

# Matching Instructional Uses with Interim Assessment Designs

Brian Gong, Center for Assessment  
October 7, 2019<sup>1</sup>

## Introduction

There is great desire to make assessments more useful and effective in supporting increased student learning and better school and district functioning. To achieve more useful and effective assessments, I explore these statements:

- Assessments must be differentiated to meet diverse uses and situations, particularly supporting better instruction.
- Assessments must be differentiated because different information is needed for different purposes and circumstances; a key aspect of different *claims* and *designs* is whether the instructional path is to *move forward* into new content or to *move back* into content that was previously instructed.
- Assessments may provide information more or less closely related to student content knowledge/skills. For many instructional decisions, content/skills assessed directly are more useful than scale scores or content/skills related to scale scores.
- When educators use assessments to inform instructional and curricular decisions and actions, contextual information is necessary. To inform curricular and instructional decisions, *contextual* information—especially about the curriculum, instruction, and the student’s learning history—must be considered to be able to interpret and use assessment information.
- Results from interim assessment may be combined with other information to inform improvements in curriculum and instruction for the next instructional cycle and systemically. Improving the next instructional cycle for an individual teacher, or the school/district’s curriculum/instructional practices are fundamentally different tasks than improving within-year learning by an individual teacher, and requires different uses of assessments and different other information.
- These points have immediate and significant implications for those who design, use, and evaluate interim assessments.

Some important implications of these assertions for interim assessments are:

1. No single interim assessment can provide all needed information to “inform instruction.”
2. Because instructional actions are diverse from each other and require specific information, evaluating the assessment’s match to the instructional need requires a more specific claim than that the assessment is designed to “inform instruction”. Those designing, using, or evaluating an interim assessment should pay especial attention to the assessment’s claim.

---

<sup>1</sup> The paper was prepared as a background paper for the 2019 Reidy Interactive Lecture Series (RILS) conference on *Improving the Selection, Use and Evaluation of interim Assessments*, September 26-27, 2019, Portsmouth, NH.

3. If the interim assessment does not provide detailed information about the test blueprint—what specific content is included, how, and when—then it is impossible to evaluate whether the test design is adequate to support the claim and intended interpretations and uses. Those designing, using, or evaluating an interim assessment should pay especial attention to this type of test blueprint information because it can be checked quickly and can be understood by educators (as contrasted with more psychometric test information).

The need for specialized, diversified assessments stands in stark contrast with the approaches that have received the most attention and effort recently: to squeeze more informative “subscores” out of summative assessments to help inform instruction for individual students, and to have sets of interim assessments support summative claims.

### Assessments Must Be Differentiated

*Assessments must be differentiated to meet diverse uses and situations, particularly supporting better instruction.*

Diverse uses and situations require different assessments. An influential paper named three broad classes of assessment, with different purposes and characteristics: summative, interim, and evaluative<sup>2</sup>. This division reflected the common wisdom that, “Assessments must be designed to fulfill a purpose; it is difficult to get a single assessment to fulfill multiple purposes well.”

This is more than an aphorism—it is a design truth. For any tool, the more specifically it is designed to do a particular task, the less suitable it is for doing a different task. This is true of interim assessments as well. This is because an assessment is a tool that is designed to collect evidence to support a claim or inference and use. Thus, different claims lead to different assessment designs. For example:

**Claim (summative end of year):** The student has achieved a general level of proficiency over the body of content (knowledge and skills identified in the state’s content standards) at the end of the year.

**Design:** Assessment blueprint includes assessment items representing the *body of content*; assessment performance is interpreted in terms of *levels of proficiency*; evidence is collected near the *end of the year*.

---

<sup>2</sup> Perie, M., Marion, S. and Gong, B. (2009), Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments. *Educational Measurement: Issues and Practice*, 28: 5-13; Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen, CO: Aspen Institute

**Claim (summative end of first marking period):** The student has achieved a general level of proficiency over the body of content (knowledge and skills identified in the state’s content standards) identified for instruction in the first instructional/marketing period.

