



National Research & Development Center to Improve
EDUCATION FOR SECONDARY ENGLISH LEARNERS

WestEd 

A Vision for Using an Argument-Based Framework for Validity Applied to a Comprehensive System of Assessments for English Learners in Secondary Grades

MARGARET HERITAGE, CAROLINE WYLIE, MOLLY FAULKNER-BOND, & AÍDA WALQUI

Introduction

The purpose of this document is to present a vision for a system of assessments for English Learners in secondary grades that brings assessment closer to the classroom and fully involves teachers in assessment development

and validation. By reimagining a system of assessments, our intention is to signal a new and equitable direction and to provoke reflection and debate among all those concerned with improving outcomes for English Learners.¹

¹ This vision for a CAS will be partially implemented over three years in an iterative study of secondary-level teachers' use of high-quality replacement units (developed in line with the pedagogical approaches described in the background paper). As teachers grapple with the CAS vision in practice, areas that need additional clarifications or need to be trimmed back to be more feasible, or require more support structures to help a Community Of Practitioners (COP) enact the validation process will likely be revealed. The CAS vision will be revised in response to data from this study.

The vision centers on the development and implementation of a learning-centered **comprehensive assessment system** (CAS): a “comprehensive set of means for eliciting evidence of student performance” (NRC, 2001, p. 20) in support of assessment use to enhance the learning of secondary-grade English Learners across content areas. The aim of this CAS

Framework is to ground assessment in the classroom and to create a coherent through line from formative assessment to assessments used for accountability. In this way, enhancing the teaching and learning for English Learners is paramount.

The Problem that the CAS Framework is Aiming to Address

In the United States, over one third (34.7%) of all English Learners are enrolled in the secondary grades (National Center for Education Statistics [NCES], 2018). Two thirds of secondary English Learners (65 percent) have been schooled entirely in the United States (NCES, 2018) and are often labeled “long-term,” a reference to the length of time they have been enrolled in school — more than six years — without meeting their state’s achievement standards to be reclassified out of the English Learner subgroup. Labels can be damaging (e.g., Brooks, 2018; Kibler & Valdés, 2016; Paris, 2012; Rosa, 2019; Umansky & Dumont, 2021) and potentially lead to negative consequences for students. For instance, based on the mistaken belief that English proficiency is a necessary precondition to engage in rigorous grade-level learning, students classified as English Learners often do not have access to the core curriculum and demanding learning opportunities (Callahan & Shifrer, 2016; Glick & Walqui, 2021; Johnson, 2019) and are frequently excluded from grade-level content courses altogether (Umansky, 2016). Assessment data reveal the adverse impact of this domino effect of lost opportunities. For example, data from the 2017 National

Assessment of Educational Progress (NAEP) show a difference of 33 percentage points in reading proficiency between non-English Learners and students currently labeled as English Learners in eighth grade (38 percent non-English Learners versus 5 percent) and a difference of 30 percentage points in mathematics (36 percent non-English Learners versus 6 percent EL) (NCES, 2018). As a consequence, their performance on other indicators such as ACT participation and postsecondary enrollment is adversely impacted (Carlson & Knowles, 2016). The magnitude of this achievement lag is untenable from the perspective of educational equity.

Educational Equity

According to the Organisation for Economic Co-operation and Development (OECD), educational equity has two dimensions: fairness and inclusion (OECD, 2012). **Fairness** means ensuring that personal and social circumstances — for instance, gender, socioeconomic status, and language status — are not obstacles to educational achievement. **Inclusion** means setting a basic minimum standard for education

that is shared by all students, irrespective of background or personal characteristics. In this regard, teachers and administrators are responsible for ensuring that all students in the United States meet the achievement standards that have been adopted by each state and have equitable learning opportunities in order to reach them.

To this end, it is essential that English Learners have access to, and engagement with, challenging and rigorous content learning that is required to meet state standards and that teaching and learning support both high levels of content and analytic learning and the development of language resources needed to learn that content. Realizing equity requires

understanding each student's needs and designing learning experiences that will help all English Learners — and all means each one — to achieve success. Assessment must play its part in providing information that will support educators to engage in ongoing practices that are likely to lead to positive outcomes for every English Learner. Since the CAS Framework has individual variations built into the design and interpretation, it is intended for use with all English Learners in all settings.

In the next section, the perspective on language development that underpins the CAS Framework's approach to assessment system design and evaluation is discussed.

Perspective on Language Development²

English Learners need opportunities in the classroom to develop situated language competencies during interactions with peers and teachers while simultaneously developing discipline-specific practices (Valdés, Kibler, & Walqui, 2014). This means that learning disciplinary concepts and analytical practices is not distinct from the linguistic means through which the understanding is developed and expressed; the demands of understanding concepts, practices, and relationships are not privileged above the demands of linguistic resources, nor vice versa. Building with their existing language resources English Learners develop and use new language resources as they make meaning of content (Walqui & Heritage, 2011).

This perspective on language development has its roots in Vygotsky's theory of the relationship between language and thought (e.g., 1978). Vygotsky maintained that thought is not merely expressed in words, it comes "into existence through them" (Vygotsky, 1986, p. 218). In this vein, he argued that the development and functioning of higher mental processes (cognition) are mediated and that language is one of the most important mediating tools that humans have at their disposal (Swain, 2006). Language as a mediating tool is used in interaction with others and with oneself (through inner speech) and results in the creation and use of higher mental processes (van Lier, 2004; Swain & Lapkin, 2011).

² For a more detailed description of this perspective, see Heritage, Faulkner-Bond, & Walqui, 2021.

The perspective on language development incorporates three additional concepts that are hallmarks of classroom practice for English Learners: apprenticeship, the ZPD, and scaffolding.

Apprenticeship operationalizes Vygotsky's emphasis on the interrelated roles of the individual and the social world. It refers to the process through which the individual becomes part of the group and develops their ways of doing things. Consequently, apprenticeship can only occur in community activity and involves active individuals participating with others in "culturally organized activity that has as part of its purpose the development of mature participation in the activity by the less experienced people" (Rogoff, 1995, p.143). English Learners are apprenticed into the language and make sense of disciplinary concepts and analytical practices.

The ZPD, a concept that also originates with Vygotsky, is defined as the distance between what the individual can accomplish during independent problem-solving and the level of problem-solving that can be achieved with the assistance of adult or in collaboration with a more expert peer (Vygotsky, 1978). In his

discussion of the importance of the ZPD for education, Vygotsky (1978) identified learners' emerging abilities as the appropriate target for instructional efforts to guide development (Levi & Poehner, 2018). English Learners' emergent content understanding and practices, including language are both targets for instruction within the students' ZPD. Creating contexts for academic learning in English in the ZPD occurs in part through the scaffolding of social interaction (Walqui, 2006).

Scaffolding is a process of "setting up" the situation to make the child's entry easy and successful, then gradually pulling back and handing the role to the child as he becomes skilled enough to manage it. (Bruner, 1983, p. 60). From this definition, we can understand that scaffolding has a more or less constant ritual structure (though flexible) and an interactional process that is jointly constructed from moment to moment and which occurs in the student's ZPD (Walqui, 2006).

The clear implication for the CAS Framework from this perspective on language is that language and content learning are not treated as separate entities; they develop together and therefore, should be assessed together.

Current Assessment System

Currently, the assessment experiences of English Learners and indeed, most, if not all K-12 students in the United States, are dominated by large-scale, year-end assessments (Figure 1) (e.g., Gordon, 2020; Volante et al.,

2020). Figure 1 shows the current system with the dominance of large-scale standardized assessments, which both overwhelm and press on the design of classroom-based assessments.

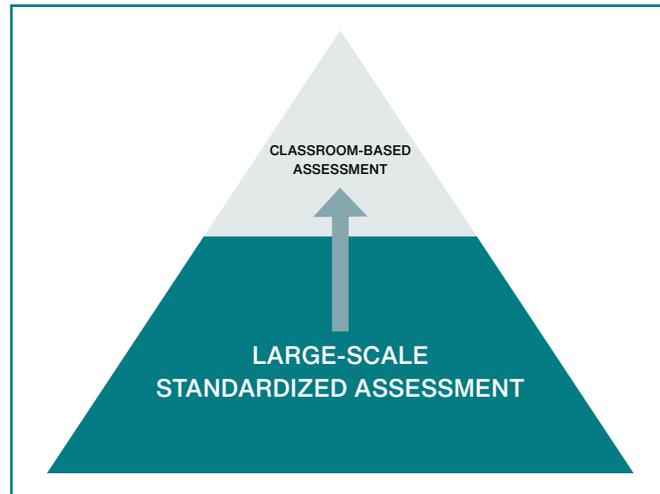
Large-scale standardized assessments are designed to support accountability reporting and decision-making, but in practice are used for a variety of other purposes, including placement, accountability, classification and reclassification (Umansky & Porter, 2020).

While the current system does include other forms of assessment such as benchmark assessments, many of them are developed or selected primarily for their perceived relevance to large-scale assessments (Volante et al., 2020). Specifically, student scores on the year-end summative assessment are often treated as the most accurate and meaningful indicator of a student’s achievement and information collected throughout the year and often designed to predict this score and gauge student progress on this assessment, rather than to collect substantive, actionable information about student learning.

Support for Classroom Learning

Large-scale assessments reveal little about students’ responses in the context of classroom learning and have limited utility for the purpose of supporting language and content learning (Bailey & Durán, 2020). The validity of any assessment is prejudiced if it reinforces approaches to teaching which are inappropriate for the specified educational goals (Black, 1993). In this regard, the validity of current standardized assessments is prejudiced on the grounds that they can reinforce teaching practices that isolate language from content and analysis, which often means segregating English

Figure 1. Current Assessment System



Learners from their non-English Learner peers, thereby removing important contextual factors that are critical to students’ development of content knowledge.

English Learners acquire additional language and content simultaneously by responding to “affordances” emerging from dynamic communicative situations (van Lier, 2000, 2004; van Lier & Walqui, 2012). For this reason, and in light of the perspective on language development described above, assessment of English Learners needs to reflect the nature of the learning context and students’ experience in learning content and language simultaneously. A validity concern in assessment for English Learners, who are both a linguistically and culturally heterogeneous group, is their interpretation of the assessment items that are potentially insensitive to the students’ backgrounds (e.g., Solano-Flores, 2006). Both the social and cultural nature of learning suggest the need for new ways to assess English Learners beyond traditional means such as standardized assessments (Durán, 2008).

Inverting the Assessment System

Concerns have been raised about the top-down nature of assessment in the United States, with calls for a more bottom-up approach that places greater and prioritized emphasis on assessment for the purpose of informing and improving learning and the teaching processes that enable learning (Bailey & Durán, 2020; Gordon, Gordon, Aber, & Berliner, 2013; Wilson, 2018³).

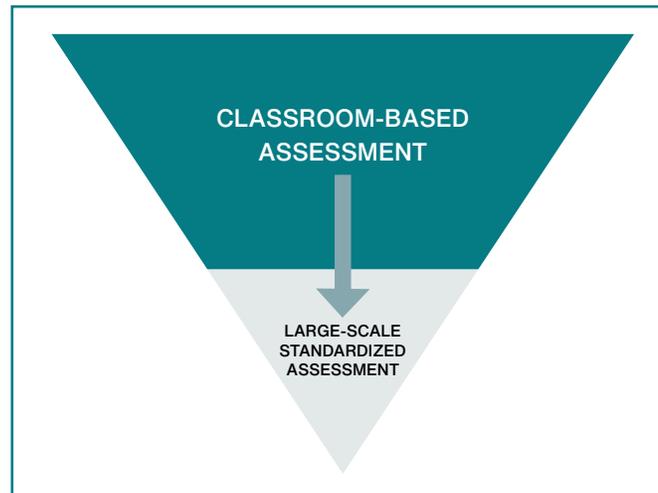
The CAS Framework adopts the novel approach of inverting the current assessment system (Figure 2) to privilege assessments at the classroom

level that inform ongoing teaching and learning for English Learners as the drivers of assessment types. Its purpose is essentially to make the large-scale summative score less of a culmination and more of a confirmation of information that is already known.

This inversion represents an approach to assessment that addresses the current problem of limited access to rigorous content for English Learners and the consequential impact of their poor performance on large-scale assessments. In Figure 2, classroom assessment is the driver of large-scale standardized assessment with the potential of ultimately enabling improved synergy between them and classroom-based assessments (cf. Wilson, 2018).

