

Empowering personalized instruction with a three-tiered approach to learning evidence

By Laine Bradshaw, PhD

Vice President of Classroom Solutions, Pearson

Millions of key instructional decisions are made each school day, but what evidence of learning is guiding those decisions? What impact would stronger learning evidence that steers better instructional decision-making have on a learner's life, a school system, and our country's educational system?

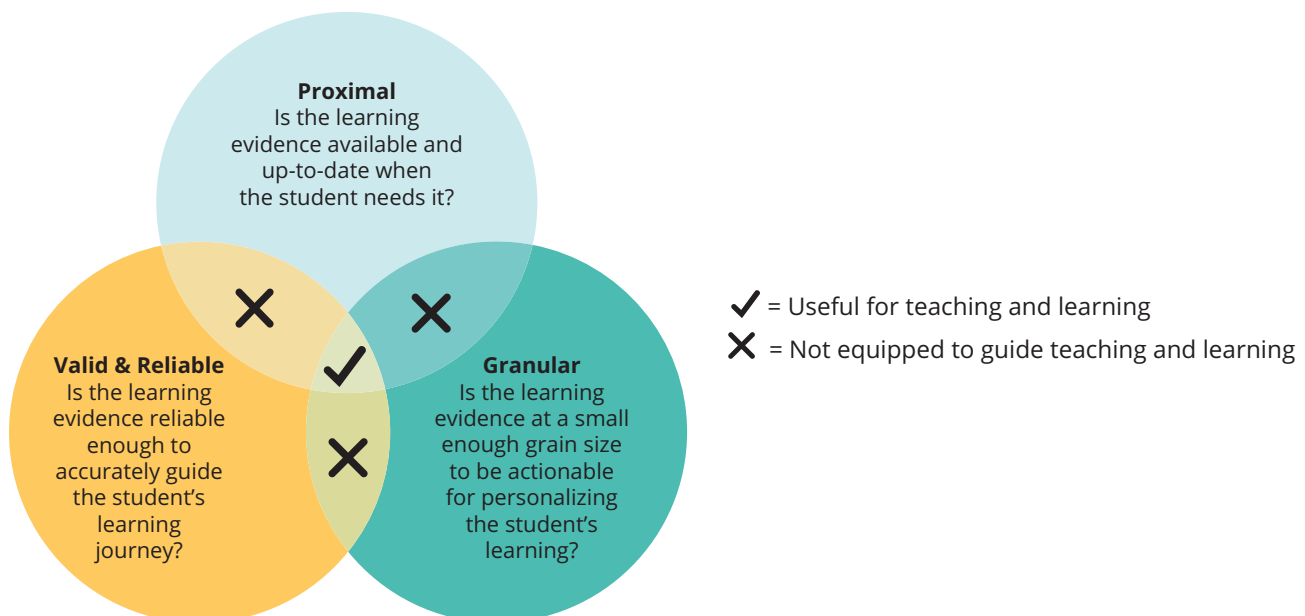
The promise of accelerated learning through personalized or differentiated instruction that is tailored to a learner's needs rests on the insight, accuracy, and availability of learning evidence. Addressing those needs starts with identifying them through assessment. Results from assessments provide learning evidence.

This guide identifies three features of learning evidence that assessments must provide to be effective tools for day-to-day teaching and learning in the in-person or remote classroom.

We share our views from the perspective of building the Navy assessment system in collaboration with state and district school leaders, educators, researchers, psychometricians, and technologists to have the ability to provide learning evidence with these three key features: granularity, reliability, and proximity. While any one of these features is not unique to Navy, the combination of the three is.

Navy is a unique classroom assessment system designed to empower personalized learning by providing granular, reliable, and proximal learning evidence for students and those who are helping them learn.

