Workshop Deliverables

Prepared in support of the
SBAC Enterprise Systems Technology
Architecture Phase 1 Summary Report

for the
Smarter Balanced Assessment Consortium

20 January 2012
### Content

1. Introduction ................................................ 4

2. Concerns, Probability and Impact Table. .................... 5

3. User Roles and Goals ........................................ 6

4. Top Three Priority Exercise ................................. 12

5. Epics – Functional Requirement. ............................. 14

   - Student. ................................................................................... 14
   - Parent. .................................................................................... 14
   - Teacher (Long-Term Sub) ...................................................... 15
   - Principal .................................................................................. 15
   - Elementary School Principal .............................................. 16
   - School Psychologist ................................................................. 16
   - Proctor ................................................................................... 16
   - Juvenile Justice ..................................................................... 16
   - Pre-Report Production Psychologist ........................................ 16
   - Database Manager ................................................................. 16
   - Human Scoring Vendor .......................................................... 16
   - Scorer .................................................................................... 17
   - Media .................................................................................... 17
   - Scoring Director .................................................................. 17

6. Personas ............................................................. 18

   - Alvin ..................................................................................... 19
   - Simone .................................................................................. 20
   - Theodore “Teddy” ................................................................. 21
   - Chanti ................................................................................... 22
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. User Scenarios</td>
<td>Item Creation Lifecycle</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Scoring - Interim Assessments</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Test Creation and Test Delivery</td>
<td>42</td>
</tr>
<tr>
<td>8. Interim vs. Summative Assessments</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>9. Architecture Outputs</td>
<td>Interoperability Diagrams</td>
<td>57</td>
</tr>
<tr>
<td>10. Priorities</td>
<td></td>
<td>59</td>
</tr>
</tbody>
</table>
1. Introduction

This document reflects the outcomes from a series of workshops conducted by the SMARTER Balanced Assessment Consortium. Three workshops were held over three different weeks to capture the needs, requirements and desires of the Consortium in the design of the overall assessment system. These structured conversations had the ultimate goal to provide the necessary information to define the enterprise architecture for the assessment system.

The workshop participants included individuals from the SBAC Members and outside experts in the assessment and technology field; making up the collaborative. The following represents the workshop schedule and topics discussed:

- Week 1 – Scoring, Distributed Scoring, Reporting and Portals/Dashboards
- Week 2 – Item Authoring, Item Banking, Digital Library and Interoperability
- Week 3 – Test Creation, Test Delivery Platform, Adaptive Testing and Integration Framework

The documents that follow are not complete documents or thoughts. These simply represent the conversations that occurred during the three weeks. Follow-up conversations were held to further discuss several components.
2. Concerns, Probability and Impact Table

At the start of the workshops, a discussion was held to determine what the concerns of the participants included. A following conversation occurred to determine what the probability of the concern would happen as well as what impact it would have on the overall assessment system.

<table>
<thead>
<tr>
<th>CONCERNS</th>
<th>PROBABILITY</th>
<th>IMPACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple implementation render coordination and integration</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Item parameters drift because field test and operational test delivery platform differ</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Technology / capacity issues will prevent schools from accessing the software</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Many educators and administrators are “technology immigrants”</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>RFP may end up costing more based on decision we have made on architecture</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Development of proposed architecture exceeds expected costs</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Not enough time for adopters to plan for adoption</td>
<td>M/H</td>
<td>M/H</td>
</tr>
<tr>
<td>System outages - caused by uncoordinated change management</td>
<td>M/H</td>
<td>M/H</td>
</tr>
<tr>
<td>Interoperability will be a major hurdle for States (sync with existing systems)</td>
<td>M/H</td>
<td>M/H</td>
</tr>
<tr>
<td>No time for comprehensive usability testing</td>
<td>M/H</td>
<td>M/H</td>
</tr>
<tr>
<td>Inconsistent UX</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>No profit motive - no sustained innovation</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>Not having pilot State / region all through period A &amp; B</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>States not engaging often and early</td>
<td>M</td>
<td>M/H</td>
</tr>
<tr>
<td>State / region resistant to adoption grow uncertainty</td>
<td>M</td>
<td>M/H</td>
</tr>
<tr>
<td>Architecture not able to accommodate emerging trends</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Scope changes</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Constant conflicting priorities (some states may like more technically depth, others breadth)</td>
<td>M/L</td>
<td>L</td>
</tr>
<tr>
<td>Availability of test design from architecture input</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>Perception of inclusiveness by community</td>
<td>M/L</td>
<td>M/L</td>
</tr>
</tbody>
</table>
3. User Roles and Goals

In order to understand all the potential users of the system, an activity took place to determine what the users were as well as the specific goals of the user. Three groups representing scoring, item creation and test creation and delivery participated in this activity.