The rationale for this inversion is articulated in the logic model in Figure 3. Within this logic

Figure 2. Inverted Assessment System

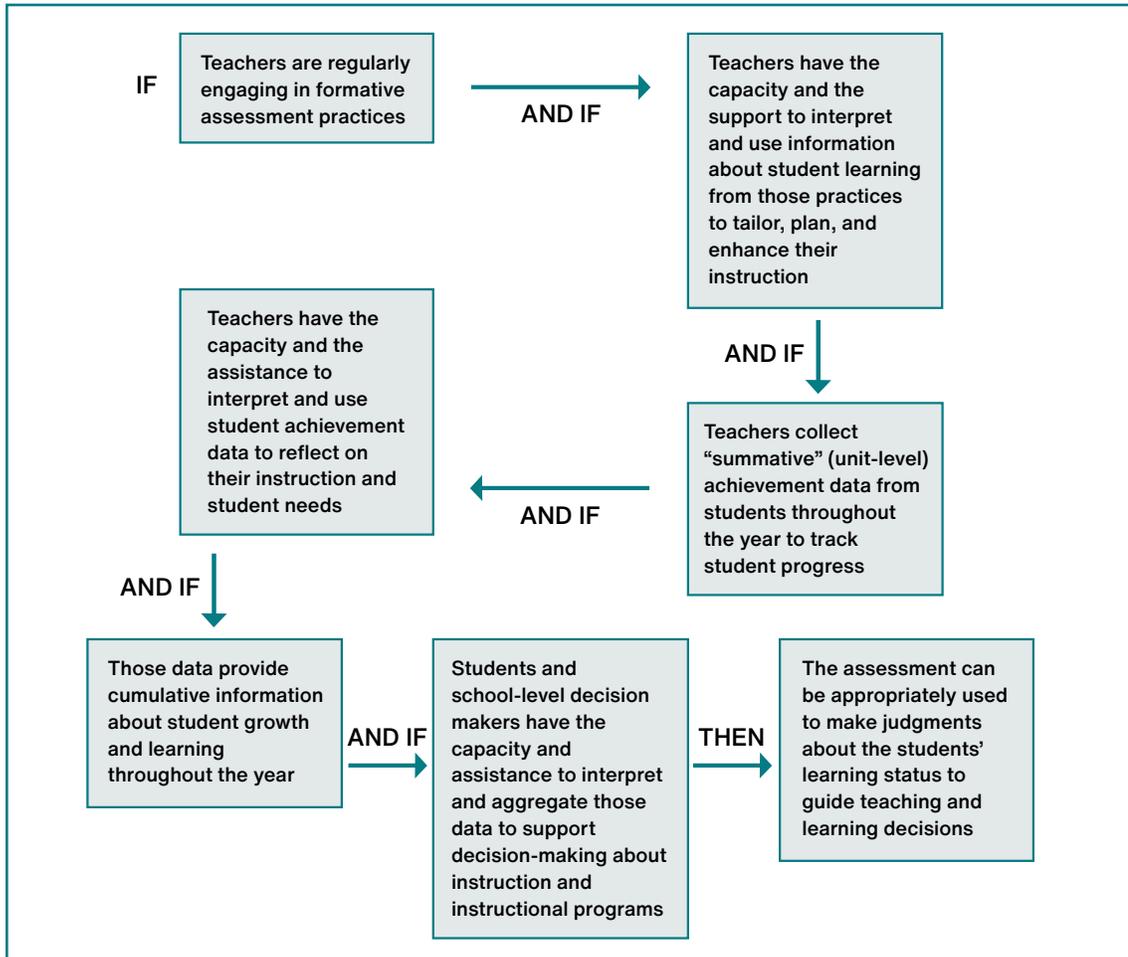


model, there is congruence across the assessments in terms of the focus on simultaneous academic content and language development and in reinforcing approaches to teaching which are appropriate for the specified educational goals for English Learners.

Such a system would require significant investments to ensure that teachers have the capacity, support, and assistance to engage in the instructional and assessment practices described. While it is beyond the scope of this paper to discuss how such changes might be accomplished, the cost savings from using fewer formal assessments offer an opportunity to redirect and reprioritize spending towards investments in the human capital of teachers instead.

³ Mark Wilson's Presidential Address at the National Council of Measurement in Education [NCME], 2017, later published as Wilson, 2018. It is noteworthy that subsequent to this address, a subgroup of the NCME, The Classroom Assessment Task Force, was established to promote classroom assessment in rebalancing efforts.

Figure 3: Logic Model for Inverted Assessment System



The CAS Framework

The CAS Framework is intended to reflect the nature of the learning contexts and the students’ backgrounds in assessment to support equal opportunities to learn and to achieve for secondary English Learners. The CAS Framework is based on the following assessment principles:

1. Focuses on the learner and learning:

Assessment provides insights into each student’s thinking, skills, and language

development. Assessment is aligned to high-quality classroom learning, consistent with Table 1, provides all students with the opportunity to show where they are in their learning through multiple modalities, and reflects meaningful, worthwhile tasks that challenge the upper reaches of students’ language competence and conceptual understanding.

2. Emphasizes rigorous learning: Assessment focuses on the concepts, knowledge, language, and analytical practices inherent in academic content standards. Assessment reflects high-quality classroom learning experiences characterized by apprenticeship, interaction, and scaffolding within the students' zone of proximal development (ZPD) that promote deep, interconnected understandings and the language to express them.

3. Produces actionable information for the user: Information is asset- and future-oriented, focusing on what students can do in terms of content and language as well as highlighting areas of need and potential growth. Information yielded is tractable for teaching and learning.

4. Supports metacognition and self-regulation: Assessment provides information that supports the ongoing development of students' metacognitive thinking about their learning (both their thinking processes and their language use), their achievement, and their approaches to learning, which in turn enables them to proactively orient their actions to achieving goals.

5. Promotes self-efficacy and learner identity: Assessment is designed with multiple entry points so that all students are able to show what they know and what they can do with language, giving students a sense of accomplishment and helping them enhance their feelings of self-efficacy and build positive learner identities.

Overview of Framework

Just as a one-size-fits all pedagogy does not meet the learning needs of every English Learner, no single assessment can accomplish all assessment purposes for English Learners. The CAS comprises a range of assessments (formative assessment, end-of-unit classroom assessment, end-of-year classroom assessment, and student portfolios). The CAS's aim is to provide decision-makers with the information they need to support English Learners in the secondary grades to engage in challenging and rigorous content learning required to meet state standards. Table 1 shows the potential users, purposes, and proposed methods of assessment within the CAS Framework.

Table 1. Assessments and Their Uses in a CAS⁴

TYPE	USER	PURPOSE	METHOD	INFORMATION
Formative Assessment-designed into ongoing teaching and learning	Teachers and students	Inform ongoing learning and teaching	<ul style="list-style-type: none"> • Observation of classroom discourse/ students engaged in instructional tasks • Analysis of student work • Student self-assessment/peer assessment • Metacognitive monitoring relative to goals 	<ul style="list-style-type: none"> • Emerging, partially formed language • Understanding of academic content • Current learning status relative to lesson goals • Any difficulties, misconceptions
Classroom summative at the end of a learning episode or thematic series of lessons	Teachers, students, parents, school-level administrators	Evaluate learning relative to medium-term goals	<ul style="list-style-type: none"> • Student work products and performances (e.g., portfolio), with associated rubric(s) • Student self-assessment and evaluative reflection • Classroom summative assessments (e.g., unit assessments, interim assessments that cover one or more units) designed/selected by teacher(s) 	<ul style="list-style-type: none"> • Cumulative record of learning • Status of student learning relative to medium-term goals (e.g., unit) • Student strengths and needs
Classroom summative at the end of the course, semester or year	Teachers, students, parents, school and district administrators	Evaluate learning relative to long-term goals	<ul style="list-style-type: none"> • Student work products and performances (e.g., portfolio), with associated rubric(s) • Student self-assessment and evaluative reflection • Classroom summative assessments designed/ selected by teacher(s) 	<ul style="list-style-type: none"> • Cumulative record of learning • Status of student learning relative to longer-term goals • Student strengths and needs
External summative at the end of the year	Teachers, students, parents, school, district, and state administrators	Federal accountability, evaluate programs, inform systemic planning	<ul style="list-style-type: none"> • District-administered standardized assessment • Statewide standardized assessment 	<ul style="list-style-type: none"> • Achievement relative to end-of-year state standards

⁴ Since the CAS represents a reimagining of what is possible, not a reordering of what currently is, there are types of assessment (e.g., district-administered, off-the-shelf assessments) that are not explicitly included in the framework. In the reimagined system, such instruments would not be necessary.

Three Properties of the System

To optimize the credibility and utility of the resulting information, an assessment system should exhibit three properties: comprehensiveness, coherence, and continuity (National Research Council [NRC], 2001).

Comprehensiveness means that a range of approaches are used to provide a variety of evidence to support educational decision-making (NRC, 2001). **Coherence** relative to the framework has two dimensions: (1) the alignment among learning goals, instruction, and assessment so that all three are moving in the same direction (horizontal coherence), and (2) the extent to which learning goals, instruction, and assessment are continually intertwined over time to promote student progress (developmental coherence) (Herman, 2010). **Continuity** refers to the need for the system to assess student progress over time, “akin more to a video recording than to snapshots provided by a system of on-demand tests” (NRC, 2001, p. 257).

The CAS Framework is designed to reflect these three properties. For example, the framework supports comprehensiveness by including assessments with different grain-sizes and purposes, by emphasizing horizontal coherence among learning goals, instruction, and assessment approaches, along with the role of learning trajectories to support vertical coherence, and by attending to continuity with the inclusion of the collection and review of evidence in a student portfolio.

Soundness of Assessment Information

There are two important concepts that concern the soundness of the information from any assessment for decision-making: validity and reliability.

Validity

Does the assessment information lead to sound interpretations, decisions, or actions that enhance learning for secondary-grade English Learners? (Moss, Girard, & Haniford, 2006.). The answer to this question depends on the validity of the inferences made from test scores for a given purpose. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) refers to validity as the “degree to which evidence and theory support the interpretation of test scores for proposed uses of tests” (p. 11), and validation as a “process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed uses” (p. 11). Significant efforts are generally expended to collect validity evidence for assessments that are used to inform high-stakes decisions such as certification or placement decisions. Certainly, assessments that have significant consequences for students should have a high degree of validity (i.e., they require more evidence) to support their use. However, as we shall see in the CAS Framework, validity is important for all assessment uses, including the use of questions teachers ask during formative assessment.

Validity is not a property of the assessment but of the inferences made from assessment results and the extent to which they justify the use of an assessment for a specific purpose. For instance, scores from a mathematics assessment may have a strong degree of validity for mathematics problem solving, but a weak degree of validity for students' computational fluency. In formative assessment, for example, a concern about validity would lead a teacher to explore whether the questions, tasks, and activities that she uses to elicit evidence of student thinking are

aligned with the learning goals of the lesson, and whether they generate student responses that provide insights into their conceptual understanding, their use of analytical practices, and their language development.

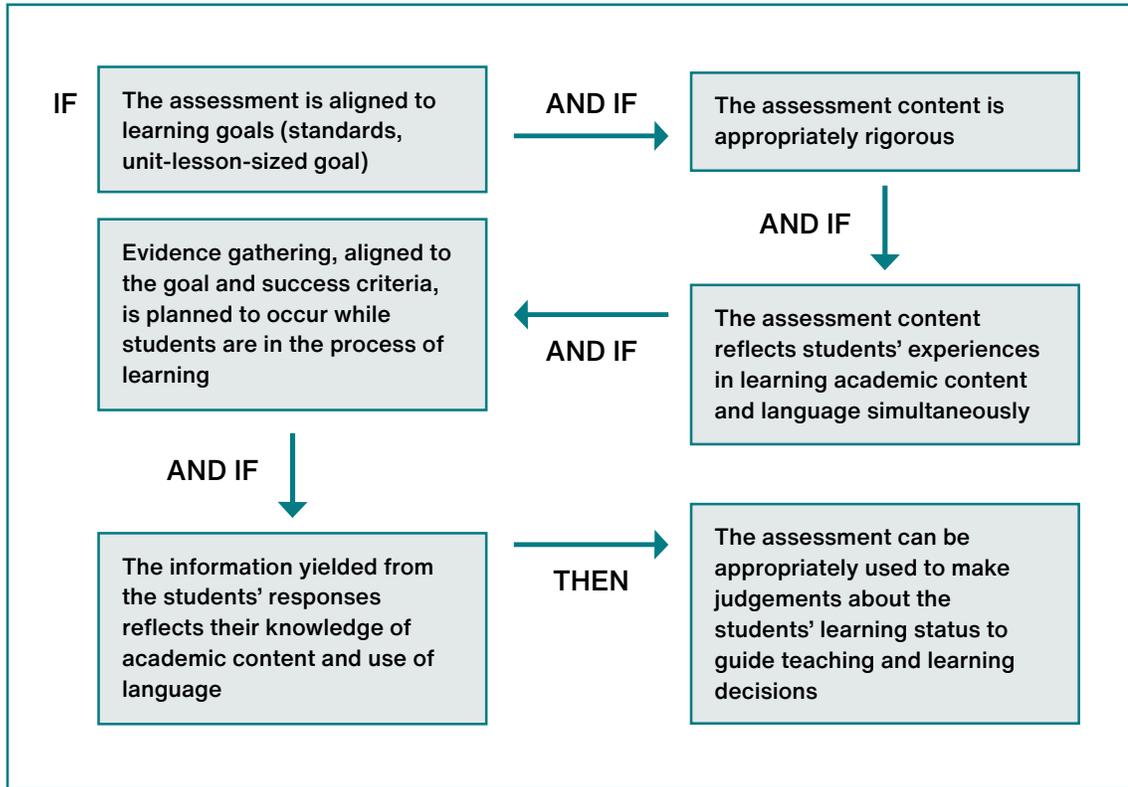
Reliability, a necessary but not sufficient condition for validity, refers to the consistency of assessment results across settings, students, and users. Appendix A provides a discussion of reliability in relation to the assessments in the CAS Framework.

An Argument-Based Approach to Validity

The validation approach for the assessments in the CAS Framework draws on Kane's (2006, 2013) structure for an argument-based approach to validity in order to delineate the evidence that needs to be accumulated and evaluated for valid interpretations and uses of the information yielded by the assessments. The CAS Framework provides users with guidance about the evidence on which the soundness of interpretations, decisions, and actions to enhance learning for secondary-grade English Learners can be judged. While this validity approach has relevance for classroom assessment in general, because of its emphasis on the assessment of integrated content and language development, it has particular application for English Learners.

An argument-based approach to validity involves two arguments: 1) an **interpretation and use argument (IUA)** and 2) the **validity argument**. An IUA specifies the proposed interpretations and uses of assessment results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances (Kane, 2013). It is useful to think of an IUA in terms of a logic model or if/then propositions that articulate the means for reaching the intended purpose (Perie & Forte, 2011). Figure 4 shows an example of a logic model for a classroom assessment of academic content knowledge and language.