**Design:** Assessment blueprint includes assessment items representing the *body of content*; assessment performance is interpreted in terms of *overall proficiency or readiness to go on*; evidence is collected near the *end of the instructional/marketing period*.

**Claim (interim end of first instructional unit):** Of the content knowledge and skills (XYZ) instructed in the first unit, the student’s knowledge/skills are strong enough on these (as identified) to be prepared to go on to instruction in the next unit, but are not strong enough on these (as identified).

**Design:** Assessment blueprint identifies knowledge and skills to be *instructed in the first instructional unit*; includes assessment items representing *that body of content*, which is less than the body of content for the whole year; assessment performance is interpreted in terms of *strength and weakness* sufficient to move on in instruction, for *particular knowledge and skills*; evidence is collected near *end of first instructional unit*.

**Note:** The design specifications for this interim assessment include:

- *Relation to past instruction:* “instructed in the first instructional unit”
- *Domain content/skills:* “body of content included in the first instructional unit”
- *Assessment information and Interpretation:* “particular domain knowledge and skills that an individual student is strong or weak on in relationship to the instructional decision of whether to move on in the curriculum”
- *Timing:* “near end of the first instructional unit”

It is evident that the interim assessment designed to provide focused information about what was learned during the first instructional unit cannot provide information to inform the claim for the summative end of year. It also is evident that while the summative end of year assessment *might* cover the *content* of the first instructional unit, the assessment would need to be designed very differently to yield the information specified by the blueprint to inform the claim of the interim assessment for the first instructional unit because the particular design specifications are quite different.

#### [Interim Assessments May Need to be Diversified](#)

Inasmuch as there are different claims for summative and interim assessments, there will need to be different designs for summative and interim assessments. However, there will also need to be different designs for an interim assessment. Interim assessments are typically conceived of as a *set* of assessments—at least two, typically three or four—administered within a year, usually before the end of the year. An important consideration of interim assessments, then, is what the *set* of interim assessments are designed to do, and how the set is designed.

In discussing interim assessments, I illustrate several different possible designs (See Fig. 1, from Gong, 2010<sup>3</sup>). These different designs “differ in terms of the content included in a set of interim assessment measures administered four times during the year, followed by the state summative test. For the purpose of this illustration, assume that the district has established a learning sequence of topics corresponding to the ten topics in the state content standards which I will label A through J, and organized them into an instructional sequence. The ten topics, A through J, are taught one a month. Some of the topics have multiple parts, such as D1, D2, D3, and D4. The state assessment includes the focal “terminal standards” students are expected to know and be able to do to embody proficiency. Note that not everything that is instructed is assessed on the state assessment. This may be because the content is a type of scaffolding or learning progression the student learns first in order to develop the end-of-year proficiency. It may be because certain knowledge and skills cannot be assessed well in the on-demand, end-of-year assessment. Although this example assumed the ten topics were all centered in the state content standards, many curricula include local choices in addition to state content standards.