<table>
<thead>
<tr>
<th>Process</th>
<th>User</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoring</td>
<td>Student</td>
<td>Understand where I am in my current year</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Be college-/career- ready</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Know how my achievement compares with that of my peers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Know if I am completing my requirements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Know what I should be doing next</td>
</tr>
<tr>
<td></td>
<td>Parent</td>
<td>Make sure my child's goals are being met and I have visibility into that progress</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contribute towards and enhance my child's learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collaborate with teachers and school staff</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Want to know how to help my child</td>
</tr>
<tr>
<td></td>
<td>Teacher (long-term sub)</td>
<td>Monitor student progress</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maximize student performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plan intervention and curriculum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collaborate with other teachers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Know whether the program meets its goals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicate with parents about student progress</td>
</tr>
<tr>
<td>Role</td>
<td>Requirements</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Teacher (short-term sub)            | Not be a babysitter to the students  
Quickly get back on path  
Get information on what my tasks are for the duration of the substitution  
Understand where the class currently is |
| Special Education Teacher           | Options for different input/output devices to accommodate needs of my students  
Want visibility of my students' performance with other teachers  
May have a different breadth of assessment  
Bypass the system if system doesn't handle exceptional cases |
| Principal                           | Want to know how my school is doing  
Want to know how my teachers are doing  
Collaborate with other principals  
Communicate to authority the school's progress and goals |
| Elementary School Principal         | Track school progress  
Able to drill down / facilitate dialog around reports  
Present monthly progress reports to cabinets  
Looking for performance data for teacher performance reviews  
Engaged parents |
| School Psychologist                 | Assess academic strengths and weaknesses  
Provide psychological services to students |
<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
</tr>
</thead>
</table>
| Proctor                                   | Give test & monitor students during tests  
Ensure state / school procedures are followed  
Initiate students' test sessions          |
| Higher-Ed Placement Officer               | To place students in colleges or to properly remediate classes  
Review admissions criteria for their colleges |
| Intervention Specialist/Tutor             | Review reports  
Review student-level data  
Access resources                           |
| Juvenile Justice                          | Determine if students are performing in their schools  
Act as a proxy in students' home schools |
| Guidance Counselor                        | Help students choose past High School choices  
Direct student data to colleges or military |
| Pre-Report Production Psychologist        | Act as bridge between scoring and reporting  
Analyze reporting, audit data to verify information is/remains accurate |
| Database Manager                          | Ensure systems can talk to each other  
Configure, calibrate student data          |
<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Scoring Vendor, Scoring Director</td>
<td>Monitor raters’ performance</td>
</tr>
<tr>
<td></td>
<td>Train raters</td>
</tr>
<tr>
<td></td>
<td>Provide information back to schools</td>
</tr>
<tr>
<td></td>
<td>Maintain scorers’ performance</td>
</tr>
<tr>
<td></td>
<td>Conduct range finding to create rater training and certification documents</td>
</tr>
<tr>
<td></td>
<td>Certify human raters</td>
</tr>
<tr>
<td></td>
<td>Manage scorer adjudication processes</td>
</tr>
<tr>
<td></td>
<td>Receive and return scores as required by contract</td>
</tr>
<tr>
<td></td>
<td>Alert client of unusual responses (plagiarism, cheating, disturbing content)</td>
</tr>
<tr>
<td>Scorer</td>
<td>Get feedback on my scoring</td>
</tr>
<tr>
<td></td>
<td>Improve at evaluating student work</td>
</tr>
<tr>
<td></td>
<td>Score responses as trained in time expectations</td>
</tr>
<tr>
<td>AI Scorer</td>
<td>Train engine to score responses to items that are already hand-scored</td>
</tr>
<tr>
<td></td>
<td>Produce reports on performance of AI engine on responses</td>
</tr>
<tr>
<td></td>
<td>Monitor AI scoring operationally, for score quality and performance</td>
</tr>
<tr>
<td>AI/Human Scoring Director</td>
<td>Manage, ensure proper flow of data between different scoring processes (AI, human, AI + human)</td>
</tr>
<tr>
<td></td>
<td>Ensure scores are returned as required by contract (valid scores, timely scores)</td>
</tr>
<tr>
<td></td>
<td>Produce reports on quality of scoring (AI, human, AI + human)</td>
</tr>
<tr>
<td>Program Evaluators</td>
<td>Evaluate district programs</td>
</tr>
<tr>
<td></td>
<td>Review statistical data to evaluate interventions</td>
</tr>
<tr>
<td>Role</td>
<td>Responsibilities</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Local Board</td>
<td>Make comparisons among schools within and across districts</td>
</tr>
<tr>
<td></td>
<td>Evaluate principals and teachers</td>
</tr>
<tr>
<td></td>
<td>Use summary data to help set policies</td>
</tr>
<tr>
<td>US ED Evaluator (Accountability)</td>
<td>Evaluate SBAC</td>
</tr>
<tr>
<td></td>
<td>Evaluate adequate yearly progress (AYP)</td>
</tr>
<tr>
<td></td>
<td>Report to the board or superintendent</td>
</tr>
<tr>
<td>Policy Makers</td>
<td>Evaluate teachers, staff, and users of the system</td>
</tr>
<tr>
<td></td>
<td>Look for audience to support funding opportunities</td>
</tr>
<tr>
<td></td>
<td>Compare and contrast state-to-state performances</td>
</tr>
<tr>
<td>Media</td>
<td>Get information on relevant school or district performance</td>
</tr>
<tr>
<td></td>
<td>Produce interesting stories</td>
</tr>
<tr>
<td>Item Creation</td>
<td></td>
</tr>
<tr>
<td>Item Reviewer</td>
<td>Ensure the item is unbiased, does not disadvantage certain groups, and is free of insensitivity</td>
</tr>
<tr>
<td></td>
<td>Ensure that items are of the appropriate grade level and standards, fair to all students, and accurate to all constructs</td>
</tr>
<tr>
<td></td>
<td>Work with item organizer or developer of item if there is a required change or revision</td>
</tr>
<tr>
<td>Item QA</td>
<td>Ensure item layout is correct: screen design, grammar correction, layout and style, animation (such as drag and drop), supported by X browsers</td>
</tr>
<tr>
<td></td>
<td>Test the item</td>
</tr>
<tr>
<td>Role</td>
<td>Responsibilities</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Content Specialist                       | Item specification, selection, assignment  
  Participate in item review  
  Ensure accessibility  
  Manage translation |
| Item Tagger                              | Ensure item has appropriate metadata  
  Tag accessible elements so special education students can consume the item |
| Item/Task Originator and Developer       | Get item and tasks into the system as efficiently as possible  
  Identify the constructs, give specification on an item  
  Improve an item to make the item better, e.g. language, content, accessibility  
  Identify stimulus to create items  
  Make sure I have copyright to the item I create  
  Provide input on accessibility issues  
  Formulate templates or models from which items are produced  
  Get feedback from cognitive lab review to see how well the templates worked and improve the process over time |
| AI Reviewer                              | Review items for their AI scorability  
  Suggest changes to enhance scoring |
| Test Creation and Delivery               | Delegated Authority TBD  
  Test Creator TBD  
  System Proctor TBD |
4. Top Three Priority Exercise