Figure 4: Logic Model for Classroom Assessment of Academic Content Knowledge and Language



Building the Arguments

The first step in an argument-based approach is to specify the **propositions** of the IUA. A proposition might be, for example, “the assessment is aligned to learning goals.” The next step is to establish the claims that support each proposition. The **claims** are the fundamental criteria for appraising the extent to which each proposition is supported with specific evidence (Herman et al., 2011). The combination of the logic model,

propositions, and claims form the IUA. Evidence is marshalled for each of the claims in the IUA. The evaluation of this evidence then forms the **validity argument**. In practice, as it is often not feasible to collect evidence for every claim at once (or at all), users should prioritize the accumulation of evidence for those claims in the IUA that are most ambitious or consequential (Kane, 2013), for example, portfolio scores to determine course placement in the next school year.

Proposition: The argument for the use of assessment for specific purposes comprises a series of propositions or hypothetical statements that link the performance on the assessment to specific interpretations of the meaning of the information yielded and the conclusions or decisions made on the basis of performance.

Claim: The fundamental criterion for appraising the extent to which each proposition is supported and needs to be substantiated with specific evidence.

Using a Community of Practice Approach to Evaluate the Validity Argument for Classroom-Based Assessment

For high-stakes assessment uses such as accountability, graduation, or certification validity, evidence is often gathered through a series of rigorous studies that may involve convening large groups of subject matter experts, or by conducting complex statistical analyses. The outcomes of these efforts are typically written into a technical report, which may be posted publicly or shared with governing bodies such as the federal government or a certifying organization. Most testing programs also retain a committee of technical advisors who review their validation plans and results to provide feedback throughout the process.

However, because the CAS Framework centers classroom-based assessment at its core, a different approach to the collection and evaluation of validity evidence is needed. Rather than being undertaken by external subject matter experts or psychometricians, we propose that an evaluation of validity evidence for classroom-based assessment is conducted in a teachers' **community of practice** (COP).

The accumulation and evaluation of validity evidence is an iterative and **educative** process. Repeated cycles of assessment review provide opportunities for teachers to increase their assessment literacy; to deepen their knowledge of simultaneous language and content development and their understanding of quality in the context of classroom assessment; to improve

their analysis, interpretation, and application of assessment information to support student learning; to make improvements to an assessment question, task, or activity for future use; and to enhance the quality of newly developed assessment questions, tasks, or activities.

Community of Practice refers to a group of teachers within a school who meet regularly to tackle a particular problem of practice and find solutions.

Educative refers to the teacher learning that arises from engaging in with peers in discussions about classroom assessment, standards, and pedagogy. Educative refers to the teacher learning that arises from engaging in with peers in discussions about classroom assessment, standards, and pedagogy.

This increased role is not intended to place undue burden on teachers who will need time and support to engage in this work. Teachers are not expected to address all their assessments simultaneously. Rather, the framework lays out a roadmap for changing assessments in the system to maximally benefit English Learners' development. Implementing the framework should be thought of as a long-term undertaking. The COP envisioned here might replace (but function similarly to) current practices such as data discussions or meetings in which teacher teams meet to review assessment data and plan instruction.

Sources of Evidence for Validity Evaluation

A variety of evidence sources is needed for the validity evaluation in the COP to determine the strength of the evidence in support of an assessment's use.

Evidence in the CAS Framework falls in one of six categories, generally:

1. Documentation related to assessment development and/or administration of the assessment (e.g., learning goals presented

to students, directions for portfolio selection)

2. Individual teacher reflection on specific aspects of the assessment (e.g., teacher reflection on whether the questions, tasks, and activities are accessible to the range of students' zone of proximal development (ZPD) present within the class)
3. Peer observation of assessment processes or supporting processes (e.g., how effectively teachers communicate learning goals to students, observation and analysis of student-to-student or student-to-teacher interactions)
4. Student feedback about assessment (e.g., survey or interviews)
5. Peer feedback on an assessment claim (e.g., peer review and feedback on the alignment between the breadth and depth of cognitive complexity and language usage represented by the unit goals and the evidence selected by students)

In the sections that follow, the types of supporting evidence are described more specifically in the context of the related claims.

Propositions, Claims, and Evidence for the Assessments in the CAS Framework

This section includes a logic model for each assessment in the CAS Framework and its propositions, claims, and supporting evidence needed for the validity evaluation.

Formative Assessment⁵

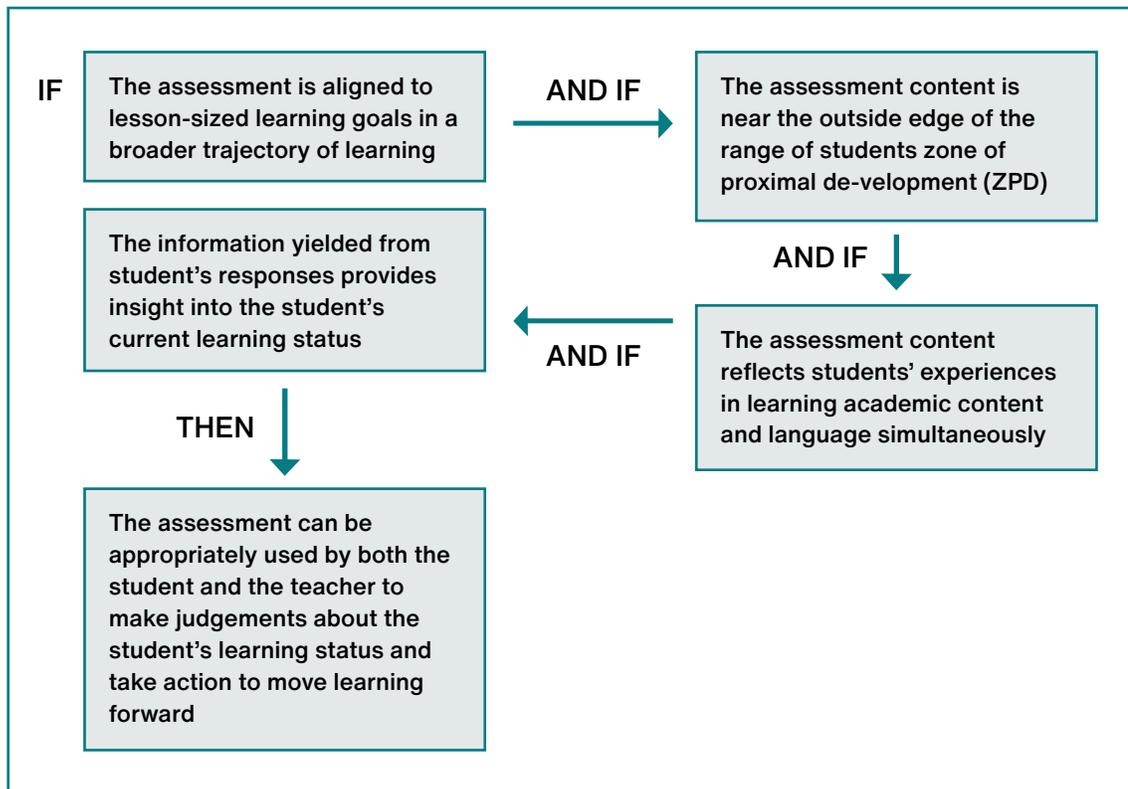
The hallmark of formative assessment that is designed into teaching is to ascertain the current learning status of individual students

⁵ We are not proposing that every question, probe, task, and activity that a teacher uses in formative assessment go through the validation process. However, undertaking the process periodically with a selected question, task, or activity can lead to important learning for teachers about design, integration of academic content and language, providing feedback, or identifying next steps in learning. This learning can be applied to subsequent questions, tasks, or activities.

during their learning, and take action to advance each student's learning toward meeting lesson goals. This does not mean that a teacher will consistently engage in one-on-one instruction, which is neither practical nor desirable. Rather, the teacher will provide multiple points of entry to questions, tasks, and activities that enable individual students to show where they are in

their learning relative to content and language lesson goals. With the evidence obtained from formative assessment, the teacher can engage in individual, small group, or whole class instruction and provide opportunities for peer-to-peer learning to advance each student's learning toward meeting lesson goals. The logic model for formative assessment is shown in Figure 5.

Figure 5. Logic Model for Formative Assessment



Proposition 1: The assessment is aligned to lesson-sized learning goals in a broader trajectory of learning that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them.

CLAIMS	EVIDENCE
<p>The teacher (s) has clearly identified goals from clusters of standards that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them</p>	<ul style="list-style-type: none"> ▪ Teacher lesson plans articulate goals and success criteria derived from the standards along with questions, tasks or activities selected or developed
<p>The teacher has a clear understanding of the intended trajectory of learning in both language and content for individual students</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation of the alignment between goals/success criteria and questions, tasks or activities, the position of the goals in a trajectory of learning (i.e., do the goals build on prior learning and can they be extended to new learning?)
<p>The teacher has designed assessment tasks that reflect all of the above and the teacher and students have a clear understanding of what meeting the learning goal(s) entails in terms of performance on the assessment (e.g., what students will say, do, make, or write)</p>	<ul style="list-style-type: none"> ▪ Peer classroom observation on aspects requested by observed teacher (e.g., alignment of goals/criteria with formative assessment opportunities, clarity of performance criteria, communication of goals and criteria to students to ensure common understanding, breadth of cognitive complexity represented by goals and assessment questions, tasks, activities)
<p>The assessment questions, tasks, or activities assess the breadth of cognitive complexity represented by the lesson goals</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation of the breadth of cognitive complexity represented by goals and assessment questions, tasks, activities

Proposition 2: The assessment content is near the outside edge of the students' ZPD.

CLAIMS	EVIDENCE
<p>The teacher is working with an understanding of the range of students' prior learning based on prior evidence and current ZPD</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on whether the questions, tasks, and activities are accessible to the range of students' ZPDs across the class
<p>The teacher is able to ask questions and provide integrated tasks or activities for assessment purposes that are matched to the range of students' ZPD</p>	<ul style="list-style-type: none"> ▪ Teacher reflection on whether action taken in response to evidence moved student learning forward (i.e., the information provided insights into the students' ZPD and the teacher was able to match an instructional response accordingly)
<p>The teacher provides students with multiple points of entry to questions, tasks, and activities for assessment purposes so that students have opportunities to reveal where they are relative to the learning goal</p>	<ul style="list-style-type: none"> ▪ Peer classroom observation on aspects requested by observed teacher (e.g., whether the questions, tasks, and activities are accessible to the range of students' ZPD, whether action taken in response to evidence moved student learning forward)

Proposition 3: The assessment content reflects the students’ experience in learning content and language simultaneously and is accessible to all students.

CLAIMS	EVIDENCE
The teacher has knowledge of the student’s emerging conceptual understanding, application of analytical practices, and language development	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on knowledge of students’ emerging conceptual understanding, application of analytical practices, and language development
The teacher has knowledge of the student’s funds of knowledge from home and community	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on knowledge of the students’ funds of knowledge from home and community
The teacher integrates knowledge of the student’s language development, conceptual understanding, analytical practices, and funds of knowledge into the assessment	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on the integration of knowledge about the students’ language development, conceptual understanding, analytical practices, and funds of knowledge into the assessment ▪ Peer classroom observation on the degree of integration of knowledge about the students’ language development, conceptual understanding, analytical practices, and funds of knowledge into the assessment
The assessment questions, tasks, or activities are accessible and fair for all English Learners (e.g., provide multiple means of engagement, representation, and expression; do not contain images or references that would be unfamiliar, confusing, or offensive to some students)	<ul style="list-style-type: none"> ▪ Peer review of accessibility and fairness of assessment questions, tasks, or activities for English Learners
The assessment is consistent with the pedagogical approach in the classroom (see Heritage, Faulkner-Bond, & Walqui, 2021)	<ul style="list-style-type: none"> ▪ Peer review on the alignment between classroom pedagogical approaches and the assessment approaches
The assessment offers the potential to learn something new, consolidate understanding, think from a new perspective, or reconsider one’s own thinking	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on the potential of the assessment to learn something new, consolidate understanding, think from a new perspective, or reconsider one’s own thinking ▪ Peer classroom observation on the degree to which assessment offers the potential to learn something new, consolidate understanding, think from a new perspective, or reconsider one’s own thinking

Proposition 4: The information yielded from the student’s responses provides qualitative insight into the student’s current learning status.

CLAIMS	EVIDENCE
<p>Students have the opportunity to provide evidence of their emerging learning in a variety of modalities (e.g., oral, written, drawings, video) and, if relevant, languages</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on opportunities provided for a variety of modalities and degree to which learning was revealed ▪ Peer classroom observation on variety of assessment modalities provided
<p>The assessment provides sufficient scaffolding and/or signals where students can scaffold for themselves or ask for support so that students are able to reveal their current learning status (e.g., think-pair-share, language/writing frames, graphic organizers, discussion-bookmark)</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on whether sufficient scaffolding was provided and degree to which it supported students to reveal current learning status or if more or less scaffolding is required ▪ Peer classroom observation on use of scaffolding
<p>The teacher has sufficient understanding of content and language to anticipate a range of student responses in order to probe further and to deepen insights into student learning</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on ability to anticipate range of responses and use that to probe more deeply ▪ Peer classroom observation on use of probes
<p>Across multiple students, the teacher is able to recognize similarities in responses in order to inform future next learning steps</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on ability to recognize similarities in responses ▪ Peer classroom observation on responses to students

Proposition 5: The assessment can be appropriately used by both the students and the teacher to make judgments about the students' learning status and take action to move learning forward.