Learning Evidence Analysis



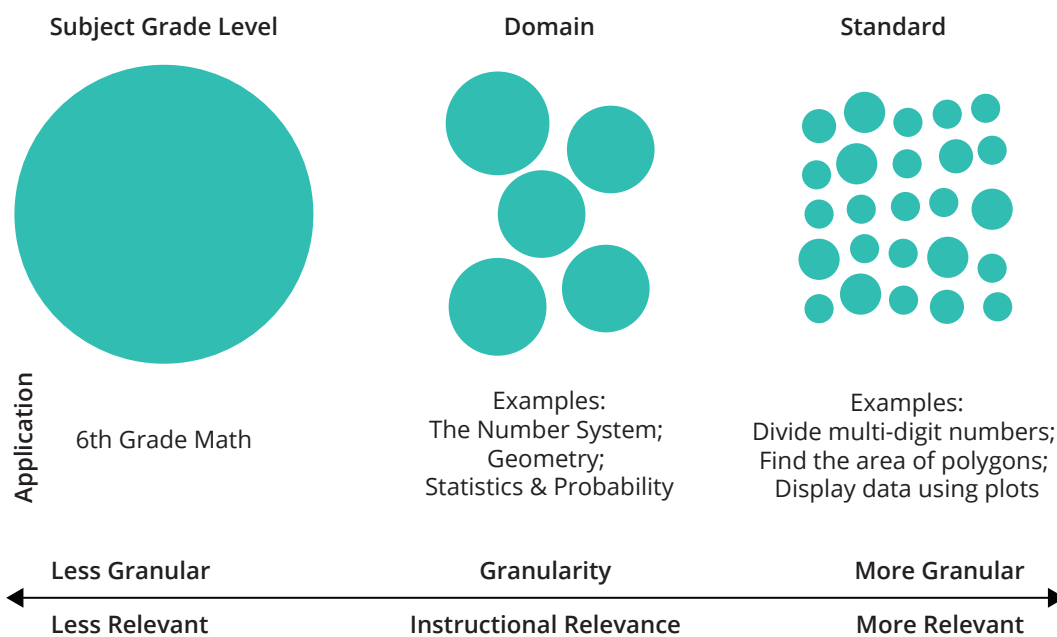
1. Granularity

To be effective for teaching and learning purposes, the granularity of the learning evidence provided by assessments must be small enough to guide instructional decisions.

Let's look at the granularity of learning evidence provided by the three primary types of assessments that schools typically use: end-of-year summative, interim (typically three times per year), and classroom (on-demand throughout the year).

Summative and interim assessments support scores for the overall subject at a grade level (e.g., a 6th grade math score) and for domains (e.g., Geometry, Statistics & Probability scores). Classroom assessments aim to measure more fine-grained learning targets, such as the academic standard (e.g., find the area of polygons, display data using plots) or smaller.

Granularity of Learning Evidence



Grade-level and domain-level granularity is too coarse to support decision-making for day-to-day or week-to-week instruction. Assessments designed to give these larger pieces of information are well suited for informing macro-level accountability and educational programming decisions, but they are not meant to be used as detailed guides for an individual student's learning. Consider a student result of being 'low' in Algebra or being in the 42nd percentile in math. A domain involves weeks or months of learning material, and a grade level is a year's worth. Learning evidence that refers to the volume of material found in a domain or grade level is too large of a focus for guiding the design of *targeted* support and instruction.

More granular learning evidence is needed to be *actionable* for instruction, and the smaller grain size of classroom assessments provides this actionability. For example, the overall Geometry domain is too broad to direct the next instructional activities, but at the standards level, a teacher can design a lesson or activity to help a student understand lines of symmetry for a figure.

The smaller granularity of evidence provided by classroom assessments adds value to the learning process because it provides a level of insight that pinpoints a student's learning needs and helps to guide next steps.

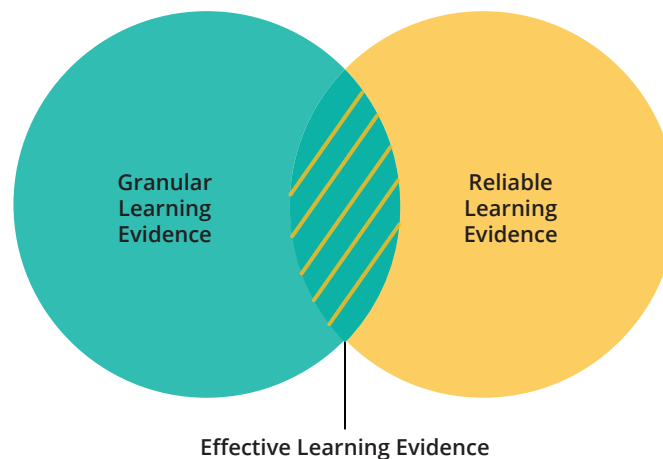
2. Validity and reliability at the right granularity

To be effective for teaching and learning, the smaller grain size of learning evidence must also be valid and reliable.