Each workshop was presented with the exercise to name the top three priorities for the overall system. The following list represents each priority provided.

- Access to assessment data with instructionally sensitive level of detail
- Adaptive engine
- AI scoring to keep costs low and reduce scoring time
- Being competitive: the difference between adopting and following policy
- Calibration application – mechanism to calibrate tests
- Comparability for linear and CAT tests
- Computer-adaptive, unconstrained by grade levels
- Easy blueprint creation
- Exchange of best practices for common core instruction (collaboration)
- Extensible item types renderers
- For canned reports, only use canned reporting functions that everyone agrees have the most value. More complex reports should be left to the state
- Item path – authoring to student to growing to long-term data use – including adaptive issues
- Meaningful/actionable reports that provides actionable information and results to teachers
- Performance summaries for each student and classroom
- Reporting capability standards
- Reporting
- Should be focused on “live” dashboards that teachers can use to guide instructional and classroom mgmt. decisions (eg: student grouping for projects or for remediation/enrichment).
- State reports intended for compliance purposes should be a side effect (design for analysis and let the compliance reports be preset views of the “cube” of data)
- SIF/APIP/Tagging criteria
- Student test delivery standards
- Teacher-created items have a very different workflow, and if this is not well understood we could have a threat to assessment validity
- The architecture should support the capture of very rich student response data generated from constructed-response and performance tasks. This supports not only scoring, but the ongoing R&D necessary to evolve such items and the adaptive algorithms that can identify what’s next in near real time for students
- Use/incorporate high-quality tech-enabled items
- User defined interim testing options (ability to define own test blueprint)
The data was then compiled. A summary of data includes:

- Feature requests x23
- System design items x10
- Items that are concerns x3
- One item said “TBD”
5. Epics – Functional Requirement

These were captured during the user goals identification workshop and from the top three priority exercise. An epic is large feature, or a grouping of smaller features. These groups include a requirement that accomplishes a single identified goal in software development. Epics are sometimes referred to as stories. The following are the epics that were identified.

**Student**
- Take tests
- Access resources
- Access practice items and model responses
- Trust that my information is secure
- Gets results quickly
- System is customized to my needs
- Can access my information any time

**Parent**
- School has accurate information on my child
- See expectation for my child
- One-stop-shop for information
- See how the school/teachers are doing, see history compared
- Be notified of ways to help my child learn
- Be notified if my child “goes of the path”
- Receive information in different ways – redundancy
- Receive information customized to my needs (eg. Language)
- Access with existing hardware
- Allow me to see assessments before my child takes the test
- Want to know how test results will be used and in what form
- My child could take charge of his/her learning
**Teacher (Long-Term Sub)**

- Administrate tests
- Drill down student profiles
- Create tests
- Give formative and interim tests
- Understand SBAC system components
- See instructional professional development
- Create educational plans tailored & effective to student
- Deviation from standards, enhancement or acceleration
- Individualize to student
- See information on how my class is doing
- How well are my teaching materials meeting my students' needs
- Verify whether students have learned info up until certain point
- Control over-assessments
- Offering differing levels of functionality/help based on my level of expertise and content knowledge
- Understand variability in assistance between elementary vs. HS
- Know about the test program scheduling
- Have appropriate access to data at diff levels
- Get info about incoming students and history so that I can prepare
- Have access to non-achievement data
- Assessment info is understandable for all parties

**Principal**

- Primary summary info on plans
- Understand explanation of results
- Interpret results
- See teacher-level data
- Evaluations (out of scope)
Elementary School Principal
- Building-level test administration
- Teacher performance reviews & performance data

School Psychologist
- Report review
- Look at attendance data

Proctor
- Access student records on behalf of students
- Ensure students receive the correct tests

Juvenile Justice
- Access real-time reports

Pre-Report Production Psychologist
- Report equating and calibrating
- Conduct report analysis

Database Manager
- Concerned with data integrity
- Facilitating / using the interoperability –loading / configuring students
- Data transforming between systems
- System maintenance

Human Scoring Vendor
- Range finding
- Score anchor papers
- Onsite or distributed scoring
- Scoring – able to integrate w/platform to get data I need; put data back in
- Reporting vendor – extract data
**Scorer**
- Monitor scores that come out of AI scores
- Score things that can't be scored by machine
- Factor in historical information in scoring because things change over time

**Media**
- Access to reporting to see relevant school and district performance
- Access to comparison data

**Scoring Director**
- Understand discrepancies between AI and manual scoring
- Identify patterns in the scoring
6. Personas

Once user roles and goals had been identified, personas were generated. Personas intend to provide specific information about a user, including background information, some demographics and the motivation of the user.