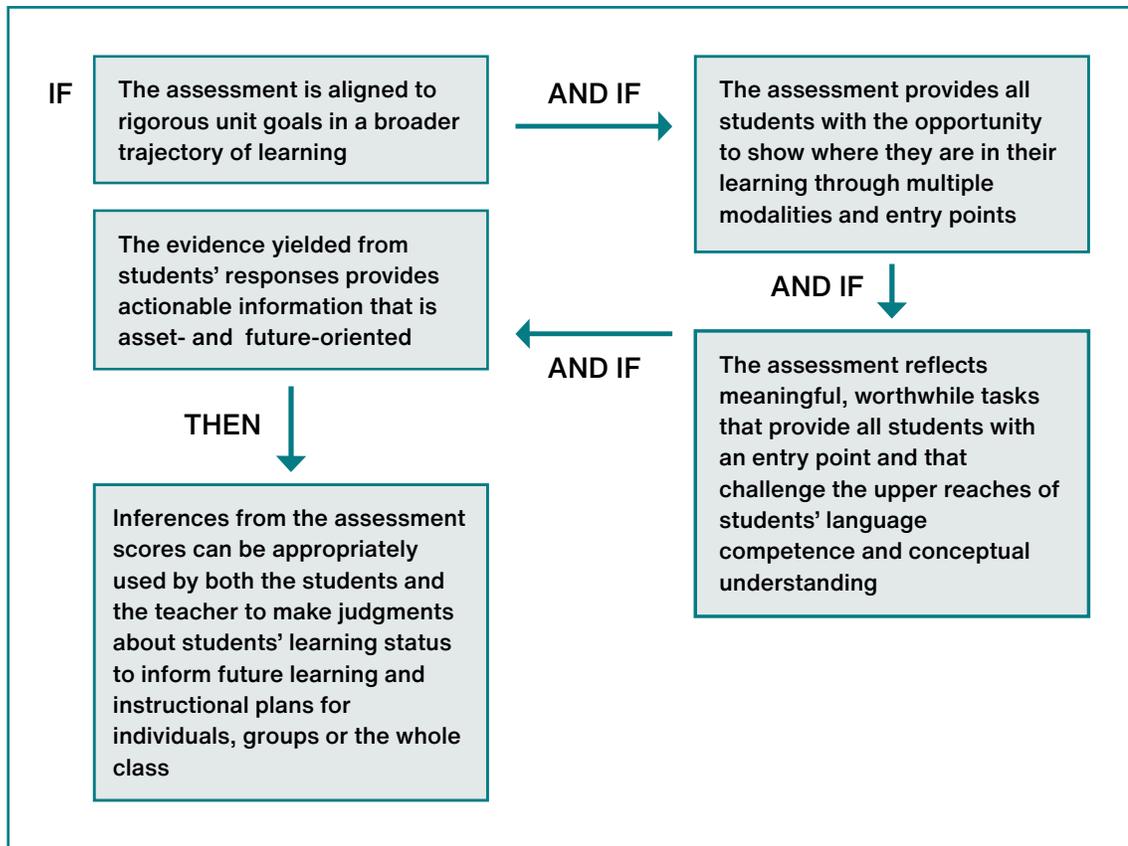
CLAIMS	EVIDENCE
The qualitative insights are sufficient to inform a judgment about next steps in learning (e.g., responses to multiple questions, analysis of student work, observations of student activity)	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on sufficiency of insights (see Appendix A)
The level of detail is adequate to make a decision about next steps to advance learning	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on level of detail to take action to advance learning and the ability of students to use feedback (and the quality of feedback for this purpose)
The student is able to engage in metacognitive activity about his/her learning and respond to external feedback or take action as a result of their own internal feedback	<ul style="list-style-type: none"> ▪ Student interview or feedback. Protocol for interview/feedback. For example, 5-point Likert scale for feedback: I understand what is expected of me in my learning; the feedback from my teacher/peers is helpful; I use feedback to improve my own learning; I am able to monitor and assess my own learning while I am learning; I feel good about my learning/progress/myself as a student. Interview questions around same topics ▪ Peer observation of how the teacher provides opportunities and supports for students to engage in metacognitive activity about their learning and respond to external feedback, or take action as a result of their own internal feedback from self-assessment
The students better understand the learning goals and performance expectations even if they are not yet fully achieved	<ul style="list-style-type: none"> ▪ Students' learning progresses in terms of language, conceptual understanding and analytical practices (e.g., subsequent performance on classroom summative assessments) ▪ Student interview or feedback on understanding of learning goals and performance expectations
The students feel positive about their own ability, identity, and/or trajectory	<ul style="list-style-type: none"> ▪ Student interview or feedback on feelings about their own ability, identity, and/or trajectory

Classroom Summative Assessment (thematic series of lessons or unit)

The purpose of classroom summative assessment for English Learners is to ascertain the students’ learning status at the end of a thematic series of lessons or a unit. It answers the question “what have the students learned to date?” With the information gained from such classroom summative assessment, the teacher is able to update her understanding of students’ learning relative to her medium-term goals

(e.g., unit). With an understanding of students’ strengths and needs in academic content and language usage, the teacher may adjust future plans to clear up persistent misconceptions and determine opportunities to revisit content or language that students may need further support within subsequent units. The logic model for classroom summative assessment (thematic series of lessons or unit) is shown in Figure 6.

Figure 6. Logic Model for Classroom Summative Assessment (thematic series of lessons or unit)



Proposition 1: The assessment is aligned to goals from state standards that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them.

CLAIMS	EVIDENCE
<p>The teacher has clearly identified goals from <u>clusters of state standards</u> that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them</p>	<ul style="list-style-type: none"> ▪ Documentation of teacher reflection and evaluation of the alignment between the assessment content and the unit learning goals and success criteria ▪ Documentation (e.g., assessment blueprint) of the alignment between the goals and clusters of state standards or unit goals
<p>The teacher and student have a clear understanding of what meeting the unit learning goal(s) entail in terms of performance (e.g., oral, written, viewing, drawing, and other representations)</p>	<ul style="list-style-type: none"> ▪ Documentation of unit level criteria for success (performance indicators) ▪ Student feedback on their understanding of what meeting the unit learning goal(s) entail in terms of performance
<p>The full, or prioritized set of unit learning goals are assessed by the assessment(s) The assessment questions, tasks, or activities assess the breadth and depth of cognitive complexity and language usage represented by the unit goals</p>	<ul style="list-style-type: none"> ▪ Documentation (e.g., assessment blueprint) of the full, or prioritized set of unit goals addressed by the assessment and the breadth and depth of the assessment content represented by goals
<p>The assessment questions, tasks, or activities are accessible and fair for all English Learners (e.g., provide multiple means of engagement, representation, and expression; does not contain images or references that would be unfamiliar, confusing, or offensive to some students)</p>	<ul style="list-style-type: none"> ▪ Peer review of accessibility and fairness of assessment questions, tasks, or activities for English Learners ▪ Feedback from families and community members about the accessibility and fairness of assessment questions, tasks, or activities for English Learners ▪ Student feedback on the accessibility and fairness of assessment questions, tasks, or activities for English Learners

Proposition 2: The assessment provides all students with the opportunity to show where they are in their learning through multiple modalities.

CLAIMS	EVIDENCE
<p>The assessment questions or tasks encompass multiple modalities and entry points and, where appropriate, students have an opportunity to select how they will demonstrate what they know and can do with language</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on whether the assessment provides all students with the opportunity to show where they are in their learning through multiple modalities ▪ Documentation of peer review on whether the assessment provides all students with the opportunity to show where they are in their learning through multiple modalities
<p>The assessment provides students with an opportunity to self-assess their own performance in terms of academic content and language development for the unit</p>	<ul style="list-style-type: none"> ▪ Student feedback on whether they were sufficiently supported to assess their own learning

Proposition 3: The assessment reflects meaningful, worthwhile tasks that provide all students with an entry point and that challenge the upper reaches of students' language competence and conceptual understanding.

CLAIMS	EVIDENCE
<p>The teacher(s) designs the assessment based on an understanding of the class's prior learning and range of new learning in academic content (conceptual understanding and analytic practices) and language usage of the current unit</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on the degree to which the assessment is based on an understanding of the class's prior learning and range of new learning in the unit
<p>The assessment content reflects high quality classroom learning for English Learners that provides all students with an entry point and that challenges the upper reaches of students' language competence and conceptual understanding</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on the degree to which the assessment was meaningful and offered worthwhile tasks that provided all students with an entry point and challenged the upper reaches of their language competence and conceptual understanding
<p>The assessment offers the potential to consolidate understanding, think from a new perspective, or reconsider one's own thinking</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on the degree to which the assessment offers the potential to consolidate understanding, think from a new perspective, or reconsider one's own thinking ▪ Student feedback on the above

Proposition 4: The inferences from the scores yielded from students’ responses provides actionable information that is asset-oriented and proleptic, future-oriented (i.e., it provides information about what students know and can do so that teachers can build with the students’ current learning status to secure progress).

CLAIMS	EVIDENCE
<p>Students have the opportunity to provide evidence of their emerging learning in a variety of modalities (e.g., oral, written, drawings, video), and, if relevant, languages</p>	<ul style="list-style-type: none"> ▪ Teacher reflection on and evaluation of opportunities provided for a variety of modalities and degree to which learning was revealed
<p>The assessment provides sufficient scaffolding and/or signals where students can scaffold for themselves or ask for support so that students are able to reveal their current learning status (e.g., think-pair-share, language/writing frames, graphic organizers, discussion-bookmark)</p>	<ul style="list-style-type: none"> ▪ Teacher reflection on and evaluation of whether sufficient scaffolding was provided and the degree to which it supported students to reveal current learning status or if more, or less scaffolding is required ▪ Peer classroom observation on aspects requested by observed teacher (e.g., degree and utility of scaffolding)
<p>The teacher has sufficient understanding of content and language to anticipate a range of student responses in order to probe further and to deepen insights into student learning</p>	<ul style="list-style-type: none"> ▪ Teacher reflection on and evaluation of the sufficiency of their understanding to anticipate students’ responses and probe further ▪ Peer classroom observation on aspects requested by observed teacher (e.g., use of probes, responses to students)
<p>Across multiple students, the teacher is able to recognize similarities in responses in order to inform future next learning steps</p>	<ul style="list-style-type: none"> ▪ Teacher reflection on and evaluation of their ability to recognize similarities in responses in order to inform future next learning steps ▪ Peer classroom observation on aspects requested by observed teacher (e.g., ability to recognize similarities to inform next learning steps)

Proposition 5: The assessment scores can be appropriately used by both the students and the teacher to make inferences about the students' learning status and take action to move learning forward.

CLAIMS	EVIDENCE
The qualitative and quantitative insights are sufficient to inform a judgment about next steps to advance learning (e.g., responses to multiple questions, analysis of student work, observations of student activity)	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation on sufficiency of insights and level of detail to take action to advance learning, and the ability of students to use feedback (and the quality of feedback for this purpose) ▪ Student learning progresses in terms of language, conceptual understanding, and analytical practices (e.g., subsequent performance on classroom summative assessments)
The level of detail is adequate to make a decision about next steps to advance learning	<ul style="list-style-type: none"> ▪ Student learning progresses as above
The students are able to engage in metacognitive activity about their learning and respond to external feedback or take action as a result of their own internal feedback	<ul style="list-style-type: none"> ▪ Student interview or feedback. Protocol for interview/feedback. For example, 5-point Likert scale for feedback: I understand what is expected of me in my learning; the feedback from my teacher/peers is helpful; I use feedback to improve my own learning; I am able to monitor and assess my own learning while I am learning; I feel good about my learning/progress/myself as a student
The student better understands the learning intentions and performance expectations even if they are not yet fully achieved	
The student feels positive about their own ability, identity, and/or trajectory	

Classroom Summative Assessment (end of course or year)

The purpose of classroom summative assessment for English Learners is to ascertain the achievement of students at the end of a course or a year relative to course objectives or state standards. Scores from these assessments may determine students' final grades or achievement levels for the course, which may, in turn, affect their placement in or grant access to, future courses. Individual teachers can use the assessment results to reflect on the

effectiveness of their teaching and make plans about any changes or improvements that are needed. Teachers, school and district administrators can examine patterns of achievement across classrooms and grades to inform decisions about policies, programs, and resources in relation to English Learners.

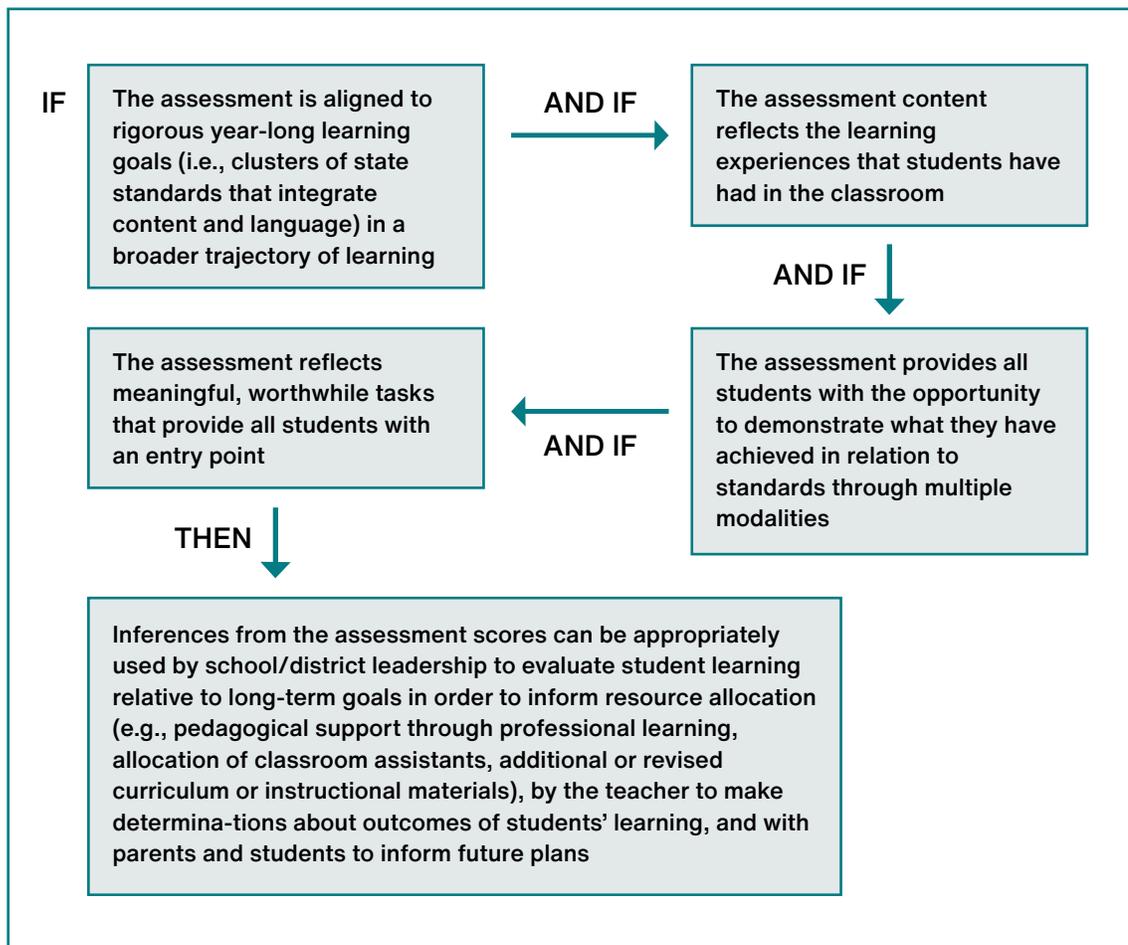
The results of these assessments can also be reported to students and parents/guardians to inform future plans for continued and optimal support for the students.