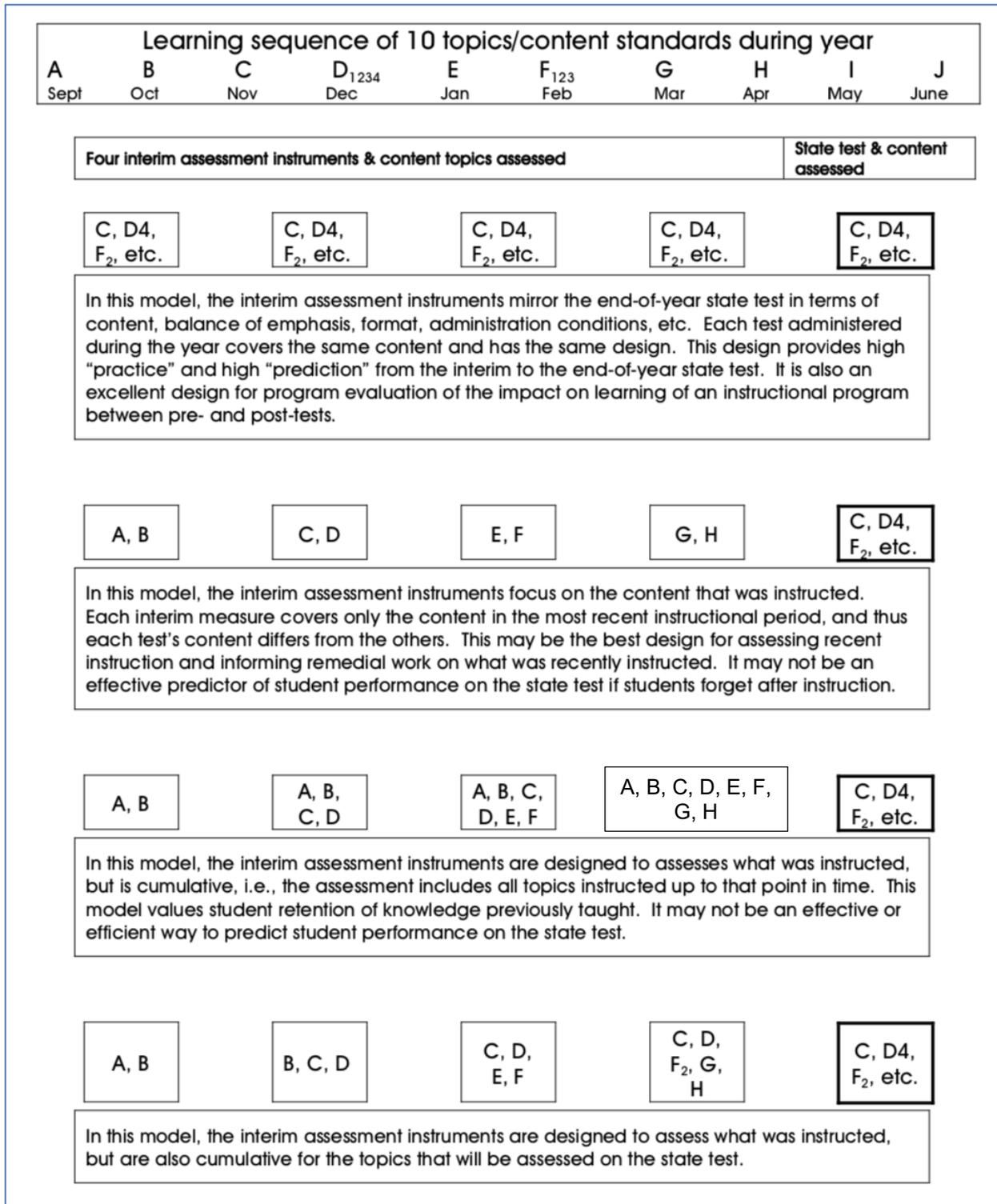
“The four interim instrument designs could all be used to predict how well a student might do on the state test, for example. However, the designs have different assumptions (notably about students’ remembering/forgetting over time) and very different score structures. For example, a score of “50% of the items correct” would have very different interpretations of how well a student was prepared, for the various designs, where 50% would be a high performance in the first design (since the student has not yet been instructed on most of the content included on the test), and a low performance in the second design (where ostensibly the content was all taught recently).

“Note that in every one of these designs, 100% of the test items are aligned to the state’s content standards (one-way alignment between items and standards).” (ibid, p. 15)

---

<sup>3</sup> Gong, B. (2010). Using Balanced Assessment Systems to Improve Student Learning and School Capacity: An Introduction. Washington, DC: CCSSO.