These personas are fictional. The personas identified represent a sampling of users for the SBAC assessment system and do not represent each extrapolation of user. For example, several students are identified instead of creating personas for every possible representation of a student. The purpose of the personas is to ensure that the enterprise architecture can manage the necessary requirements of each application within the architecture.
Alvin

Basic Information
- 7TH Grade Student
- Public Middle School
- Boise, ID

Background
- Uses FB all the time
- Owns a smartphone
- Interests: video games, soccer
- Does not like to read
- Medium proficiency
- Passes his classes/tests
- His parents are engaged in his education

Motivation
- Does not want to be harassed
- Peer acceptance
- “More likely to do more if he could earn something” à leads to progress (game theory)
Simone

Basic Information
- 11TH Grade Student
- Hispanic High School
- San Diego, CA

Background
- Level 3 ELL Student
- Has some academic issues due to her English skills.
- Has been in the US for the past 2 years
- Oldest of 3 daughters
- Speaks Spanish at home
- More proficient in English than her parents
- Heavy texter, not a smartphone user
- Above average student in her school
- Would have advanced proficiency if not for her English language skills
- Has an active guide at school who helps her navigate/get value out of the system

Motivation
- Wants to pass the high school exit exam
- Wants to be the first in her family to complete college
- Would like to get a stable career
- “Fitting in” is big on her mind
- More likely to use a system that shows her program to mastery via the interim system
Theodore “Teddy”

Basic Information
- Age 45
- D.O.E as Math and Science Curriculum Consultant
- Lives in West Valley, UT

Background
- Consultant for 3 years
- Has 3 children: 17, 4, 3
- Uses Skype to call his kids
- Wife teaches dance
- He is a scorer
- His job is mapping state standards onto Common Core State Standards (CCSS)
- Former high school math teacher for 15 years
- Believes in making common core stronger
- Overworked, on the road a lot, does a lot of webinars

Motivation
- Successful implementation of CCSS
- Believes he can help students
- Feels a strong link to his former colleagues
- Feels he could make more impact
- Took a pay cut to be in this role
- Wants to see more students graduate/ready for college
- Wants to be more effective covering a broad range of levels
- Outside of his comfort zone *
- Reaching out to content experts (esp. to subjects *), inside/outside his state
- Wants to show teachers how to use the system
- Active participant in the community
Chanti

Basic Information
- Age 38
- Elementary School Principal
- Greenwood, SC

Background
- Has been principal for 2 years
- Was a teacher for 12 years
- Earned degree in Elementary education from USC
- Later earned her Masters of Ed. D
- Has never left Greenwood, SC (outside of college)
- Has 2 kids: 12, 11
- Has 26 friends on FB
- Uses a blackberry issued by the school
- Knows every kid by name
- Very committed to her profession

Motivation
- Principal is not her last stop
- Very connected to the community
- Cares about her teachers
- Believes in personalized learning and technology
- Her school lacks the technology emergence
- Concerned about the true growth of her kids (100 of them at her school)
Ella

Basic Information
- Age 41
- Hand-Scoring Manager
- Works for ACME, Inc. (Vendor)

Background
- Working in state contract: MS
- M.A. in English language
- Has 7 years in hand-scoring, worked her way up
- Manages up to 30 other scorers, the scoring process
- Works on a Dell desktop at work
- Feels confident in her team
- Gets confused by MS-Office updates
- Does not have “smart” tv with internet browsing at home
- Reads traditional print books
- Would not machine score
- Skeptical about A.I. scoring
- Feels passionate about her job

- Attends range-finding meetings
- Client interfacing: vendor sends scores from the state
- Works for office assessment
- Manages/aware of rater training exercises
- Has a son with language disability
- Believes the system
- Under pressure to score in quantity + quality + shorter time
- Worried about her job: machine scoring
- Acts as liaison between DOE and suspected cheating/altered papers
- Has a low tolerance for machine mistakes
- Spends her evenings helping her son with his homework that he has to write.

Motivation
- Secret ambition: to write a novel
- Keeping her job
- Wants company to produce high-quality scoring
- Wants to demonstrate that human scoring has a place
- Does not have/use the systems at home
- Understands the value of including constructive-response and multiple-choice assessments
- Does not believe other scoring methods can achieve to her levels
- Has strong opinions on prompts (questions)
- Resisting the increasing requirements on paper tests
Roger

Basic Information
- Age 40
- Works for Vendor
- Lives in Nova Scotia

Background
- Has been working for item-creation vendor for 13 years
- Taught for 7 years as an elementary ELA teacher
- Writes for ELA content
- Gadget savvy, builds his own computer
- Works while he is on the move, can reconnect to continue his item creation

Motivation
- Wants to be able to travel, have flexibility to do personal travel. Works from home.
- Interested in how kids learn
- Reads about the latest theories and research
- Wants to contribute and appreciate the opportunity afforded by this new system
Naomi

Basic Information
- Age: 30
- Item/Task Originator
- Middle School Math Teacher

Background
- Teaching math for 5 years, taught common core for 2 years
- Knows differentiation
- Completed online item writing
- Submitted sample items to the consortium
- Has a desktop computer at school
- Works long hours as a teacher
- Has a dog