Teachers and school administrators can examine the results of these classroom end-of-course or end-of-year summative assessments with those of the end-of-unit assessments and vice versa to evaluate the degree of consistency between them. In the event of inconsistencies, teachers and administrators will need to investigate the

potential contributory reasons in order to determine whether revisions to one or both sets of assessments are needed.

The logic model for classroom summative (end of course or year) is shown in Figure 7.

Figure 7. Logic Model for Classroom Summative Assessment (end of course or year)



Proposition 1: The assessment is aligned to rigorous long-term learning goals (the competencies expected by the state standards or end-of-course goals).

CLAIMS	EVIDENCE
<p>The teacher (s) has clearly identified goals from state standards that have been taught throughout the year or end-of-course goal that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them</p>	<ul style="list-style-type: none"> ▪ Documentation (e.g., assessment blueprint) of the alignment between the assessment content and clusters of state standards or end-of-course goals, and the cognitive and language complexity of assessment questions and tasks
<p>The assessment questions and tasks are an appropriate representation of the goals being assessed (e.g., not so broad that it contains dimensions that are irrelevant to the goals, nor so narrow that it fails to include important dimensions of the goals)</p>	<ul style="list-style-type: none"> ▪ Documentation (e.g., assessment blueprint) of the representation of goals to be assessed
<p>The assessment questions, tasks, or activities are accessible and fair for all English Learners (e.g., provide multiple means of engagement, representation, and expression; does not contain images or references that would be unfamiliar, confusing, or offensive to some students)</p>	<ul style="list-style-type: none"> ▪ Documentation of external review of accessibility and fairness of assessment questions and tasks for English Learners (e.g., peers, district subject-matter experts, administrators, assessment experts, representative families, and community members vital to families) of all the above

Proposition 2: The assessment reflects the learning experiences that students have had in the classroom.

CLAIMS	EVIDENCE
<p>The assessment reflects the integration of academic content learning (conceptual understanding and analytic practices) and language use</p>	<ul style="list-style-type: none"> ▪ Documentation of the degree to which the assessment reflects: <ul style="list-style-type: none"> – the integration of academic content and language use – meaningful, worthwhile tasks that offer an entry point for the range of students in the class/grade-level – the contextual factors of students’ learning experiences (e.g., different degrees of scaffolding, sensitivity to students’ cultural backgrounds, using oral language to interact with others to get things done) ▪ Students’ feedback on how their classroom learning experiences and their lived experiences are reflected in the assessment ▪ Documentation of external review (e.g., peers, district subject-matter experts, administrators, assessment experts, representative family members, and community members vital to families) of all the above
<p>The assessment reflects meaningful, worthwhile tasks that offer an entry point for the range of students in the class/grade-level</p>	
<p>The assessment reflects the contextual factors of students’ learning experiences (e.g., different degrees of scaffolding, sensitivity to students’ cultural backgrounds and lived experiences, using oral language to interact with others to get things done)</p>	

Proposition 3: The assessment provides all students with the opportunity to demonstrate what they have achieved in relation to long-term goals through multiple modalities (e.g., video, audio, graphic representations, writing), and, if relevant, languages.

CLAIMS	EVIDENCE
<p>The assessment accounts for the potential range of students present in the class/grade-level so that all students have the opportunity to show what they have achieved in their learning</p>	<ul style="list-style-type: none"> ▪ Documentation of whether the assessment encompasses the full range of where students may start and end the year ▪ Documentation of the degree to which sufficient scaffolding was provided, the degree to which it supported students to reveal current learning status, or if more or less scaffolding is required
<p>The assessment questions or tasks encompass multiple modalities, and, where appropriate, students have an opportunity to select how they will demonstrate what they know and can do with academic content and language</p>	<ul style="list-style-type: none"> ▪ Documentation of the degree to which the assessment provides for a variety of modalities and the degree to which learning was revealed ▪ Student feedback on opportunities to select how they will demonstrate what they know and can do with academic content and language (e.g., Were the choices offered sufficiently broad so that students could find an optimum mode? Was the opportunity worthwhile?)
<p>The assessment provides students with an opportunity to assess their own performance in terms of academic content and language development related to the assessment construct</p>	<ul style="list-style-type: none"> ▪ Student feedback on opportunities to assess their own performance (e.g., Was it valuable? Were students given adequate time for self-assessment? Did they need more support embedded in the assessment to do this?) ▪ Documentation of external review (e.g., peers, district subject-matter experts, administrators, assessment experts, representative family members, and community members vital to families) of all the above

Proposition 4: The assessment can be appropriately used by school/district leadership to evaluate student learning relative to long-term goals (e.g., standards, end-of-course goals) in order to inform resource allocation by the teacher to make determinations about outcomes of students’ learning and with students and parents to inform future plans.

CLAIMS	EVIDENCE
<p>The inferences from the scores are sufficient to inform stakeholders’ judgments about learning relative to long-term goals and resource allocation</p>	<ul style="list-style-type: none"> ▪ Questions and tasks have appropriately developed rubrics or scoring guides and exemplar performances that illustrate each level of the rubric ▪ Application of moderation⁶ process within a COP calibrates teacher judgments against the assessment rubric/scoring protocol to help ensure a common understanding of quality and support reliability across scorers ▪ If sub-scores are to be reported (e.g., a separate score for content understanding and for language use), the assessment blueprint (see Proposition 1) should a) reflect how items map to the scoring categories with a sufficient number of items to warrant a sub-score and b) show that sub-scores actually represent different constructs and are not just unreliable estimates of the total score
<p>The inferences from the scores are sufficient to inform the teacher’s revisions to curriculum/ instructional plans for future use</p>	<ul style="list-style-type: none"> ▪ Teacher surveys on impact of data on future curriculum/instructional plans
<p>The inferences from the scores are sufficient to support teacher reflection on the quality of their instruction</p>	<ul style="list-style-type: none"> ▪ Teacher surveys on impact of data on reflections on quality of instruction
<p>The student is able to engage in metacognitive activity with respect to future learning</p>	<ul style="list-style-type: none"> ▪ Student feedback on: <ul style="list-style-type: none"> – the degree to which the assessment helped them think about their own learning – whether the assessment results were useful in assisting them to set goals about areas of improvement or other future learning goals

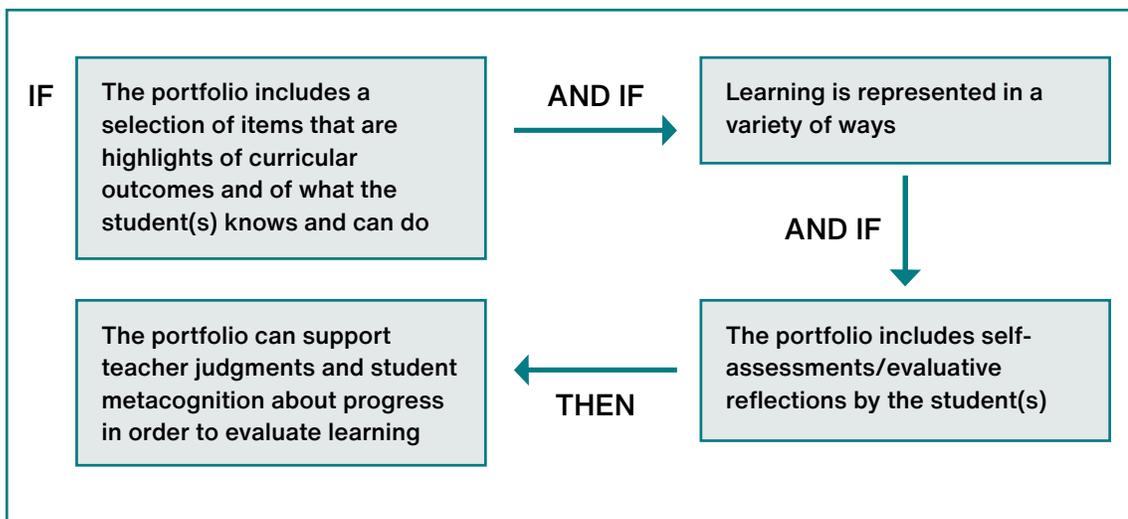
⁶ **Moderation** is a structured process: groups of teachers discuss samples of student work at different levels of quality in conjunction with associated standards and rubrics. The process is intended to develop consistency of interpretation of each level of the rubric across teachers (e.g., Connolly et al., 2012; Wyatt-Smith et al., 2010).

Year-Long Student Portfolio

The year-long student portfolio has four purposes: (1) monitor learning over the course of the year; (2) evaluate achievement at the end of the year; (3) support students to engage in metacognition, reflecting on their own learning path; and (4) inform localized high-stakes decisions (e.g., engagement in rigorous grade-level

learning, access to the core curriculum and demanding learning opportunities, classification and reclassification) about individual students when teacher-based judgments about individual student’s strengths and weaknesses differ from performance on other assessments, including those for the purpose of accountability. The student portfolio logic model is shown in Figure 8.

Figure 8. Logic Model for Year-Long Student Portfolio



Proposition 1: The portfolio includes a selection of items that are highlights of curricular outcomes and of what the student(s) knows and can do.

CLAIMS	EVIDENCE
<p>The teacher(s) has clearly identified unit goals and long-term goals from clusters of state standards that integrate the development of key conceptual understandings, analytic practices, and the language needed to express them</p>	<ul style="list-style-type: none"> ▪ Teacher description of the content and language unit learning goals, how they align with standards, and how they are situated within a trajectory of learning for the class
<p>The teacher has a clear understanding of the intended trajectory of learning for the class from the beginning of the year to the end</p>	<ul style="list-style-type: none"> ▪ Teacher reflection and evaluation of the progression of learning for the class over the year
<p>The teacher has clearly articulated criteria for portfolio selections and has communicated them to, or co-developed them with, students</p>	<ul style="list-style-type: none"> ▪ Documentation of what was communicated to, or developed with, students about unit learning goals and criteria for portfolio selections ▪ Peer observation of communication/ co-development ▪ Student interview or survey about their understanding of the portfolio process (learning goals, performance criteria, and criteria for item selection, etc.)
<p>The portfolio selections represent the breadth and depth of cognitive complexity and language usage embodied in the medium-term goals (e.g., unit) and the long-term goals (e.g., clusters of standards)</p>	<ul style="list-style-type: none"> ▪ Peer teacher review and feedback on the alignment between the breadth and depth of cognitive complexity and language usage represented by the unit goals and the evidence selected by students

Proposition 2: Learning and progress is represented in a variety of ways.

CLAIMS	EVIDENCE
The portfolio includes work created in multiple modalities	<ul style="list-style-type: none"> Teacher creates a unit-specific checklist (to both support opportunity to learn (OTL) and representation of thinking and language use) that identifies the expectations for portfolio selections, (e.g., multiple modalities, opportunities to demonstrate academic content and language development, variety of work products, evidence of student growth, evidence of best effort) Peer teacher review and feedback on the alignment between the content of the unit and opportunities for students to demonstrate learning and progress in a variety of ways, including specific periodic tasks Evidence of ongoing use of checklist by both teacher and students as way of monitoring OTL and representation of thinking and language use Evidence of student use of teacher created protocol of student reflections Peer teacher review and feedback on the tasks that intend to allow students to demonstrate their ability to synthesize their understandings and use language for that purpose in a particular domain or subdomain
The portfolio includes work to show the process of learning language and content (e.g., video/audio of group collaborative problem solving)	
The portfolio includes evidence of student growth (e.g., early and final drafts)	
The portfolio includes work that represents a student’s best effort selected by the student against specific criteria	
The portfolio includes structured student reflections on what they have learned and how their learning changed relative to key learning goals	
The portfolio includes tasks (e.g., a product, performance, or presentation) that allow students to demonstrate their ability to synthesize their understandings and use language for that purpose in a particular domain or subdomain	

Proposition 3: The portfolio includes students’ self-assessments/evaluative reflections.

CLAIMS	EVIDENCE
The student self-assessment is guided by protocols and rubrics linked to unit and end-of-year goals	<ul style="list-style-type: none"> Teacher review and reflection on the depth of student reflections, the degree to which they are connected to unit and end-of-year goals, and show evidence of increasing self-efficacy, motivation, and learning identity Peer teacher review and feedback of teacher reflections and sample of student self-assessments/reflections
Throughout the unit portfolio, students reflect on their sense of self-efficacy, motivation and learning identity	

Proposition 4: The portfolio can support teacher judgments and student metacognition about progress in order to evaluate learning.