The effectiveness of targeting a learner's needs is lost if the evidence is not accurate. Unreliable granular learning evidence can easily misguide the way instructional resources and supports are provided for students. For example, providing additional lessons on figure symmetry for a student who already understands those concepts takes time away from focusing instruction on the areas in which the student does need support. Inaccurate granular learning evidence may evoke an action, but not an action that accelerates student learning or makes good use of time and resources. Thus, effective assessment for teaching and learning requires both reliability and specificity.

Right Learning Evidence

Effective learning evidence is evidence that has both granularity and reliability to accurately pinpoint student needs for personalized student learning.



Where can we find reliable granular learning evidence?

Summative and interim designs do not yield reliable scores for actionable grain sizes of learning evidence at the standards-level or smaller. Though summative and interim assessments may report standards-level information, that information is not a reliable score or reporting category. Although this point is widely agreed upon by psychometric professionals, it may not be clear to educators or learners who are using the results because the learner's report may not indicate the reliability difference in the grade-level, domain-level, and standards-level information.

Assessments that produce results at multiple grain sizes may have general claims of being "valid and reliable", but it is important to ask what that validity and reliability refer to. "Valid and reliable" may mean that grade- and domain-level results are supported by sufficient validity and reliability evidence, while the standards-level results are not. Some interim assessments even recommend standards-level learning paths based on unreliable standards-level information, which results in learners spending valuable time getting instruction that isn't aligned to their needs.

It is important to understand the reliability of the learning evidence given on a score report or used to determine a personalized learning path. *State, district, and school leaders can ask assessment providers how reliable the specific results, at standards-level or smaller grain sizes, are on a given report.*

Obtaining learning evidence that is both granular and *reliable* is a challenge. Typical classroom assessments that focus on providing granular evidence are not also focused on providing reliable granular evidence. Classroom assessments typically utilize a small number of items—written by educators or provided by licensed item banks—per granular learning target, such as a standard or a skill. The number, or percentage, of items a student answers correctly is used to produce a subscore for the granular learning target. Then, a threshold, or cut score, is set on the subscore to describe the result (e.g., 70% or above is proficiency in the skill, 90% or above is mastery of the standard). Psychometric professionals also widely agree that these approaches using subscores based on a small number of items do not provide reliable learning evidence.

To fill the gap in assessments that provide both reliable and granular information for teaching and learning, we created Navy. Navy is a diagnostic classroom assessment system that utilizes principled assessment design practices and psychometrics to ensure technical merit like summative and interim assessments do, but utilizes a small number of items to produce granular results like typical classroom assessments do. Navy achieves this combination by using a new kind of *diagnostic psychometrics* to classify students into learning groups (competency vs. non-competency) on multiple learning targets (academic standards), rather than using scaling-based psychometrics to determine an overall score or percentile rank for a general area of study (domain, grade level). Leveraging these diagnostic methods, learning group classifications can be made reliably with short assessments. Providing trustworthy granular learning evidence without lengthy assessments is key for practical use as a grade level often contains a large number of learning targets.

Coordinating both the psychometric technique and the assessment design with the desired use of the assessment results is a key step in gaining reliable results at the desired grain size. It is helpful to start the design process by asking: What results do we want to use from this assessment? What results do we need to be able to trust to successfully guide our next steps? Then, the assessment design and psychometric methods can be built to support those intended uses and fulfill those needs.

This chart depicts reliable granularity for different assessment types:

Granularity of Reliable Reporting Categories

Granularity	Summative	Interim	Typical Classroom	Navy Classroom
Grade level	●	●		
Domain	●	●		
Standard				●

● = Reliable reporting category for the assessment type

What about validity?