Figure 1: Some possible interim assessment designs (Gong, 2010)



That discussion several years ago focused on “predicting how well a student might do on the state test” at the end of the year. Consider some more contemporary applications:

- Which design would be most appropriate for use in a competency-based education model where a student is expected to demonstrate “competency” before “moving on” to the next unit? (The second design is the most common.).
- Which design, if any, would be appropriate for providing evidence that a student “can apply what has been learned, integrating new knowledge and skills with previously learned ones, in increasingly sophisticated ways”? (The third design.).
- And if the purpose of mastering individual competencies is to be able to put them together into a graduation exhibition at the end of the year? (Design 4).

Because there are different intended interpretations and claims, there should be different assessment designs and specifications. This is true for differentiating between summative and interim assessments in almost all cases, and also often between different interim assessments within a set of interim assessments.

### Assessments Must Be Differentiated Because Different Information Is Needed for Different Purposes and Circumstances

*A key aspect of different assessment claims and designs is whether the instructional path is to move forward into new content or to move back into content that was previously instructed.*

Assessments must be differentiated because different information is needed for different purposes and circumstances; a key aspect of the different information are the *claims* made in relation to the *domain* or student’s content/skills, and whether the instructional path is to *move forward* into new content or to *move back* into content that was previously instructed.

For interim assessments that take place during the year, a key question is whether the teacher is going to move forward with the planned curriculum or not. The curriculum for the year provides a plan for what will be taught and learned, along with a sequence that presumably is because learning in some intentional order facilitates learning all the desired content and skills, and a calendar to reflect the plan that students will have the opportunity to learn all that has been identified. The interim assessment is poised at the point where the teacher and students can ask, “Should we go on? Or is there something that we should spend more time on? And if so, what?”

#### Assessments When the Teacher is Not Sure Whether to Go Forward or Back

Let’s consider when the teacher is not certain whether to move forward in the curriculum or to work with (some) students more on what was previously instructed. There might be many reasons for such uncertainty, such as concern for how the decision will affect her curriculum

plan for the year, or how she might manage with multiple, different instructional plans for different students, etc. However, in the situation where the teacher is not sure whether the student(s) are in fact, ready to go on, then the teacher needs more information. And that information should usually be obtained through assessment. Subsequent assessments should address the reasons for uncertainty—are the results not clear, or are they not credible for some reason? The specific assessment(s) chosen to inform will be specific to these conditions—and likely different from assessments designed to inform under other conditions.

A common example of multiple assessments to inform whether to go forward or back is when successively more focused assessments are used to home in on a particular diagnosis of student strengths and weaknesses. For example, a state summative assessment may provide an overall indication of student proficiency—especially near to when the state summative assessment was administered; a quarterly interim assessment may narrow the content domain; a more focused assessment may provide insight into strengths and weaknesses on particular content and skills that were instructed; and an even more focused assessment may provide insight into not only strengths/weaknesses, but cognitive reasons for the student’s performance. thus helping inform an instructional decision of whether to proceed or not.

#### Assessments When the Teacher Has Decided to Go Back

When the teacher has decided to work with student(s) on material that has been previously instructed, it usually is because there is an indication the student(s) do not know the previously instructed content sufficiently well to be able to learn new material at the same time as addressing the previously instructed content.

In going back to help students do better on material that has been previously instructed, there are two things that are quite different from going forward, and that also tend to call for differentiated assessment needs and therefore different assessment designs:

- It is unlikely that all students will have the same needs. A main need is for the assessment to reveal for individual and groups of students what the strengths and weaknesses are somewhat specifically so instruction can be designed to fit the circumstances.
- The appropriate instruction will most likely be something different than how students were previously instructed, i.e., usually the best instruction will not be to repeat what was done before.

Previously, I have used a medical analogy to portray the relationship between assessment and instruction<sup>4</sup>.

--	--	--	--	--

---

<sup>4</sup> Gong, B. (2015). Evaluating the Quality of Accountability Systems: Beyond Reliability and Validity. ETS Angoff Lecture, Princeton, NJ.