Motivation
- Does not like the items that she had seen
- Works collaboratively with her colleagues to derive items, which she enjoys
- Sees this as additional income during the summer
- Wants to develop her professional network outside of her school, because she lives in a small town
- Wants to improve her assessment skills
Freddie

Basic Information
- Item Developer
- Consultant for a Vendor

Background
- Specializes in Flash technology
- Has no experience in content development
- Background in graphical design
- Needs direction from content specialists
- Benefits from collaboration – more collaborative than the typical software developer

Motivation
- Likes to define short-term projects
- Needs to be part of something creative
- Likes being attached to the education field
- Likes to brag about his accomplishments
- Very “intense” in his manner of working
Trudi

Basic Information
- Content Specialist
- Senior Level in Test Development Company

Background
- Provides content leadership
- Works with accessibility professional to solve issues

Motivation
- Has an enormous amount of pride in her work
- Feels responsible for the quality of her product
- Wants to escape the political environment in schools
- Likes to feel independent
- Thrives on the cyclical heroism in her job
- Thrives in the interpersonal and collegial interaction
7. User Scenarios

This section contains the following user scenarios generated in the SBAC architecture workshops:

- Item creation lifecycle
- Scoring
- Test creation and delivery

A user scenario is a tool that provides a broad understanding of typical user interaction with the components in a system. While the tool does not provide an exhaustive illustration, it does provide valuable insight for making architectural decisions. The diagrams and flows intend to provide the business requirements and logic for the scenarios.
**Item Creation Lifecycle**

1. **Identify need for item**
   - Input by test creation system
   - Input by claim
   - Input by item bank

2. **Generate requirement**
   - Template
   - Identify item characteristics
   - Choice of attributes to fulfill

3. **Author item**
   - Performance tasks move forward as a group

4. **Item review**
   - Includes AI review to ensure scorability
   - Reviewers have to be trained

5. **Internal processing**
   - Internal checks done?

6. **External review**
   - Technology review
   - Accessibility review
   - Tolerance check in T. E. items

7. **Need traceability**
   - Items need to be secured to authority

8. **Can item be seen by students?**
   - Adjust item as specified

9. **Needs pilot?**
   - Item pilot
   - Item field test
   - Item try-out

10. **Needs field test?**

11. **Needs try-out?**

12. **Passed review(s)?**

13. **Item rejected**

---

**Stimuli does not need testing**

- Committee members

**Need traceability**

- Items need to be secured to authority

---

**Available to public**

- Can be seen by students
- Item is scorable
- Item cannot be edited
- New items can be derived (as new)

---

**Integration tester**

- Tagging is done before item is seen by students
- Versions of an item need to be saved with each item; including notes (on which version, viewed by whom)

---

**Content specialist**

- Media writer

**Content analyst**

- Item writer

**Reviewer Editor**

- Committee members

---

**Early on: accessibility issues review**

- Input by test creation system
- Input by claim
- Input by item bank
Scoring - Interim Assessments

This shows the initial breakdown of interim tests to help identify which business process to further investigate.

INTERIM BRAINSTORM
Each of these areas was then elaborated on to create user scenarios and highlight points that are of architectural concern.

**Interim Assessment – Constructive Response Item**

1. **STUDENT SEES QUESTION**
2. **STUDENT ENTERS RESPONSE**
3. **COMPLETELY SCORED BY MACHINE?**
   - **Y**
     - **TEST IS CONCLUDED, RESPONSE IS SCORED**
   - **N**
     - **RESPONSE IS STORED FOR TEACHER TO SCORE**
       - *(Can see how to assign appropriate scoring)*
     - **SCORES MAY BE REVIEWED BY TEACHER* (Revised or commented)*

* This is an optional step
An algorithm has:
- Set of scores
- Proficiency levels
Teacher Creates an Interim Test – Human Scoring, Constructive

TEACHER CREATES AN INTERIM TEST - HUMAN SCORING, CONSTRUCTIVE

1. **Teacher Identifies Need for a Test**
2. **Teacher Specifies Content Domain of Test** - Subject area, Grade level, Sets of content standards
3. **Teacher Specifies Additional Criteria**
4. **Teacher Selects Test Items**
5. **Teacher Administers Test**
6. **Student Takes Test**
7. **Test Is Scored**
8. **Test Results Are Reported** (Exported/Published in a particular format)
9. **Test Results Are Monitored** - Item, Student Info
SUMMATIVE, COMPUTER-BASED SCORING

STUDENT SEES QUESTION → STUDENT GIVES RESPONSE → RESPONSE COMPLETELY SCORED BY MACHINE?

Y → CONTINUE TO NEXT QUESTION UNTIL TEST IS COMPLETE

N → RESPONSE IS DISTRIBUTED TO SCORERS

- Distribution of read-behind scorers
- Monitoring

Assume computer-based scoring same as “AI scoring”

Need to take into account latency times here and ensure architecturally that they will not interfere with C.A.T.
Primary, Paper-Based Interim Test

**PRIMARY, PAPER-BASED INTERIM TEST**

1. TEST IS CREATED
2. TEST IS PRINTED
3. TEST IS ADMINISTERED
4. IS TEST SCANNED + COMPLETELY SCORED BY MACHINE?
   - **Y**: TEST RESULTS ARE RECEIVED
     - results are stored somewhere
   - **N**: TEST IS HAND-SCORED
5. OPTION TO INPUT RESULTS TO SBAC SYSTEM

