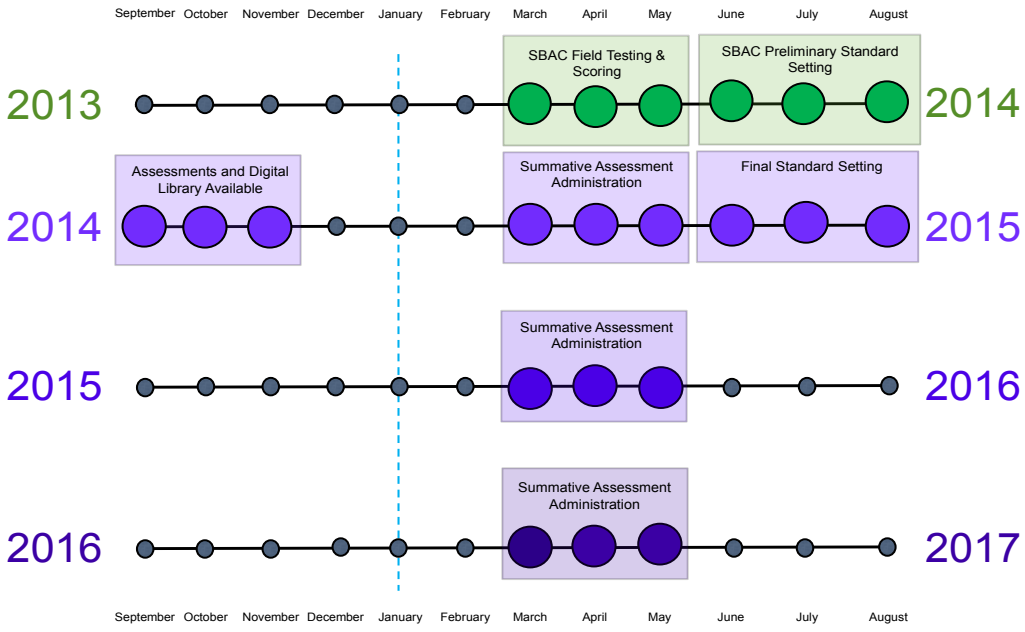
Growth and Assessment Transition:
Evaluating Implications for the Hawaii Educator Effectiveness System

One of the five components in the Hawaii Educator Effectiveness System (EES) is academic growth, calculated using Student Growth Percentiles (SGP). To produce SGPs, multiple test scores (at least two) are required for each student. These test scores are based on summative assessment results in consecutive years.

Like many states, Hawaii is in the process of transitioning to the new Common Core State Standards, which necessitates a corresponding transition for the state assessments. Figure 1 presents the general timeline that all states using the Smarter Balanced Assessment Consortium (SBAC) assessments will face during the transition period.

Figure 1. SBAC States Timeline

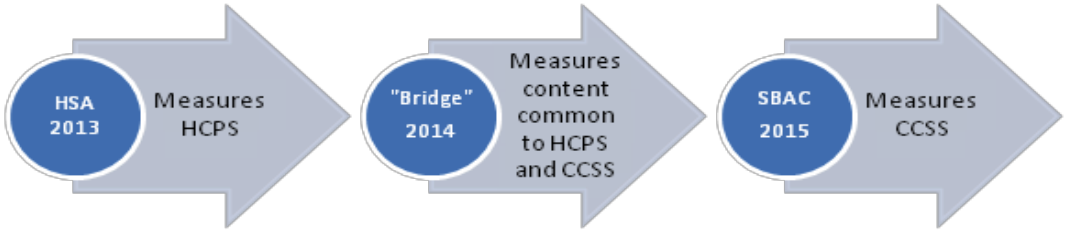


As indicated in the timeline, all SBAC states are currently in the process of field testing the SBAC assessments. For all states, legacy tests or transition assessments are also being administered during this field testing period. In the 2014-2015 school year, it is anticipated that all SBAC states will make the final and complete transition over to the SBAC summative assessments.

The assessment transition in Hawaii is depicted in Figure 2, which shows the progression from Hawaii State Assessments (HSA), which measures the Hawaii Content and Performance Standards (HCPS), to

the 'bridge' assessment, measuring both HCPS and Common Core State Standards (CCSS), and finally to the Smarter Balanced Assessments, which are designed to fully reflect the CCSS.

Figure 2. Assessment Transitions in Hawaii



The central consideration for Hawaii is whether defensible and appropriate measures of academic growth can be obtained for the EES with different tests used as the basis of the SGP calculation. Although the normative properties of the SGPs, which do not require a specific test or scale to produce growth estimates, may support the continued use of SGPs during and after the transition period, inferences about growth and how they support accountability decisions may need to be re-examined during this period. The EES TAG recommends specific analyses to be conducted to assess both the properties of the assessments used in SGP calculations and the resulting growth outcomes. The table located on the following page describes the focal questions to be analyzed, provides examples of the analyses or evidence that can be produced to address these questions, describes the desired outcome to promote confidence in continued use of growth scores during the transition period, and outlines implications to consider if desired outcomes are not reached.