CLAIMS	EVIDENCE
The assessment rubric/scoring protocol describes learning expectations for the year in sufficient detail to support evaluation of students' portfolio in terms of both content and language and helps the teacher distinguish among levels of performance in terms of both conceptual development, application of analytical practices, and language development	<ul style="list-style-type: none"> Use of the assessment rubric/scoring protocol across teachers Application of low-stakes moderation process within a COP to calibrate teacher judgments against the assessment rubric/scoring protocol to help ensure a common understanding of quality
The teacher's interpretation focuses on what students can do in terms of content and language as well as highlighting areas of need and potential growth for teaching and learning	<ul style="list-style-type: none"> Individual reflection of assessment data to support interpretation of what students can do in terms of content and language as well as highlighting areas of need and potential growth for teaching and learning
Teachers reflect on the coherence between their judgments of student learning based on their portfolios compared to their judgments of student performance on the periodic tasks	<ul style="list-style-type: none"> Calibration discussion in COP about alignment between judgments of student performance on the portfolio performances and periodic tasks

Support for COP Review of Validity Evidence

To support COP discussions of the validity evidence, two protocols are included in Appendix B and Appendix C.⁷ The protocols provide guidance on:

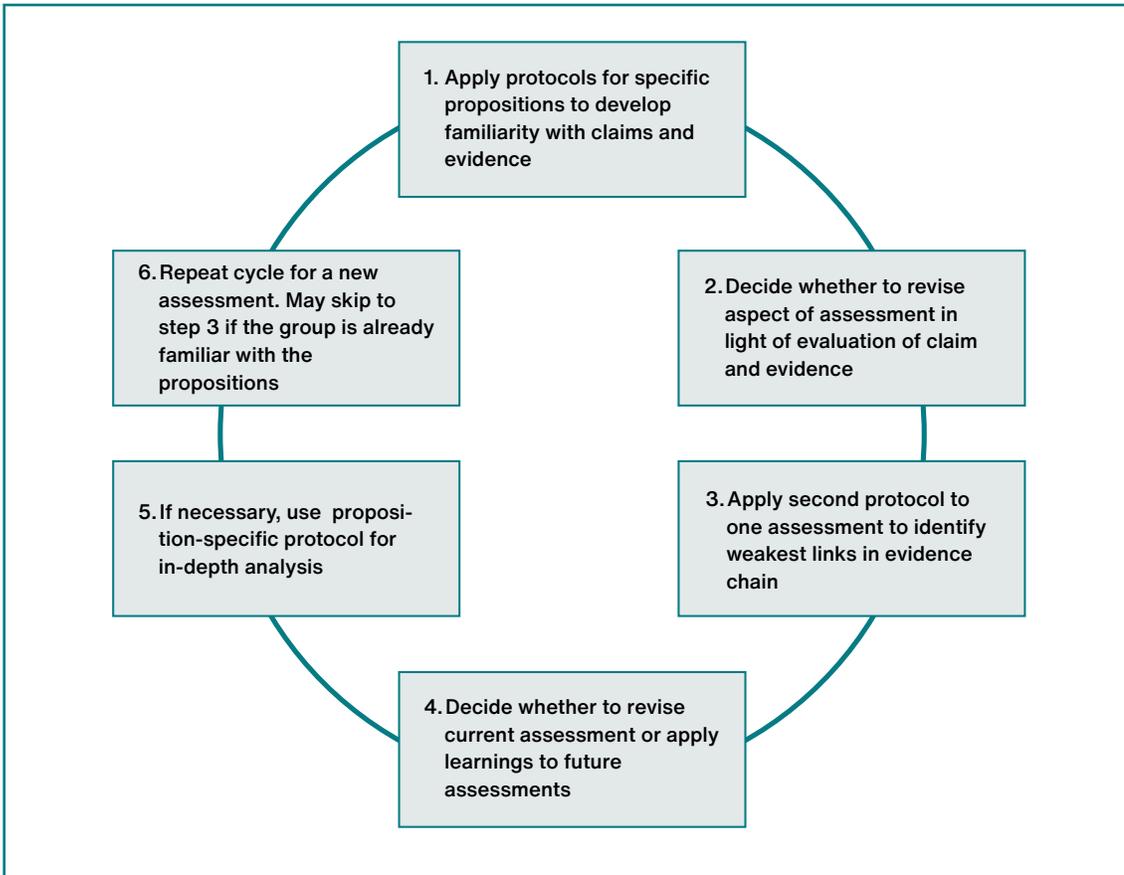
- how to analyze propositions and claims to help ensure a common understanding among teachers
- what a teacher would present to the COP

- what the members of the COP should listen for
- what questions the members of the COP would ask
- how to plan outcomes they are collectively moving towards

A process for using the protocols is shown in Figure 9.

⁷ The CAS will be partially implemented over three years in an iterative study of secondary-level teachers' use of high-quality replacement units. This study will be an opportunity to try out and refine both protocols, create additional resources for teachers as needed, and potentially collect video of COP discussions to support teacher learning about these processes.

Figure 9: Process for Using Protocols



Proposition Specific Protocols: The first protocol guides COP members through three stages of reviewing and evaluating evidence for a specific proposition: (1) the Initial Planning stage will identify the targeted proposition, ensure common understanding, and identify the types of relevant evidence that may need to be collected for the evaluation of the validity argument; (2) the validity review is the core work of the COP to review evidence for propositions' claims and identify strengths and areas to improve; and (3) during the Action Review, COP members examine whether recommended changes were made or to reflect on how the discussion from the Validity Review impacted ongoing assessment development.

The COP members may decide to apply this protocol initially to some, none, or all of the propositions associated with a specific type of assessment, depending on their familiarity with each of the propositions for the assessment on which they are focused. For example, the COP members may decide that they first want to better understand what it might mean to develop assessments that allow students to use multiple modalities. The group could use the protocol for Classroom Summative Proposition 2 and bring examples of assessments that either use multiple modalities, or that members think could be modified to incorporate multiple modalities. These discussions can help teachers to develop assessment literacy in collaborative

ways. Other aspects of shared assessments are not discussed at this stage because the group elected to focus on a specific aspect.

Assessment Specific Protocols: Once the COP members have developed a level of familiarity with each proposition and its claims and evidence, they can use the protocol for the particular assessment (e.g., classroom summative). This protocol addresses all five propositions and supports a discussion about the chain of evidence that starts with the unit goals and ends with how the summative unit assessment will be used. This protocol will support COP discussions as the members articulate the chain of evidence for a specific assessment. The COP may need to take some time to decide what of the possible forms of evidence they are going to collect. For instance, will they survey students? Will they engage in peer teacher review? Careful planning will be needed prior to collecting evidence for review.

When the evidence has been amassed, the purpose of the discussion is to identify places where the evidence might not support a specific claim, or where an entire proposition cannot be supported. The COP members can then decide either to discuss how to improve the assessment, or decide when that work could be done. It is important to capture sufficiently detailed notes about the decisions that could assist future revisions if they are not going to be made immediately.

COP Discussions Using Both Types of Protocols: The entire COP discussion is intended to be formative: the logic model and propositions describe the ideal nature and conditions of an assessment. The evidence describes the current status of the assessment, and the COP discussions examine the gap between intentions and

reality. Lessons learned from a review of one assessment can be applied to future assessments. The COP members might decide to focus their efforts on revising an upcoming assessment and then examine it in terms of the validity evidence to determine whether the revisions helped them get closer to the ideal model as described by the propositions. The COP might also decide to return to the elaborated protocol because they identified a specific proposition as an area of weakness across their assessments and want to work on that area in a targeted way.

It is important to note that it would not be practical or even advisable that a COP engage in this level of review for every classroom assessment. An in-depth review of one or two per semester may be sufficient if individual and COP time is also spent on revising other assessments that may have the same weaknesses.

External Large-Scale Summative Assessments

The primary purpose that external large-scale assessments serve is for accountability in relation to federal and state policy. Others have already written extensively on how to apply an argument-based approach to validity to large-scale summative assessments, generally (e.g., Bachman, 2005; Kane, 2013); we do not attempt to duplicate those efforts here. Instead, this section focuses on aspects of validation that are unique and central to the particular role and goal of large-scale summative assessment in the context of this CAS for English Learners in secondary grades.

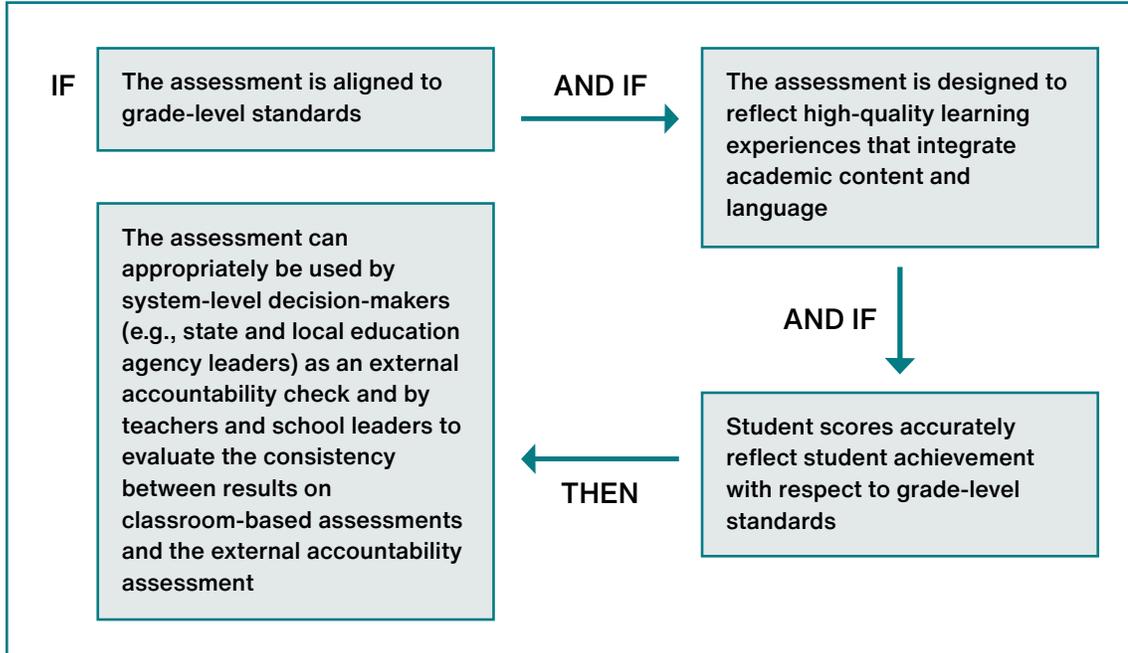
One such aspect is the centrality of integrated language and content within the CAS Framework. Whereas teachers have the control

to use this integrated approach in designing their own classroom-based assessment opportunities, large-scale summative assessments are currently segregated along these lines – they measure either language or content, but not both. We therefore include propositions, claims, and evidence that would be necessary to demonstrate assessment validity for this purpose. Neither the process nor the types of evidence that would be required are different from current practice on the surface – that is, the same kinds of alignment studies and standard setting panels would be required – but the construct of measurement would need to be adjusted to reflect the language and content integration.

A second important difference in our vision is the relationship of large-scale summative assessments to learning and to teacher-generated assessments elsewhere in the system.

In the CAS Framework, these large-scale assessments are the only type of assessment that is external to the classroom. The responsibility for validating interpretations and uses for standardized assessments falls to assessment developers and test score users. Teacher-generated validity evidence from other assessment forms can serve as external criteria against which the validity of large-scale scores may be interpreted or compared. Figure 10 shows the logic model for large-scale summative assessment.

Figure 10. Logic Model for Large-Scale Summative Assessment



Propositions, claims and sources of evidence associated with this logic model are described next.

Proposition 1: The assessment is aligned to clusters of grade-level learning standards.

CLAIMS	EVIDENCE
The assessment is designed from clearly identified content and language development goals derived from clusters of state standards or end-of-course goals	<ul style="list-style-type: none"> The assessment blueprint specifies how content and language development will be assessed Documentation from alignment study (judgments from subject matter experts, including teachers with expertise in language development and English Learner education)
The assessment blueprint reflects a comprehensive sampling of content, analytic practices, and language use	<ul style="list-style-type: none"> Questions and tasks are consistent with the blueprint and comprehensively reflect the expectations of the content and language standards
The assessment questions or tasks are an appropriate representation of the goals being assessed (e.g., not so broad that it contains dimensions that are irrelevant to the goals nor so narrow that it fails to include the important dimensions of the goals)	<ul style="list-style-type: none"> Documentation of the breadth and depth of cognitive complexity and language usage reflected in the assessment questions or tasks
The assessment questions or tasks are accessible and fair for all English Learners (e.g., provide multiple means of engagement, representation, and expression; do not contain images or references that would be unfamiliar, confusing, or offensive to some students)	<ul style="list-style-type: none"> Documentation from accessibility and fairness reviews by experts

Proposition 2: The assessment is designed to reflect high-quality learning experiences that integrate academic content and language.

CLAIMS	EVIDENCE
The assessment reflects the integration of academic content learning (conceptual understanding and analytic practices) and language use	<ul style="list-style-type: none"> Documentation from expert review to examine integration of content and language
The assessment reflects meaningful, worthwhile tasks that give all students the opportunity to demonstrate their achievement relative to the standards	<ul style="list-style-type: none"> Documentation from expert review on the tasks relative to this claim Documentation from student cog labs to determine whether tasks have multiple entry points (students have a range of English language proficiency)
The assessment provides opportunities for students to demonstrate what they can do using oral and written language in an academic context	<ul style="list-style-type: none"> Assessment design provides opportunities for all students to demonstrate what they can do (e.g., multi-stage adaptive assessment)

Proposition 3: Student scores accurately reflect student achievement with respect to grade-level standards.