<b>Assessment Diagnosis</b>		<b>Interpretation of implications Prescription</b>		<b>Curriculum and Instruction Treatment</b>
Student's strengths and weaknesses		What student needs		<i>What</i> should be done to address student needs, and <i>How</i> to do it

Assessment provides the information to make a diagnosis of the student's situation—strengths and weaknesses. The diagnosis, through interpretation of the implications of the student's situation in conjunction with other information, leads to a prescription of what the student needs. That prescription must be matched with a treatment in terms of what should be done by way of curriculum and/or instruction to address the student needs, as well as how to deliver the treatment. And finally, the instruction (treatment) has to be carried out in a skillful and appropriate way to have the desired effect of increased student learning.

A challenge is that students' weaknesses and needs may be relatively small or quite extensive. As an illustration, the table below provides one taxonomy of student needs in the instructional support framework described as "intensification." Intensification involves providing more extensive, targeted, and intensive supports, as contrasted with typical remediation, which often involves haphazard repeating of what was done previously.

<b>Example of Diagnosis-Prescription-Treatment "Intensification" Framework</b>		
<b>Diagnosis</b>	<b>Prescription</b>	<b>Instructional Treatment</b>
<b>Situation of Student (level of need)</b>	<b>Support Needed by Student</b>	<b>Instructional Intervention</b>
Keeps up	Regular Instruction	None
Struggles with specific assignments	Extra feedback on work, thinking	Classroom Q&A, partner, teacher's ear
Not bringing enough from earlier lessons each day	Extra support with regular program to integrate previous knowledge with current topics	Homework clinic, tutoring, attention beyond regular class
Misconceptions, gaps disrupt participation	In-depth concentration on troublesome concepts	Sustained instruction with special materials beyond regular class
Several months to a year behind; misconceptions, gaps from months/years	Intensive supplemental instruction on content previous to course	Designed double-period ramp-up course; extended day; summer school
Source: Adapted from Daro, P. In Hill, R. (2010) <a href="https://www.slideserve.com/mele/nationwide-numeracy-news">https://www.slideserve.com/mele/nationwide-numeracy-news</a> For more on "intensification," see also Treisman, U. & Savage, D. (2019). Intensification vs. Support: Equity-		

The diagnosis requires assessments that are as diverse as targeted to particular class assignments and related content, to persistent misconceptions for troublesome concepts within the instructional unit, to connecting previous knowledge with the content currently being learned. Another way to view this is that instructional interventions could be quite varied—from a focused feedback comment by the teacher within one classroom period to intensive supplemental instruction that takes period over months. Just as the instructional actions are quite diverse, the assessments to inform them are also quite diverse.

### Assessments When the Teacher Has Decided to Go Forward

Two particular uses of assessments that vary with instructional intents are discussed in this section:

- Preassessment to learn about student knowledge, skills, and dispositions
- Assessment to match different definitions of “what’s next”

Preassessments receive less attention than remediation, but are perhaps even more important in fitting instruction to student strengths and weaknesses. It seems obvious that when “going back,” a teacher must differentiate instruction to be most effective in promoting student learning because the class of students likely is in different places in terms of understanding, and therefore assessment is needed to inform the teacher about what are the various student needs. It should be equally obvious that it would be useful to the teacher to know about what students know or don’t about upcoming curricular topics in the next instructional unit so that the teacher might incorporate student strengths into instruction and plan how to differentiate instruction to reflect the range of student knowledge and interest prior to starting instruction. Preassessments are oriented to the content domain of the upcoming instructional unit and students’ engagement with that content. Unless planned for formal measurement of pre- to post-instruction growth, preassessments may be less formal and less thorough than an assessment that would be administered after instruction.

But what is the upcoming instructional focus? Educators, when setting learning goals and designing the curriculum know that there may be many different domain aspects chosen. And setting different content/skills to be learned means that the assessments should be designed to reflect those different claims about different domains. Some different classes of “going forward” are described in the table below.