**Policy Discussion**
MONITORING - PERFORMANCE OF RATER-SUMMATIVE

CERTIFIED RATER GETS ASSIGNED TO ITEM

RATER SCORES A RESPONSE:

STUDENT RESPONSE

RATING OF THE STUDENT RESPONSE IS SCORED
- Rater info. also stored

ANCHOR PAPER

RATING OF ANCHOR PAPER VS. ANCHOR RATING
- Info. about how rater did is stored

READ-BEHIND PAPER

RATING OF READ-BEHIND VS ITS FIRST TIME RATING
- Info is stored in system

-- May need to distinguish between read-behind and second-read

-- Need to deal with adjudication rules for producing scores of record

- Non-quantitative data is also captured:
  - Accuracy of rating
  - Speed of rating
Mixed Paper-Based + Computer-Based Interim Assessment

**MIXED PAPER-BASED + COMPUTER-BASED INTERIM TEST**

- Test is created
  - content
- Paper elements + computer elements of the test is determined
- Computer element is created

- Test is administered
- Paper element is hand-scored
- Paper element is scanned
- Computer elements are machine-scored
- Computer elements are human-scored

- Test results are merged to form a score
- Test scores are stored in the system

---

If there is an item that requires a combination of AI and human scoring, some scoring management is needed.

- Accept/ - access options to be determined?
  - who can/cannot see it?
Paper-Based Summative Test

PAPER-BASED SUMMATIVE TEST

- Test is Created
- Deliver Test to School (By Vendor)
- Administer Test (By School)
- Vendor Scores Test
- Test Scores are Available In System
Monitor Vendor-Scored Summative Test

MONITOR VENDOR-SCORED SUMMATIVE TEST

System receives a batch of data → Analytics are carried out → Action needed? → Y Alert → Corrective action is taken → This could be re-marked again: versions of scores need to be saved → N Analytics are scored

-- Raters
-- Item

-- Capture event info
-- Who resolved issue
-- Capture any relevant notes

Note: We often rely on HS vendors to do this work and report their statistics on rater monitoring
Performance Task - Interim Assessment

PERFORMANCE TASK - INTERIM ASSESSMENT

See constructive response for scoring / selective

There will be a mix of AI and human scoring, with AI scoring some parts of tasks and humans scoring others. This will require the architecture to accept different scoring sources.
Monitoring – Crisis Papers

Monitoring -- Crisis papers

- Rater concern is triggered
- Based on content or pattern e.g. plagiarism

- Rater alerts supervisor
- Paper info
- Student info
- Rater's comment

- Rater supplies relevant information to client
- Client takes action

- Notify district?
- Client sends relevant info to district
- Client marks issue as resolved

- Client = State
Test Creation and Test Delivery

To help the group understand the processes of test creation and test delivery, we used this diagram to loosely illustrate the steps involved. The following diagrams break these steps down into more elaborate details.

UNDERSTANDING TEST CREATION & TEST DELIVERY

TEST CREATION

- Identify need
- Define blueprint
- Acquire items, simulate
- Publish
- Ready the test

-- Things that need to be set up to deliver tests

TEST DELIVERY

- Schedule test
- Authenticate student
- Administer test
- Scoring & Reporting

-- Include registration
-- Apply program characteristics
We explored how states may use the SBAC test creation and delivery system differently. While not comprehensive, the group provided six examples of typical SBAC system set up, and explored who may use these set ups. The dots under each state are color-coded. For example, State 1’s set up has test item bank and test delivery system.

UNDERSTANDING TEST CREATION & TEST DELIVERY (CONTINUED)

ITEM BANK ARCHIVE
Contains all stages of items:
-- Dev
-- Ready for field test
-- Operational
-- Retired
Item is typically more thorough and complete

TEST BANK
Contains tests that are:
-- Ready for field test
-- All tests
-- Published, that can be delivered
-- Some portion may be complete

TEST
Contains:
-- Algorithms
-- Metadata

Test bank contains information to deliver a test
**Test Creation – SBAC Owned**

This diagram represents both interim and summative test creation.

**TEST CREATION - SBAC-OWNED SYSTEM**

1. **Identify need**
2. **Define blueprint**
3. **Acquire items, simulate**
4. **Package**

Assumptions:
- Item exposure control is in place
- Delegated Authority

Arrow indicating: -- Select one, or build one
**Test Package**

Contains

Eligible items: operational and field test items. Eligible items could be defined as operation and field test items. Eligible and operational items are not mutually exclusive.

- Test design: test definition or manifest (used in the technical creation of the test)
- Order of test: order of items as specified in a test.
- Media assets
- Item metadata
- Multiple packages to prevent security compromises. Multiple packages can contain components of the same test.
- Blueprint
- Test items
- Structure
- Blueprint, test items, structure are not mutually exclusive with the other items on the list.
Test Bank

Contains

- Test packages
- Items (that are ready for test)
- Test definitions

Blueprint

This illustrates what a blueprint may contain. If we imagined a table of contents for a blueprint, it would contain the items shown below.

- Table of Contents
- Grade level
- Subject area
- Grade-level subcategories
- Allocation of content across scoring categories
- Functional and non-functional
- Quantities (number) of test items
- Quantities (number) of graphics
- Duration of test
- Skills assessed
- Define categorization
- Grouped items
- Description of pilot space
- High-level information on scoring
- Define implementation constraints (does not contain exceptions)
- Accessibility
- Design and style manual: references the styles to use for this test and the subgroup
Test Creation – State, LEA or School

Two scenarios are represented in the diagram. One is for a blueprint and one is when there is no blueprint. The “no blueprint” scenario was discussed for formative assessment.