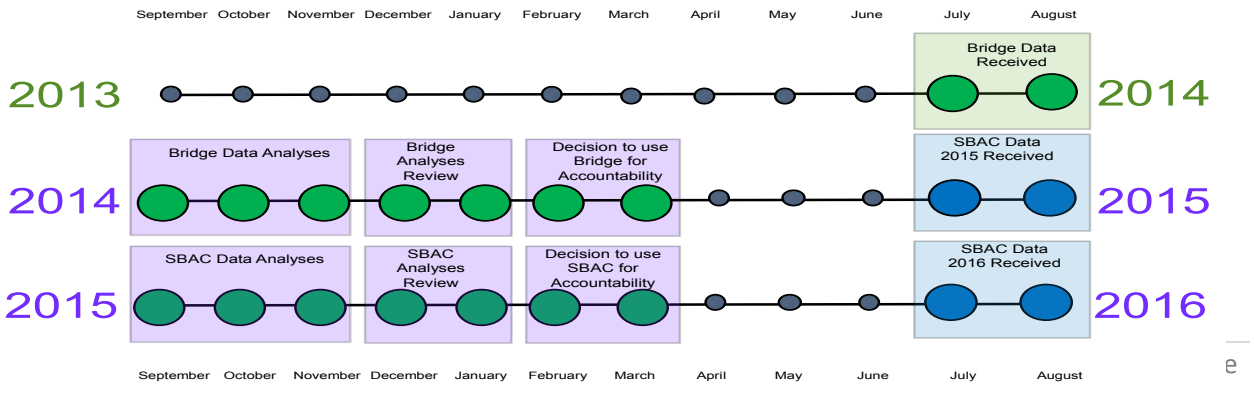
Focal Question	**Evidence/ Analyses**	**Desired Outcome**	**Implications if Desired Outcomes not Achieved**
Are items and tests well-aligned to standards and reflect what teachers should be teaching and students should be learning?	Review of test blueprints Alignment procedures/studies	Alignment and blueprint information for each test reveal that the content standards assessed, distribution of emphasis, and cognitive complexity are adequate to measure the target construct	Significant construct shifts found complicate the communication and interpretation of "growth" results reported to teachers and schools. This finding would be a key factor to weigh if the analysis conducted to answer the next focal question fails to achieve desired outcomes.

What is the relationship between prior year test scores and current year test scores, for different tests during the transition period?	Correlations of test score distributions for HSA to Bridge and Bridge to SBAC	A moderate to strong positive correlation exists in all analyses.	Low correlations found may illuminate differences in how the CCSS have been implemented across schools, and may also signal significant construct shift found across assessments. More investigation required to identify sources contributing to low correlations found between assessments.
Does each assessment used in the growth calculation have an adequately 'high ceiling' and 'low floor' to measure a broad range of performance?	Percent of students obtaining the maximum or minimum score Model fit analyses	Relatively few students (e.g. 5% or less) achieve the maximum or minimum score Model fit is within conventional thresholds and does not substantially change each year	Significant floor or ceiling effects found call into question the accuracy of information used to summarize growth achieved by students in a given classroom. Although adjustments available in the SGP package to reassign SGPs, this does not "fix" the results from the assessments.
Are growth estimates relatively consistent and stable? Are ratings achieved on growth relatively consistent and stable?	Year to year correlations of aggregate growth scores (MGPs) at the school and class level and by school. Compare ratings from 2013-2014 with ratings from the 2014-2015 and 2015-2016 years. Flag schools and teachers with where ratings differ by two categories.	Where n sizes are sufficiently large, correlations are positive and generally consistent for MGPs produced with different tests. It is important to note here that stability will also improve if data are pooled across years and when the margin of error is factored into the ratings. These adjustments (pooling and applying the margin of error) should be made to the data to review their impact on stability.	A high level of instability found in the MGPs and associated classification of teachers and schools on growth threatens the credibility of the data since volatile signals make it difficult to assess improvements achieved over time. May still illuminate important differences in fidelity of CCSS implementation, but this should be evaluated against additional information gathered to substantiate whether stakeholders agree that sites flagged align with other sources confirming uneven implementation issues.

Are growth estimates relatively unrelated to extraneous variables?	Correlation between MGPs and variables such as prior year status scores and poverty.	Where n sizes are sufficiently large, correlations are not greater than moderately positive	High correlations found between extraneous variable and growth may threaten perceptions about the credibility or fairness of the system with regards to uneven resources available across districts or setting a zero-sum evaluation system that reflects largely status-drive outcomes. However, findings may also illuminate important realities and differences in fidelity of CCSS implementation. Additional investigation required to identify reasons for these high correlations to emerge, especially since low correlations with extraneous variables are documented with the HSA.

The confidence in the use of growth scores in the EES during the transition period is more a matter of degree than an 'all or nothing' certainty but needs to be weighed carefully against implications for not achieving desired outcomes with each analysis conducted. Following the release of findings from all analyses conducted by HI DOE, the TAG recommends that HI DOE review any implications and weigh the benefits and costs of using SGPs for the EES as well as for Strive HI during the transition period to inform their final and overall judgment. The TAG can also serve as another external body reviewing results from each analysis conducted to help inform HI DOE's final decision. Figure 3 presents a suggested and tentative timeline for conducting analyses, reviewing results, and making final decisions on using growth results from either the "Bridge"or SBAC assessments for the EES.

Figure 3. Bridge and SBAC Growth Results Review Timeline



This timeline is considered to be tentative since the suggested timeframes and accompanying activities reflected in Figure 3 may shift contingent upon input received from the Systems Accountability Office (SAO) and the HI DOE leadership.