<b>Next instructional focus (what student is asked to do)</b>	<b>Next assessment domain/content focus</b>
“Going forward” learning is focused on a more advanced topic in a sequence of content.	Next content in curriculum sequence

“Going forward” learning is focused on an increase in performance on the same thing, towards more error-free performance.	Same content as current instructional unit (fewer errors)
“Going forward” learning is focused on increasing expertise on the same thing (e.g., ability to apply or analyze due to more powerful mental model, increased fluency, greater independence).	Same content as current instructional unit but tasks and scoring allow more sophisticated solutions, less scaffolding
“Going forward” learning is focused on increasing integration across content and skills, within the same content area or across content areas (e.g., math-science).	Same content as current instructional unit <b>and</b> previous units in the same and other content areas
“Going forward” learning is focused on applying the knowledge and skills learned to problems or projects, often by student choice.	Same content/skills applied

The first “going forward” option—moving on to the next topic in the curriculum—sounds the most logical and is favored by educators using an individualized “move on when ready” type of curriculum. The second is the easiest to implement in terms of instruction: “Go back and work on what you didn’t get right the first time”; it could apply to all students except those who got a perfect score on the final assessment. The third—increased expertise—reflects a view that it is not just whether a student can solve a problem, but also *how* the student solves the problem. In this view, students should be facile and confident, independent and flexible because they have a deep understanding and have automatized much of the routine thinking. Often it takes deliberate instruction and practice to develop this type of expertise, rather than going on to new content. The fourth “going forward” focus involves integration, understanding and making connections within what was learned in the content area (e.g., math) and/or with other content areas (e.g., math-science or ELA-social studies). The fifth instructional choice emphasizes applying what was learned in problems or projects, emphasizing the contextualization of what was learned, and often directed by what the students want to work on.

These are five different curricular directions that might be chosen when “going forward” from what was just instructed are intended as illustrative, not definitive. The main point is that there are several different things that may be done when “going forward” from a current instructional unit, and that assessments to inform or reflect those different choices will also be different. To inform instruction with specific claims, diversified and targeted assessments will be more appropriate rather than a single interim assessment.

### Assessments may provide information more or less closely related to student content knowledge/skills

*For many instructional decisions, content/skills assessed directly are more useful than scale scores or content/skills related to scale scores.*

Assessments differ also in the type of information produced. The information produced is closely related to what can be directly interpreted and done. Three types of information centrally important to interim assessments are:

- student performance reported as *scale scores* (and derivatives),
- student performance, assessed as scale scores but reported as *content knowledge and skills* the student has mastered (and derivatives), and
- student performance that gives direct evidence of student thinking and performance in the content domain(s) of interest.

### *Scale Scores*

Performance reported as a scale score almost always is a number (e.g., between 1 and 36 for the ACT, between 200 and 800 for SAT mathematics). The scale typically represents a construct or latent trait such as “mathematical ability” or “mathematical performance.” A higher score indicates a higher performance, or more ability. A well-developed scale is very powerful and useful. When the test has been carefully constructed and psychometrically formed, the test’s scale has the advantage of being stable and comparable—even with different specific assessment items, with different students, at different times. A scale also allows technical characterizations of the performance, such as how reliable and precise the measurement likely is. However, a critical drawback of scale scores is that they usually are not interpretable in terms of specific content or skills. Thus, it is common to say, “Student A scored 420 on the test, and Student B scored 500 on the test. I know Student B scored higher than Student A, but what does Student B know, or what can student B do that Student A cannot?”

Scale scores alone have limited utility in informing instructional decisions.

### *Content Interpretations Associated with Scale Scores*

Whereas a performance on a scale is almost always a number, it is useful to report that performance in terms of content knowledge and skills. It is most useful to report qualitatively and/or quantitatively different content knowledge and skills for different scale scores, e.g., a score of X is associated with the student knowing certain content knowledge/skills, whereas a higher score of Y is associated with the student knowing additional or more sophisticated or advanced content knowledge and skills. Many achievement tests provide such content-related interpretations, at least for some key scale scores; other achievement tests provide content-related interpretations for every or almost every possible score on the scale. Usually these scale-content interpretations are created through a combination of content analysis of the assessment, statistical analysis, and professional judgment. In large-scale assessments, a very common form of related content interpretations to scores on the scale is *standard setting*, wherein a *performance or achievement level description* is related to a *threshold cutscore*. Most interim assessments have lists of content knowledge/skills related to each scale score. Sometimes the content knowledge/skills are ordered into *learning progressions*, and the developmental sequence is related to increasing scores on the scale.