TEST CREATION - STATE OR LEA OR SCHOOL

WHEN THERE IS AN AVAILABLE BLUEPRINT

Select a blueprint → Package

WHEN THERE ARE NO AVAILABLE BLUEPRINTS

DA identifies a need → Define blueprint → Validate blueprint → Package

-- Stats
-- Time
-- Item types
Define Blueprint

This describes how a blueprint is defined. The stick figures illustrate the users who are involved in each process above.

**DEFINE BLUEPRINT**

- **Verify outcome**
  - -- What data is needed for which group

- **Parameter analysis**
  - -- Test costing tool?
  - -- Negotiations

- **Define scope categories**
  - -- Finalize scoring outcomes
  - -- Know what is feasible
  - -- Capture free-form text, decisions into the system

- **Build blueprint**
  - -- Pick algorithm
  - -- Define content structure

- **Vetting process**
  - -- Feasibility checks
  - -- Reviews on content, costs (length of time)
  - -- Simulation
  - -- We have item bank
  - -- Look at the items
  - -- Functional QA
  - -- Results added to tech personnel
  - -- Evidence maintained
  - -- Versioning
  - -- Publishing

---

End users = teachers, principals
SMEs = Psychometricians, content specialists
Test Creation

This illustrates the steps to create a test.

**TEST CREATION - ADAPTIVE TEST, SUMMATIVE**

1. **Define test design blueprint**
   - Identify purpose
   - Alignment to common core learning standards
   - Develop framework (draft content specification)
   - Number of items
   - Characteristics of test

2. **Specify content coverage of test event**
   - Build pool to optimize test
   - Pool analysis
   - Run algorithm in pool to validate there are sufficient items
   - Item selection criteria
   - Use IRT model

3. **Test simulation**

   *Activities in test simulation:*
   - Run algorithm in pool to validate there are sufficient items
   - Item selection criteria
   - Use IRT model

4. **Does pool have sufficient items?**

   *Not enough / non-existent*

   - **Create items for test**
   - **Populate pool**
   - **Adjust blueprint**
Test Simulation

The purpose of a test simulation is to ensure that the test blueprint contains sufficient items in the item pool to build the test. Simulating a test provides information and prompts a call to action, should the pool not have sufficient items of the specified criteria. Simulating a test does not test the content (items) itself.

Activities involved:

- Testing the algorithm
  - Delivers blueprint consistently
  - Specifies pool depth
  - Reporting on optional pool
  - Provides statistical functions of test
  - Validates expected result
  - Tests item-by-item or multi-stage
  - Tests for latency
- Testing the pool
- Testing the student results
- Employing a selection and scoring algorithm
- Adding timing data to simulations. With a mix of CR and MC items, you must know how the expected response time will factor into the overall test length.

Types of Tests

- Linear
- Adaptive
  - Staged adaptive (testlet)
- Paper-pencil test
- Item adaptive
- Cognitive diagnosis (adaptive)
- Pick vs. Do – Student picks an answer, versus performs an activity that is required for a score
Ready the Test – Steps Involved

READY THE TEST - STEPS INVOLVED

*Goal: Reduce irrelevant varieties across students*

- Ready the proctors
  - TAMS
  - Tutorial
  - Practicing giving the test
  - Role permissions

*Goal: Don’t let taking the test get in the way of the results*

- Ready the students
  - Make them familiar with how to run test
  - Experimenting with accessibility options
  - Practise tests
  - Not secure items
  - Practice item bank

- Ready the environment
  - Check if hardware is ready to run test
  - Check software
  - Check internet connections

- Ready the helpdesk
  - Provide support with as little human interaction
  - Diagnostics
  - Pre-emptive help/fix

- Ready the IT infrastructure
  - Notification
  - Network setup
  - Checklist
Test Delivery

TEST DELIVERY

-- Establish SBAC window
-- Verify states admin within window
-- What gender
-- When
-- What kids
-- Where

-- Possibly register new students
-- Proctor acts first, but has to do this

-- Whole test
-- By student

Configurable options:

[A] Student could see whether they have answered correctly
[B] Can configure if irregularities need to be logged
[C] Configure whether students take the survey or not. Can switch on / off for both interim or summative
[D] Can configure whether the test responses are send off for analysis, for example, to vendors.

Choose test from list of valid for range
Get list of students' accessibility info
Get eligible proctors

Visibility into test actions:
-- Timing
-- How far along
# 8. Interim vs. Summative Assessments

In order to provide clarity around interim and summative assessments, a discussion was held to determine what is the same between interim and summative assessments and what is different. Consensus was made around each of these points in the following table.

