What administration/delivery mode(s) do we want or need?	Paper-pencil, online, or both? Computer adaptive?	
3. Does the assessment have the flexibility to be administered at the group or individual level?	Can the test be administered to meet the needs of students with disabilities and English learners?	
4. What elements does the system include that ensure uniform administration for all students, except those receiving appropriate accommodation?	Can the test be administered at grade level only, or above or below also?	
5. What logistics related to the administration of the test can we support?	Will we do local scanning? Will we scan at each school or at a central location? What materials are needed to administer and score the tests?	
6. What technology resources are available and what will be needed?	Do we have the requisite number of computers? How long do data need to be stored? Will we track results over multiple years?	
7. How quickly can the tests be scored?	If open-ended items are included, what training will be provided to scorers?	
8. How will we score the tests?	Can instructional time be used for administration and scoring the tests?	

In this section, "Reporting," a critical framing question for the conversations can be, "How does reporting support the purpose(s) of the interim assessment system?"

REPORTING		
Question	Relevance / Follow-up questions	Notes / Response
1. Does the interim assessment allow for comparisons to other populations?	Does it allow for district-wide or statewide comparisons?	
2. Can the system help determine trends regarding systematic gaps or what is missing from the prior grade? Gaps associated with particular teachers?	Do assessment results seamlessly flow into a student information system? Do some stakeholders have a concern that results will be used for teacher evaluation, and if so, how do we address that concern?	
3. How quickly are reports generated and returned to schools and teachers?	Are there also parent reports? Reports that students can use?	
4. How easy are the reports to interpret and use? Is there a good representation of the change in a student's skills?	Are there examples available from other districts in our state? What inferences can be drawn from scores on the assessment?	
5. Are data reported in terms and formats that can be quickly and easily understood, disaggregated, and utilized?	What is the mode of reporting? Are there links with electronic student-level data systems? Does the assessment provide subscores? Is it set at the same level of difficulty as the state test?	
6. Who has access to the data? How secure are the data?	What are the privacy masking requirements of the assessment data? Are user-level data access rights clearly defined and documented?	
7. Can the system effectively track students who change schools within the district?	How often are student and teacher rosters updated? Can students who change districts be tracked if the same interim assessments were used in their previous district?	
8. Will the assessment integrate into a school's response to intervention (RtI) framework?	Do the assessment scores flow seamlessly into the databases and systems that support RtI?	

EVALUATION

Question	Relevance / Follow-up questions	Notes / Response
1. How do we know whether the program has worked?	It helps to establish measurable goals for subsequent evaluation.	
2. What is our plan to review the effectiveness of this program? When will that review take place? What should be evaluated during that review?	Clear goals and timelines are a must. What is our exit strategy if the evaluation does not show that the program is effective?	
3. Since many contracts with vendors are on a year-to-year basis (due to fiscal concerns), how can we tie renewal of the product to evaluation results?	Explore possibilities with vendors. A competitive or challenging economic time may present unprecedented opportunities for negotiation.	

Appendix A: Resources

Crane, E. W. (2008). Interim Assessment Practices and Avenues for State Involvement. Washington, DC: Council of Chief State School Officers. Available at http://www.ccsso.org/Publications/Download.cfm?Filename=InterimAssessmentPractices-web.pdf.

Perie, M., Marion, S., & Gong, B. (2007). The role of interim assessments in a comprehensive assessment system: A policy brief. Available at http://www.nciea.org/publications/PolicyBriefFINAL.pdf.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice 28*(3), 5–13.

Other References

Crane, E. W., Rabinowitz, S., & Zimmerman, J. (2004). Locally tailored accountability: Building on your state system in the era of NCLB. [Knowledge Brief] San Francisco: WestEd. Retrieved March 11, 2008, from http://www.wested.org/online_pubs/KN-04-01.pdf.

Siskind, T. & Potter, D. (2006). To Members of the State Board of Education and Members of the South Carolina Education Oversight Committee, memorandum, 9 November 2006, re Adoption List for Formative Assessments.

Appendix B: Eight uses of interim assessment, excerpted from Crane (2008)

In all, eight uses of interim assessment were identified: *diagnosis, prediction, preparation, placement, student evaluation, school intervention, promotion/graduation,* and *local accountability*. The purposes of interim assessments can stretch from the borders of formative assessment to those of summative assessment. For example, interim assessments are used formatively when they are used for diagnosis. At the opposite boundary of interim assessments are those with more summative uses, such as promotion/graduation and local accountability. The remaining uses are suggestive of more of a "middle ground" of purposes for these tests. The assignment of these uses along a continuum is intended as a rough guide; how tests are being used in real instances—the purposes to which the tests are actually put—should drive the thinking and any characterization about them.

At the formative margin:

– Where the goal is identification of weak areas of performance for subsequent remediation, the goal of the interim assessment is *diagnosis*.

In the interim assessment "middle ground":

– Some jurisdictions (and many test publishers) claim that interim assessment provides a good forecast of student performance on the high-stakes test. In this case, the test is used for *prediction*.

– Where interim assessments are cited as helping students become familiar and comfortable with a test that may mirror the mandatory high-stakes test later in the school year, the jurisdiction's goals include *preparation*.

– Some jurisdictions use scores on interim tests for *placement,* to help inform what courses the student should be taking next.

– Perhaps the oldest use of interim assessment is for *student evaluation*, as traditional texts and curricula often have included chapter and unit tests that provide an evaluation of student understanding of recently completed material.

– Some jurisdictions require schools in need of improvement to use interim assessment. In this case, interim assessment serves as one initiative in a *school intervention* strategy.

At the summative margin:

– Where the progress shown across interim tests affects students' passage to the next grade, the tests are used for *promotion*. Likewise, state *graduation* requirements, such as the body-of-evidence-based approaches of Rhode Island and Wyoming, may include demonstrating learning over time through progress on district interim assessments.

– Interim assessment's purposes may be directed at the school as well as at students, as school districts may include interim assessment data as an input in a system of *local accountability* (Crane, Rabinowitz, & Zimmerman, 2004).

Appendix C: Expanded listing of technical features in interim assessments, adapted from the 2007–2008 Formative Assessment Adoption List for South Carolina (Siskind and Potter, 2006)

– experimental or quasi-experimental study design

– adequate description of links to instruction

– adequate description of sampling plan

– indication of study duration

– adequate description of data analysis, including statistical techniques used

– adequate description of study's findings and their practical significance

– adequate sample size or repetitions in the study

– adequate statistical and psychometric information—description of field test sample; reliability indices; standard error for each score point; decision consistency indices; item difficulty for each item; item discrimination for each item; Differential Item Functioning (DIF) statistics for each item; evidence of bias and sensitivity review for all items; evidence that each item was edited for spelling, grammar, usage conventions, and cuing and item writing principles; information about score derivation; information about the interpretation of test scores

CCSSO

Council of Chief State School Officers

Building an Interim Assessment System:
A Workbook for School Districts