CLAIMS	EVIDENCE
The assessment questions, tasks, or activities are accessible and fair for all English Learners (e.g., provide multiple means of engagement, representation, and expression; do not contain images or references that would be unfamiliar, confusing, or offensive to some students)	<ul style="list-style-type: none"> ▪ Expert review of accessibility and fairness of assessment questions, tasks, or activities for English Learners ▪ Evidence from students on the accessibility and fairness of assessment questions, tasks, or activities for English Learners (e.g., cog labs, interviews)
Scoring of student responses is reliable and free of bias	<ul style="list-style-type: none"> ▪ Constructed response items have appropriately developed rubrics or scoring guides and exemplar performances ▪ Raters are sufficiently trained and calibrated before engaging in live scoring
Total scores and reportable sub-scores are sufficiently reliable	<ul style="list-style-type: none"> ▪ Documentation of psychometric analysis (reliability, factor analyses, DIF)
The score scale reflects the full distribution of where students are likely to start and end the year	
Cut-scores are identified in a defensible manner	<ul style="list-style-type: none"> ▪ Documentation of standard setting process and participants

Proposition 4: The assessment can appropriately be used by system-level decision-makers (e.g., SEA and LEA leaders) as an external accountability check and by teachers and school leaders to evaluate the consistency between results on classroom-based assessments and the external accountability assessment.

CLAIMS	EVIDENCE
The inferences from the scores are sufficient to inform stakeholders' judgments about learning relative to standards	<ul style="list-style-type: none"> ▪ If sub-scores are to be reported (e.g., a separate score for content understanding and for language use), the assessment blueprint (see Proposition 1) should reflect how items map to the scoring categories along with psychometric evidence of sufficient reliability of the sub-score to warrant separated reporting
Stakeholders can use assessment information in conjunction with other data sources to inform resource allocation and other policy decisions	<ul style="list-style-type: none"> ▪ Surveys of teachers, principals, and other stakeholders of how state and local leaders use assessment data
Teachers and school leaders can compare assessment information from the accountability assessment and classroom summative assessments to evaluate the consistency between them	<ul style="list-style-type: none"> ▪ Documentation of the multiple sources of summative assessment data by teachers and school leaders to inform comparisons and to identify whether additional reviews of classroom summative assessment are warranted

Validity of the System

In addition to collecting and evaluating evidence for the validity of interpretations and uses of the information yielded by the individual assessments, judgments about how well the assessments work together as a system to accomplish systemic goals will also be necessary. Common goals for an assessment system might include (1) supporting all students to master specific skills and knowledge deemed valuable and necessary for success after graduation, (2) minimizing achievement gaps or differences among groups who are believed to have the same underlying ability distribution, and (3) meeting external requirements or standards related to accountability or equity, for example. Specific additional and central goals for this CAS are for English Learners in secondary grades to have rigorous and engaging learning opportunities, and for teachers of these students to participate with skill and confidence in the practice of designing, interpreting, and validating assessments that provide meaningful information about English Learners' knowledge, skills, and powerful use of language.

The CAS Framework lays out a roadmap for changing assessments in the system, and it is unlikely that all assessments will simultaneously undergo a validation process. An important component of this long-term undertaking is that as the validity evidence for each assessment is evaluated, the entire system should be kept in review in terms of the degree to which it exhibits the three properties of an assessment system: comprehensiveness, coherence, and continuity (3Cs) discussed at the beginning of the document. Such a review may result in changes to the

assessments therein in order to ensure that they all emanate from the same set of goals related to language and content learning, communicate a shared vision of what is important for students to know and be able to do with content and language, and push teaching and learning in a common direction so that the demands of understanding concepts, practices, and relationships are not privileged above the demands of linguistic resources, nor vice versa.

Some questions to guide a review of the CAS in relation to the 3Cs include:

- Are all the assessments aligned to the complementary learning goals (of different grain sizes depending on the assessment)?
- Do the assessments reflect the same model of teaching content and analytic learning and the development of language resources needed to learn that content to English Learners?
- Is this model of teaching clear and consistent at all levels (e.g., formative assessment to classroom summative to external assessment)?
- Do the assessments at all levels reflect the same conceptualization of how students learn content and language simultaneously?
- Do the assessments provide various ways for English Learners to demonstrate their competence in content and language?
- Are the assessments linked conceptually so that change over time can be observed and interpreted?

These questions are not intended as a complete set, but rather as starting points to consider system validity.

The key point about the system is that combination of assessments should provide mutually complementary views of student learning that together reinforce important goals and teaching and learning practices while strengthening the validity of the system as a whole (Herman, 2010).

Policy Implications

Implementing this inverted assessment system has policy implications, which are outside the scope of this document, but will need to be attended to, nevertheless. These include addressing the current system-wide distortions caused by the punitive use of accountability results for teachers and schools, and also fundamentally changing the nature of accountability.

A common argument for heavily privileging large-scale standardized assessments is their relative cost and reliability compared to more person-oriented approaches such as those represented in the framework. While it is true that machine-scored items may be cheaper to purchase, score, and administer than portfolios, it is also true that such instruments have a direct — and often negative — impact on what teachers do in the classroom and what English Learners have an opportunity to learn. Multiple choice assessments, portfolios, and any kind of assessment do not operate in a vacuum; they are administered and used in the larger context of teaching and learning and their costs and benefits should be evaluated in that larger context. The CAS Framework is intended to show the many benefits of implementing an assessment system that prioritizes quality teaching and learning that will lead to successful outcomes for every English Learner.

References

- American Educational Research Association (AERA), American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, 2(1), 1-34.
- Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21(1) 49–97, DOI: 10.1080/03057269308560014
- Brooks, M. D. (2018). Pushing past myths: Designing instruction for long-term English Learners. *TESOL quarterly*, 52(1), 221–233.
- Bruner, J. (1983). *Child's talk*. New York, NY: Norton.
- Callahan, R. M. (2005). Tracking and high school English Learners: Limiting opportunity to learn. *American Educational Research Journal*, 42(2), 305–328. <https://doi.org/10.3102/00028312042002305>
- Callahan, R. M., & Shifrer, D. (2016). Equitable access for secondary English Learner students. *Educational Administration Quarterly*. <https://doi.org/10.1177/0013161X16648190>
- Carlson, D., & Knowles, J. E. (2016). The effect of English language learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management*, 35(3), 559–586. <https://doi.org/10.1002/pam.21908>
- Connolly, S., Klenowski, V., & Wyatt-Smith, C. M. (2012). Moderation and consistency of teacher judgement: Teachers' views. *British Educational Research Journal*, 38(4), 593-614.
- Durán, R. P. (2008). Assessing English-language learners' achievement. *Review of Research in Education*, 32(1), 292–327.
- Glick, Y., & Walqui, A. (2021). Affordances in the development of student voice and agency. The case of bureaucratically labeled Long Term English Learners. In: A. Kibler, G. Valdés, & A. Walqui (Eds.), *Reconceptualizing the role of critical dialogue in American classrooms. Promoting equity through dialogic education*. New York, NY: Routledge.
- Gordon, E. W. (2020). Toward Assessment in the Service of Learning. *Educational Measurement: Issues and Practice*, 39(3), 72–78. <https://doi.org/10.1111/emip.12370>
- Gordon, E. W., Gordon, E. W., Aber, L., & Berliner, D. (2012). *Changing paradigms for education. Assessment, Teaching, and Learning*, 2(2).
- Heritage, M., Faulkner-Bond, M. & Walqui, A. (2021). *A new direction for assessing English learners in the secondary grades*. San Francisco: WestEd

Herman, J. L. (2010). *Coherence: Key to next generation assessment success*. Los Angeles, CA: University of California.

Herman, J. L., Ashbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Guidance for developing and selecting assessments of student growth for use in teacher evaluation systems*. Los Angeles, CA: University of California.

Johnson, A. (2019). A matter of time: Variations in high school course-taking by years-as-EL subgroup. *Educational Evaluation and Policy Analysis*, 41(4), 461–482. <https://doi.org/10.3102/0162373719867087>

Kane, M. T. (2006). *Validation*. *Educational Measurement*, 4(2), 17-64.

Kane, M. (2013). *The argument-based approach to validation*. *School Psychology Review*, 42(4), 448-457.

Kibler, A. K., & Valdés, G. (2016). Conceptualizing language learners: Socioinstitutional mechanisms and their consequences. *The Modern Language Journal*, 100(S1), 96–116.

Levi, T., & Poehner, M. E. (2018). Employing dynamic assessment to enhance agency among L2 learners. In J.P. Lantolf, Poehner, M.E. & M. Swain (Eds.), *The Routledge handbook of sociocultural theory and second language development* (pp. 295–309). New York, NY: Routledge.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109-162.

National Center for Education Statistics (2018). *NAEP data explorer*. Retrieved September 10, 2019, from <https://nces.ed.gov/nationsreportcard/data/>

National Research Council. (2001). *Knowing what students know*. Washington, DC: National Academies Press.

Organisation for Economic Co-operation and Development. (2012). Equity and equality of opportunity. *Education Today 2013: The OECD Perspective*. Paris, France: Author. https://doi.org/10.1787/edu_today-2013-11-en

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97.

Perie, M., & Forte, E. (2011). Developing a validity argument for assessments of students in the margins. In M. Russell & M. Kavanaugh (Eds.), *Assessing Students in the Margins: Challenges, Strategies, and Techniques* (pp. 335-381). Charlotte, NC: Information Age Publishing.

Rogoff, B. (1995) Observing sociocultural activity on three planes: Participatory, appropriation, guided participation, and apprenticeship. In J. Wertsch, P. del Rio, and A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 139–164). Cambridge, UK: Cambridge University Press.

Rosa, J. (2019). *Looking like a language, sounding like a race: Raciolinguistic ideologies and the learning of latinidad*. New York and Oxford: Oxford University Press.

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and practice*, 22(4), 26-33.

Swain, M. (2006). Linguaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). London: Continuum.

Swain, M., & Lapkin, S. (2011). Linguaging as agent and constituent of cognitive change in an older adult: An example. *Canadian Journal of Applied Linguistics*, 14(1), 104–117.

Umansky, I. M. (2016). Leveled and exclusionary tracking: English Learners' access to academic content in middle school. *American Educational Research Journal*, 53(6), 1792–1833. <https://doi.org/10.3102/0002831216675404>

Umansky, I. M., & Dumont, H. (2021). English Learner Labeling: How English Learner Classification in Kindergarten Shapes Teacher Perceptions of Student Skills and the Moderating Role of Bilingual Instructional Settings. *American Educational Research Journal*, 0002831221997571. <https://doi.org/10.3102/0002831221997571>

Umansky, I. M., & Porter, L. (2020). State English learner education policy: A conceptual framework to guide comprehensive policy action. *Education Policy Analysis Archives*, 28(0), 17. <https://doi.org/10.14507/epaa.28.4594>

Volante, L., DeLuca, C., Adie, L., Baker, E., Harju-Luukkainen, H., Heritage, M., Schneider, C., Stobart, G., Tan, K., & Wyatt-Smith, C. (2020). Synergy and tension between large-scale and classroom assessment: International trends. *Educational Measurement: Issues and Practice*, 39(4), 21-29.

Valdés, G., Kibler, A., & Walqui, A. (2014, March). *Changes in the expertise of ESL professionals: Knowledge and action in an era of new standards*. Alexandria, VA: TESOL International Association.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Vygotsky, L.S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

van Lier, L. (2000). From input to affordance: social interactive learning from an ecological perspective. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). New York, NY: Oxford University Press.

van Lier, L. (2004). *The ecology and semiotics of language learning. A sociocultural perspective*. Dordrecht, NL: Kluwer Academic.

van Lier, L., & Walqui, A. (2012). Language and the common core standards. In K. Hakuta & M. Santos (Eds.), *Understanding language: Commissioned papers on language and literacy issues in the common core state standards and next generation science standards* (pp. 44–51). Palo Alto, CA: Stanford University.

Walqui, A. (2006). Scaffolding instruction for English language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism*, 9(2), 159–180.

Walqui, A., & Heritage, M. (2011). Instruction for diverse groups of English language learners. In K. Hakuta & M. Santos (Eds.), *Understanding language: Commissioned papers on language and literacy issues in the Common Core State Standards and Next Generation Science Standards* (pp. 94–103). Palo Alto, CA: Stanford University.

Webb, N. M., & Shavelson, R. J., Generalizability theory: Overview. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol 2, pp. 717-719). Chichester, UK: Wiley.

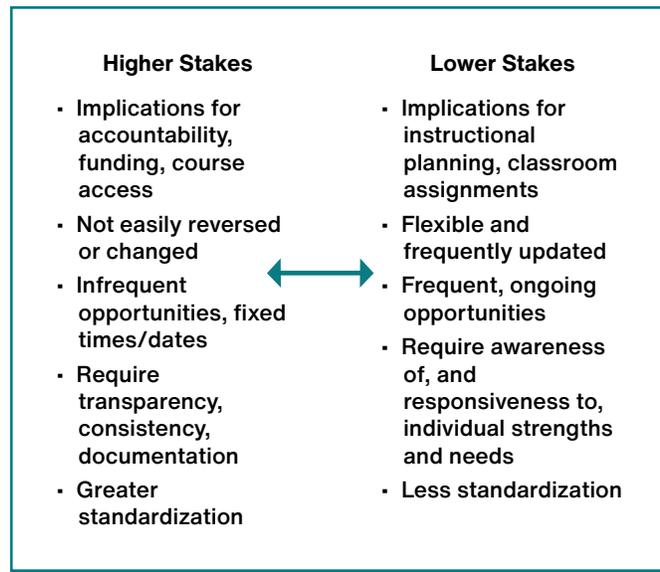
Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). In C. R. Roa & S. Sinharay (Eds.), Reliability coefficients and generalizability theory. *Handbook of statistics*, 26, pp. 81-124.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.

Appendix A: Reliability

Reliability, a necessary but not sufficient condition for validity, refers to the consistency of assessment results across settings, students, and users. For example, if Juanita completes a writing task related to a specific prompt today, tomorrow, or next Wednesday, then we would expect her ability to answer the question to be essentially the same on all three occasions. Without such consistency, we cannot have confidence that student scores are meaningful representations of student knowledge and skills. The question of how high the reliability for an assessment depends on the consequences and stakes of the use of the results (see Figure 1 for characteristics of higher and lower stakes decisions).