Some main limitations to content interpretations associated with scale scores are:

- Often the theoretical/conceptual basis for the content organization is lacking and the empirical evidence relating the content interpretation to the scale is weak
- The content/skill grain-size of the assessment score interpretations must match the instructional need. For example, assessment results at the level of an individual content standard might be too specific, just right, or not specific enough, depending on the instructional use.
- The interpretation is almost always probabilistic, not individual—that is, the relationships between content interpretation and scale score is for the “typical student” or “most students” or “what is most likely for most students,” and may or may not be accurate for any individual student
- The assessment administered to the particular student may gather limited evidence to support the content-based interpretation; the user must trust how the scale and related content interpretations were established.

Content/skill interpretations based on scale scores may be useful in some instructional decisions, depending on how well the content/skill relationships to the scale were established.

#### *Content Interpretations Based on Direct Evidence of Student Thinking and Performance in the Content Domain(s) of Interest*

For an assessment to provide more detailed content information about students, three things are needed:

- The assessment must provide evidence by eliciting performance on those content/skills of interest; there must be items or other opportunities to observe student performance aligned to the content/skills
- The assessment must provide scoring or other guidance on how to interpret the performance. More content-based scoring guidance will yield more content-based interpretations. For example, “right/wrong” scoring loses content-specific information unless very closely related to items with specific content/skills.
- The more the assessment elicits and creates a record of student thinking and procedures related to specific content, the more specific the evidence to inform interpretations about what the student knows/can do, and why. While some multiple-choice format assessment items may elicit complex student thinking, there is no visible record of student thinking. An advantage of performance tasks and other item formats that require students to “show their work” and especially to document their reasoning provides more specific evidence to inform content-specific interpretations.

Many “go back” instructional decisions would benefit from this type of content-specific assessment information, and does not require a scale.

## When Educators Use Assessments to Inform Instructional and Curricular Decisions and Actions, Contextual Information Is Necessary

*To inform curricular and instructional decisions, contextual information—especially about the curriculum, instruction, and the student’s learning history—must be considered to be able to interpret and use assessment information.*

Information in addition to that provided by an assessment must be considered when deciding what to do instructionally. As discussed in the section above, assessment functions to inform a diagnosis, while the treatment requires knowledge at least about what might be helpful instructionally, what the student’s learning history is, and what the teacher’s instructional repertoire is. For example, an assessment might accurately portray the student as knowing X and having an incomplete understanding of Y, but the instructional plan would be very different if the assessment were done before instruction versus after instruction. The contextual information of what had been instructed and how is essential in thinking of future instruction. Another example is that a teacher usually has an “instructional repertoire”—a set of things the teacher knows when and how to apply and feels confident in using in supporting student learning. Teachers’ available repertoires are shaped by their circumstances—a teacher might know an effective instructional move or strategy, but not be able to employ it because of classroom organization, time, budget, or technology support constraints. Teachers’ instructional repertoires differ, and so assessment ideally would be matched with repertoires. It is essential that, just as assessment should be sensitive to instruction, instruction should be sensitive to assessment. (Gong et al., 1992)

## Results from Interim Assessment May Be Combined with Other Information to Inform Improvements in Curriculum and Instruction for the Next Instructional Cycle and Systemically

*Improving the next instructional cycle for an individual teacher, or the school/district’s curriculum/instructional practices are fundamentally different tasks than improving within-year learning by an individual teacher, and requires different uses of assessments and different other information.*

While interim assessments may be used to inform instructional decisions when going forward or back within the year, another powerful use is to inform improvements in curriculum and instruction for the next instructional cycle, i.e., for the next group of students. When educators consistently seek to improve their curriculum and instructional plans over time, for every instructional cycle—year to year, section by section—then the improvement can become systemic and cumulative. Contrast the situation where faulty curriculum or instructional plans result regularly in deficits in student learning, but that is paired with very effective remedial instruction. Focusing on the remedial instruction—“going back”—may help each group of students but does not address the root causes of the learning problems. Using information

from assessments to improve the curriculum and instruction has a chance of improving learning on the first pass and eliminating or reducing the need for remediation.