<table>
<thead>
<tr>
<th>Interim is the same as summative</th>
<th>Interim is different from summative</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Reflects student understanding of CCSS+</td>
<td>Purpose is to inform instruction more than summative++</td>
</tr>
<tr>
<td>Adaptive +++</td>
<td>less high stakes/relaxed stakes+++</td>
</tr>
<tr>
<td>On same psychometric scale++</td>
<td>*Non-secure/diff security++++++</td>
</tr>
<tr>
<td>It's another theta hat-estimating ability</td>
<td>*Voluntary for states (subscription, OSS, Creative Commons)+</td>
</tr>
<tr>
<td>Assesses same constructs+</td>
<td>*Greater flexibility for customization++</td>
</tr>
<tr>
<td>Type and level of items+</td>
<td>Needs flexibility in picking/constructing blueprints+</td>
</tr>
<tr>
<td>Look and feel</td>
<td>*Flexibility classroom level+</td>
</tr>
<tr>
<td>Delivery system</td>
<td>*May be fixed form or unstructured</td>
</tr>
<tr>
<td>Item bank and test bank could be same</td>
<td>*Options for mini-summative or informative (e.g., students could see if they get questions right)</td>
</tr>
<tr>
<td>Produced and authorized by SBAC</td>
<td>State option+</td>
</tr>
<tr>
<td>Feedback on student performance</td>
<td>*Administered more often during year+</td>
</tr>
<tr>
<td>Computer delivered</td>
<td>Time synchronization harder (when taught and when tested)</td>
</tr>
<tr>
<td>*Overall Structure</td>
<td>Possibility to integrate other administration options (clickers, iPads)</td>
</tr>
<tr>
<td>*Process of item creation, test assembly, test delivery and reporting</td>
<td>Specifications</td>
</tr>
<tr>
<td>Monitor item exposure</td>
<td>Results are used for different purposes</td>
</tr>
<tr>
<td>“mostly” disconnected from instructional modalities and models</td>
<td>Smaller concurrent population</td>
</tr>
<tr>
<td>Flexibility for alternate delivery systems</td>
<td>Teachers may score own students' work</td>
</tr>
<tr>
<td></td>
<td>*Teacher ability to customize and choose items</td>
</tr>
<tr>
<td></td>
<td>*Based on local curricula and instruction units+</td>
</tr>
<tr>
<td></td>
<td>Initially will need to populate with a pool of items. Operationally—interim items will be items released from summative</td>
</tr>
<tr>
<td>Interim is the same as summative</td>
<td>Interim is different from summative</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Uses non-secure tasks</td>
<td></td>
</tr>
<tr>
<td>Reporting detail</td>
<td></td>
</tr>
<tr>
<td>Level of precision / diagnostic orientation vs. classification orientation</td>
<td></td>
</tr>
<tr>
<td>Use of information differs+</td>
<td></td>
</tr>
<tr>
<td>Interim provides actionable information on student. Teacher is also recipient</td>
<td></td>
</tr>
<tr>
<td>*Don’t need as much item metadata</td>
<td></td>
</tr>
<tr>
<td>Could be diagnostic-predictive of summative</td>
<td></td>
</tr>
<tr>
<td>Type of feedback to teacher and parent / stu</td>
<td></td>
</tr>
<tr>
<td>*Local deployment options</td>
<td></td>
</tr>
<tr>
<td>Student data tracked over course of yr</td>
<td></td>
</tr>
<tr>
<td>Interim may focus on a narrow set of assessment targets</td>
<td></td>
</tr>
<tr>
<td>Different tolerance for item exposure</td>
<td></td>
</tr>
<tr>
<td>Test experience is not secure</td>
<td></td>
</tr>
<tr>
<td>Could be different media +</td>
<td></td>
</tr>
<tr>
<td>Alignment to CCSS and alignment options for local standards</td>
<td></td>
</tr>
<tr>
<td>*Support k12 all content areas</td>
<td></td>
</tr>
<tr>
<td>Ability for students to develop portfolios of work</td>
<td></td>
</tr>
<tr>
<td>Undefined (by SBAC) testing windows</td>
<td></td>
</tr>
<tr>
<td>State-defined Proctor authentication+</td>
<td></td>
</tr>
<tr>
<td>Looks forward (predictive) more than backward (reactive)</td>
<td></td>
</tr>
<tr>
<td>*Option for making tests with various item pools, teacher’s own items, or both</td>
<td></td>
</tr>
<tr>
<td>*Differences in frequency</td>
<td></td>
</tr>
<tr>
<td>Initial size of bank is smaller</td>
<td></td>
</tr>
<tr>
<td>*Contains teacher-created items</td>
<td></td>
</tr>
<tr>
<td>SBAC value decisions needed:</td>
<td></td>
</tr>
<tr>
<td>Item source issues, scaling, calibration of items, use of diff tools, Rights of items (DRM), what is being reported</td>
<td></td>
</tr>
</tbody>
</table>
9. Architecture Outputs

As preliminary conversations during the workshops, initial discussions were held around some beginning architecture outputs. These conversations centered on interoperability. These diagrams do not represent the final interoperability and integration recommendations and were simply used for illustrative purposes in determining needs between the various applications.

Interoperability Diagrams

Interoperability Discussion 09/27/2011
10. Priorities

We asked each person to come up with what they felt were their top three most important aspects of the architecture. The intention was to feed this into a sliders exercise at some point in the future. These are purposely anonymous. In preparation for the sliders activity, we asked SBAC members to prioritize a list of items.

From the survey, these results were compiled:
Priority 1
(most important)

Priority 2

Priority 3

Priority 4

Priority 5

Priority 6
(least important)

Number of Responses

Time to market
Feature scope
Interoperability
Deployment and hosting flexibility
Performance, reliability, availability, quality
Innovation
From the results, a discussion was generated to determine the priorities. These below represent the importance of the dimension relative to cost.

- **Time to market**: Cost is relatively less important
- **Feature scope**: Keeping costs is more important
- **Interoperability**: Keeping costs is more important
- **Deployment and hosting flexibility**: Cost is relatively less important
- **Performance, reliability, availability, quality**: Cost is relatively less important
- **Innovation**: Keeping costs is more important