Figure A-1. Characteristics of Higher and Lower Stakes Decisions



For assessments that take the form of tests (e.g., those administered at the end of the year to assess achievement of standards), this consistency is measured in the form of **reliability**, which usually involves calculating a reliability coefficient⁸ to determine how well assessment results agree over repeated uses of the assessment. The expectation is that a student who takes the same test on different occasions or in different settings would earn roughly the same score. The higher the stakes of the decisions made from test results, the higher the level of reliability will need to be.

In the case of classroom-based assessment, including formative assessment, Smith (2003) proposes **sufficiency of information** to determine reliability. He suggests that teachers can think about the question “does this assessment provide me with enough information to make a judgment about each student’s level of accomplishment with regard to this learning?” (Smith, 2003, p. 26) to guide instructional next steps. Similar to reliability coefficients, the amount of evidence that a teacher needs will vary based on intended use of the assessment. Contrast three cases: (1) use of a quick poll of student ideas in order to adjust the lesson in the moment based on immediate feedback; (2) an end-of-lesson exit ticket with questions for students to identify something they understand, are puzzled by, and are curious about in order to add to teacher observations and

⁸ Coefficients at or above 0.80 are often considered sufficiently reliable to make decisions about individuals. A higher value, perhaps 0.90 is preferable if decisions have a significant consequence (Webb, Shavelson, & Haertel, 2006).

support a more comprehensive plan for the next lesson; and (3) a longer, more formal task such as problem-solving in mathematics to help the teacher plan out the remainder of time spent on a unit. In each case, the evidence must be sufficient for teachers to feel confident their judgment about students' learning status.

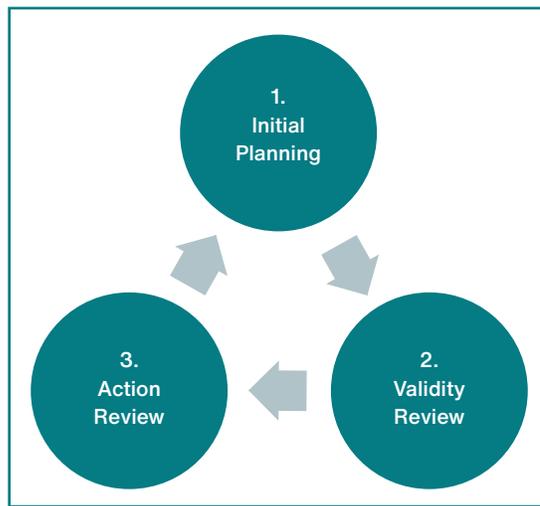
For assessments that take other forms, such as portfolios, consistency is studied through the lens of **generalizability** (Webb & Shavelson, 2005). For generalizability, consistency stems primarily from the prompts, raters, and rating tools (e.g., rubrics) used to produce scores. An important goal is confirming that scores are meaningful representations of student knowledge, rather than idiosyncratic representations of, say, rater preference, or the features of a particular assignment. In the case of Juanita's writing task: if her teacher reads Juanita's written response either tonight, tomorrow, or next Wednesday, then we would expect her to draw the same conclusions about Juanita's strengths and needs (Herman, Ashbacher, & Winter, 1992).

Appendix B: Protocol to Guide a COP Validity Discussion Focused on Single Proposition for a Specific Type of Assessment

Purpose: To help the COP members develop familiarity with the process, members will engage in an in-depth review of the claims and evidence associated with each proposition.

Process: The process is divided into three parts. The Initial Planning stage will identify the targeted proposition, ensure common understanding, and identify the types of relevant evidence that may need to be collected for the evaluation of the validity argument can take place. The validity review is the core work of the COP and entails reviewing evidence for proposition claims and identifying strengths and areas to improve. The final step is the Action Review to examine whether recommended changes were made, or to reflect on how the discussion from the Validity Review impacted ongoing assessment development.

Figure B-1. Three-Part Validity Review



Documentation: We recommend that COP members develop approaches to document discussions and action steps, both to ensure that identified revisions are made to already-reviewed assessments and that plans to revise future assessments are also documented and periodically reviewed. This documentation might be in the form of a shared Google doc that all COP members have access to, or individual teacher journals in which they can capture reflections and action steps.

1. Initial Planning

Evidence Sources

- Documentation of classroom practice (e.g., learning goals presented to students, directions for portfolio selection)
- Individual teacher reflection (e.g., teacher reflection on whether the questions, tasks, and activities are accessible to the range of students' zone of proximal development (ZPD) present within the class)
- Peer observation (e.g., how effectively teachers communicate learning goals to students)
- Student feedback (e.g., survey or interviews)
- Peer feedback (e.g., peer review and feedback on the alignment between the breadth and depth of cognitive complexity and language usage represented by the unit goals)
- Teacher moderation to help ensure a common understanding of quality and to calibrate scoring

Step 1: Identify the type of the assessment for the COP discussion

Step 2: Select the specific proposition for the discussion (we recommend a COP work through the propositions sequentially)

Step 3: Ensure a common understanding among group members of the proposition, claims, and evidence.

Step 4: Review the list of potential evidence sources against the specific evidence described for the targeted proposition.

Step 5: Determine who will present evidence, how and when evidence can be collected, and schedule the Validity Review.

2. Validity Review

Example of Teacher Presentation (based on all claims for Formative Assessment)

- Reflect on your lesson plan and how successful you think you were at articulating lesson-sized goals from the standards. Did they lead to the learning you [the teacher] expected and or not?
- Reflect on one question/task/activity that successfully helped you [the teacher] ascertain the current learning status of individual students' learning in terms of both language and academic content? What was a less successful question/task/activity?
- Reflect on your lesson and how successful you [the teacher] were at integrating your knowledge of the student's language development, conceptual understanding, analytical practices and funds of knowledge into the assessment. Was there one time that was more successful than others? Was there another that was less successful?
- What does this work reveal about the status of student learning relative to goals? Are there other modalities that might be more effective at revealing specific students' learning status? What would be our next steps for each student based on the evidence?
- What judgments were you able to make about student learning based on your evidence? Which judgments do you feel confident about and which are you less confident about?

Step 1: Remind the group of the ground rules for discussion. The focus of the discussion should always be on the assessment and ways in which it can be improved or how strengths in an assessment can be applied to other assessments. The critique should never be of an individual.

Step 2: Invite a group member to present evidence for the claims associated with the targeted proposition, drawing on the sources of evidence identified during the initial planning. While the group member makes the initial presentation, other group members do not interrupt.

Step 3: Once the initial review is complete, group members can ask clarifying questions to help the presenter provide more specific details from a particular source of evidence or to make connections between specific claims and supporting evidence more explicit. This stage of the discussion is focused only on clarification, not evaluation of the evidence.

Step 4: The fourth step is to develop a consensus judgment of the validity of evidence for the proposition. Review the proposition and summarize the evidence for each claim. Where is the evidence convincing? Where could the assessment or assessment process be improved? Where does the evidence need to be strengthened? How can we do this?

Step 5: The final step is to discuss and document how understandings about the proposition and/or nature of the assessment could be applied both to revisions of the specific assessment under review and to future assessments of this format. Whether in a shared Google doc or individual teacher journals documenting reflections and action steps is critical for moving to the Application Review stage.

3. Application Review

Step 1: Reconvene to review what the identified revisions were from the Validity Review meeting.

Step 2: Confirm whether revisions were made to the assessment as planned. For a formative assessment question or probe, this might be just making a note in a lesson plan for the following year about a revision. An end of unit revision that is used by several teachers might require a more substantive revision.

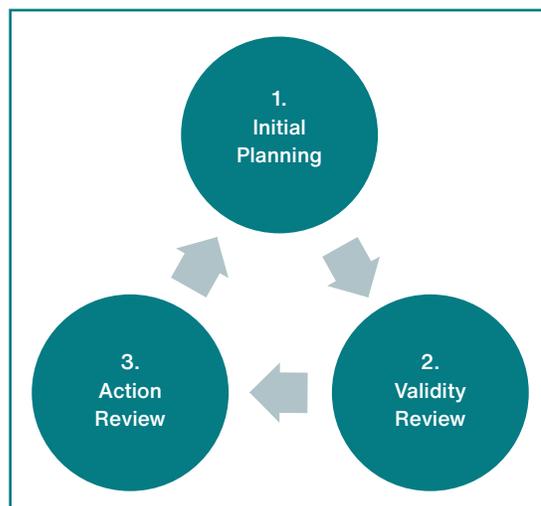
Step 3: Provide an opportunity for COP members to discuss any ways in which the previous Validity Review discussion has impacted their assessment work since the meeting.

Appendix C: Protocol to Guide a COP Validity Discussion Focused on the Set of Propositions for a Specific Type of Assessment

Purpose: To help the COP members engage in a comprehensive review of the claims and evidence associated with the full set of propositions for a specific assessment.

Process: The process is divided into three parts. The Initial Planning stage will identify the targeted assessment, ensure common understanding and identify the types of relevant evidence that may need to be collected for the evaluation of the validity argument can take place. The Validity Review is the core work of the COP to review evidence for proposition claims and identify strengths and areas to improve. The final step is the Action Review to examine whether recommended changes were made or to reflect on how the discussion from the Validity Review impacted ongoing assessment development.

Figure C-1. Three-Part Validity Review



Documentation

We recommend that COP members develop approaches to document discussions and action steps, both to ensure that identified revisions are made to already-reviewed assessments and that plans to revise future assessments are also documented and periodically reviewed. This documentation might be in the form of a shared Google doc that all COP members have access to or individual teacher journals in which they can capture reflections and action steps.

1. Initial Planning

Evidence Sources

- Documentation of classroom practice (e.g., learning goals presented to students, directions for portfolio selection)
- Individual teacher reflection (e.g., teacher reflection on whether the questions, tasks, and activities are accessible to the range of students' zone of proximal development (ZPD) present within the class)
- Peer observation (e.g., how effectively teachers communicate learning goals to students)
- Student feedback (e.g., survey or interviews)
- Peer feedback (e.g., peer review and feedback on the alignment between the breadth and depth of cognitive complexity and language usage represented by the unit goals)
- Teacher **moderation** to help ensure a common understanding of quality and to calibrate scoring

Step 1: Identify the specific the assessment for the COP discussion

Step 2: Identify which claims across all the propositions are the most critical to address

Step 3: Ensure a common understanding among group members of the propositions, claims and evidence.

Step 4: Review the list of potential evidence sources against the specific evidence described for the targeted proposition.

Step 5: Determine who will present evidence, how and when evidence can be collected, and schedule the Validity Review.

2. Validity Review

Example of Teacher Presentation (based on all claims for Formative Assessment)

- Reflect on your lesson plan and how successful you think you were at articulating lesson-sized goals from the standards. Did they lead to the learning you [the teacher] expected and or not?
- Reflect on one question/task/activity that successfully helped you [the teacher] ascertain the current learning status of individual students' learning in terms of both language and academic content? What was a less successful question/task/activity?
- Reflect on your lesson and how successful you [the teacher] were at integrating your knowledge of the student's language development, conceptual understanding, analytical practices and funds of knowledge into the assessment. Was there one time that was more successful than others? Was there another that was less successful?
- What does this work reveal about the status of student learning relative to goals? Are there other modalities that might be more effective at revealing specific students' learning status? What would be our next steps for each student based on the evidence?
- What judgments were you able to make about student learning based on your evidence? Which judgments do you feel confident about and which are you less confident about?

Step 1: Remind the group of the ground rules for discussion. The focus of the discussion should always be on the assessment and ways in which it can be improved or how strengths in an assessment can be applied to other assessments. The critique should never be of an individual.

Step 2: Invite a group member to present evidence for the focus claims for the assessment and on the sources of evidence identified during the initial planning. While the group member makes the initial presentation, other group members do not interrupt.

Step 3: Once the initial review is complete, group members can ask clarifying questions to help the presenter provide more specific details from a particular source of evidence or to make connections between specific claims and supporting evidence more explicit. This stage of the discussion is focused only on clarification, not evaluation of the evidence.

Step 4: The fourth step is to come to develop a consensus judgment of the validity of evidence for the assessment. Review the set of propositions and summarize the evidence for the targeted claim. Where is the evidence convincing? Where could the assessment or assessment process be improved? Where does the evidence need to be strengthened? How can we do this?

Step 5: The final step is to discuss and document how understandings about the propositions, claims and evidence and/or nature of the assessment could be

applied both to revisions of the specific assessment under review and to future assessments of this format. Whether in a shared Google doc or individual teacher journals documenting reflections and action steps is critical for moving to the Application Review stage.

3. Application Review

Step 1: Reconvene to review what the identified revisions were from the Validity Review meeting.

Step 2: Confirm whether revisions were made to the assessment as planned. For a formative assessment question or probe, this might be just making a note in a lesson plan for the following year about a revision. An end of unit revision that is used by several teachers might require a more substantive revision.

Step 3: Provide an opportunity for COP members to discuss any ways in which the previous Validity Review discussion has impacted their assessment work since the meeting.



© 2021 WestEd. All rights reserved.

Suggested citation: Heritage, M., Wylie, C., Faulkner-Bond, M., & Walqui, A. (2021). *A vision for using an argument-based framework for validity applied to a comprehensive system of assessments for English Learners in secondary grades*. WestEd.



The National Research & Development Center to Improve Education for Secondary English Learners is funded by the Institute of Education Sciences with a dual charge: 1) to identify and describe the systemic barriers that prevent secondary English Learners from successfully accessing the general curriculum, and 2) to develop and test innovative educative curriculum materials that enable secondary English Learners to reach their full potential in community, college, and career.



The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C200008 to WestEd. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

WestEd is a nonpartisan, nonprofit research, development, and service agency that partners with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont, Georgia, and Washington, DC, to Arizona and California, with headquarters in San Francisco. More information about WestEd is available at [WestEd.org](https://www.wested.org).