There are other program evaluation uses where assessments may help inform program improvement by monitoring curriculum implementation, coherence, or effectiveness. For example, a district might use interim assessments to evaluate whether there are curricular differences in student knowledge between students coming from two different elementary schools into a single middle school. A department might use a common interim assessment (e.g., unit final exam) to inform reflections on what curriculum and instructional methods were more effective across teachers in the department teaching the same course.

Assessment information alone is insufficient to inform this type of curricular and/or instructional improvement. Assessment results must be combined with conscious awareness of what was done in a systematic enough way to associate variations in curriculum and instruction with improvements in learning over time. Essentially, assessment information must be incorporated into program evaluation, and combined with creative curriculum and instructional design, as well as reflective practice.

### Implications for Designing and Evaluating Assessments

*If one accepts the argument that assessment uses—particularly instructional uses—require different assessment designs, then there are immediate and significant implications for those who design, use, and evaluate interim assessments.*

In this paper I have argued and tried to illustrate how assessments typically yield only the information which they were designed to give. It requires careful work to design and develop an assessment that provides accurate and useful information. It is also true that before adopting an assessment, the assessment should be evaluated for its match to the desired interpretation and use, as well as for its technical quality.

In this paper I have emphasized the assessment's *claim* and the *design*, particularly the content *domain* addressed in the test's blueprints and how those may be different over *time* and the multiple administrations of an interim assessment.

- Assessments are most powerful and useful when designed intentionally for particular purposes and uses. When there are substantially different purposes and uses, then it usually requires different assessment designs.
- Interim assessments are typically administered multiple times over a year. The set of interim assessments should be considered as a whole, as well as individually, so that the set achieves the overall purpose, e.g., predicting end of the year performance, informing a judgment about student competency, providing feedback about what was just taught, etc.

- Different instructional uses—such as deciding how to “move forward” or “move backward” in instruction—usually require different assessment designs.
- Improving “next-cycle instruction” requires very different information than assessments intended to inform instruction with the same students and same curriculum.

Three main implications of this argument are:

1. No single interim assessment can provide all needed information to “inform instruction.” **Users should identify their specific “go forward/go back” strategies and match their instructional needs to the interim assessment’s specific claims.**
2. Because instructional actions are diverse from each other and require specific information, evaluating the assessment’s match to the instructional need requires a more specific claim. Those designing, using, or evaluating an interim assessment should pay especial attention to the assessment’s claim and information provided. The assessment’s claim and information provided should match the intended use and need of the user. **If the assessment’s claim(s) are too broad, general, or include multiple different uses, then the user should very carefully examine the evidence that the assessment can support its claim(s).**
3. If the interim assessment program provides detailed information about the test blueprint—what specific content is included, how, and when—then it is possible to evaluate whether the test design is adequate to support the claim and intended interpretations and uses. Those designing, using, or evaluating an interim assessment should pay especial attention to this type of test blueprint information. **If the assessment’s documentation does not include this detailed information, then the user should be skeptical about whether the assessment in fact can support its claim(s).**

Classroom assessment—particularly interim assessments—would benefit from more specific claims, use case scenarios, and technical documentation, both to inform design and to permit evaluation. In particular, elaborating interpretive and validation arguments may be very fruitful, as is customary for large-scale assessments, although the evidence may be different. Let us hope that educators, test vendors, and researchers will together be able to provide such information about the many diverse assessments needed to support improved classroom teaching and student learning.