Reliability is a key part of validity; the use of an assessment can't be valid without the result being reliable. However, there is more to validity than reliability. To be effective for teaching and learning, granular learning evidence needs to be supported by a strong validity argument that addresses the entire assessment enterprise. This includes the design process, item construction, blueprint and form creation, psychometric performance evaluation, outcome examination, and consequence considerations.

A key to validity of fine-grained learning evidence is that the assessment design process happens at that fine-grained level. *More detailed learning evidence necessitates a more detailed assessment development process.*

Three important questions expose the validity of the content for assessments measuring more fine-grained learning targets:

Question 1: Do the items on the assessment elicit the right content, in the right proportions?

Are the important parts of the learning target measured? Do the parts that are measured adequately represent the breadth of the learning target? The answers to these questions should be "yes" in a strong validity argument. Are some parts over- or under-represented? A "yes" to this question weakens validity.

Question 2. Do the items have the right rigor?

Does the depth of knowledge measured by the items sufficiently reflect the knowledge levels required for the learning target? When strong validity evidence is present, the answer to this question is "yes" and indicates the right rigor is present to meet the expectations of the learning target.

Question 3. Do the items measure irrelevant factors?

Does answering the items correctly require knowledge, skills, and abilities outside of the learning target? A "yes" to this question threatens validity, and in turn, can threaten fairness.

Unlike reliability that can be quantified by a metric with target thresholds, acceptable validity is supported by a collection of evidence. Significant individual threats to validity can render overall validity unacceptable.

3. Proximity through flexibility, immediacy, and "refreshability"

When using learning evidence to inform good instructional decision making, the evidence must be available at decision-making time. To have evidence proximal to learning, the assessment system needs the right flexibility, immediacy, and "refreshability".

Flexibility of administration

To gather proximal learning evidence, assessment must happen during the learning process and be administered at the right time on the student's learning journey. *As-needed assessments are required to provide as-needed learning evidence.*

Schools pace and sequence learning in different ways. Within a school, classrooms may have variations in their pace and sequencing, and within a class, different students learn at different paces and on different paths. *To support learning for all students, classes, and schools, classroom*

assessments need to have flexible administration that allows for learning targets to be assessed in any combination — and at any time. To this end, Navvy is designed as a modular, on-demand system to provide the flexibility needed to personalize assessment.

Immediacy of results

While administering assessments proximal to learning is an important part of the equation for timely learning evidence, getting the results proximal to learning is another. Results are needed in real-time so that educators, students, and families can immediately act upon them to guide next steps for learning.

“Refreshability” of learning evidence

Like a watch keeping time as it passes, a key part of assessments that support personalized learning is the ability to stay up to date with the student’s learning as it changes. Personalized learning isn’t an event; it’s a process during which a student is provided instruction, their understanding is assessed, their needs are identified, additional instruction is given according to their needs, and their understanding is assessed again, with the iterative process of providing targeted supports and checking for understanding continuing as needed to help the student learn. An assessment system that supports teaching and learning enables this process by providing multiple valid and reliable assessments over the same granular learning target where results stay up-to-date, or fresh. *The ability to refresh the learning evidence for a student is key to personalizing learning.*

Applying Learning Evidence

Learning evidence operates in a larger system of assessment and an even broader system of teaching and learning. For learning evidence to be effective, it must be effectively applied within these ecosystems.

For Navvy, we have some core beliefs and practices that are central to our system design and theory of actionable learning evidence. We begin from a foundation of celebrating each learner as an individual on their personal learning journey. Learning doesn’t always move in a straight line or at an equal pace, and that’s okay. Each student has their own path and will move at their own pace. We believe each journey requires personalized guidance, and good guidance makes the journey more successful. Good guidance requires more than being granular, reliable, and proximal; it also needs to be delivered and used in healthy ways. We believe learning should be heartily celebrated along the journey, guidance should be encouraging and promote positive learning mindsets, and learning evidence should—in addition to bettering instructional decisions made by educators—empower students with ownership of their learning.

Visit **navvyeducation.com** to learn more about how Pearson can partner with your school or district.

800-627-7271 | PearsonAssessments.com

© 2022 Pearson Education, Inc. or its affiliates. All rights reserved. Pearson and Navvy are trademarks, in the US and/or other countries, of Pearson plc. SCHAS3109 EL 09/22

