



## SUMMARY: VALIDATING APPROACHES FOR NON-TESTED GRADES AND SUBJECTS

### BACKGROUND

Over the past year Education Analytics (EA) partnered with the Los Angeles Unified School District (LAUSD) to conduct the *Validating Approaches for Non-Tested Grades and Subjects* (NTGS) study.<sup>1</sup> The research project's focus on supporting teacher-developed assessments was born from a shared interest in supporting teacher assessment literacy at LAUSD and responding to the increasing national need for a scalable – yet technically rigorous – assessment development process that prioritizes teacher involvement.

The major objectives of this work are to 1) develop high-quality assessments across non-tested grades and subjects to support reliable and valid Student Learning Objectives (known in LAUSD as Data-Based Objectives or “DBOs”); 2) facilitate assessment literacy learning that can be applied to classroom instruction and other teacher-developed assessments; and 3) further a shared understanding of subject standards, curriculum, and instructional goals among teacher participants.



### SUMMARY OF RESULTS

First, the accomplishment of the pilot's leading objective – facilitating teacher-developed NTGS assessments that can support truly data-driven Student Learning Objectives – is evidenced by both the successful implementation of a rigorous assessment development cycle and the end product of vetted, operational assessments. Psychometric analysis demonstrated that the operational assessments have high reliability and that the items are able to accurately differentiate between students with varying academic ability levels.<sup>2</sup> See page 2 for a brief explanation of these psychometric concepts.

<sup>1</sup> This study was supported with funding from the Gates Foundation.

<sup>2</sup> The Cronbach's  $\alpha$  of the three operational assessments administered in June averaged 0.83. The item analysis also exhibited validity evidence supporting the interpretation and use of these test scores. The item difficulty indexes are in appropriate ranges that match with students' ability levels: most items have relatively large point-biserial correlation coefficients (also known as "item discrimination indexes".)



### WHAT IS TEST RELIABILITY?

Generally speaking, **test reliability** reflects the overall quality of a test. It refers to “the consistency of scores across replications of a testing procedure” (AERA, APA, & NCME, 2014, p.33). There are several types of test reliability coefficients available; for this analysis, we reported the reliability coefficients of classical test theory, (**Cronbach’s  $\alpha$** ). **Cronbach’s  $\alpha$**  ranges between 0 and 1. The closer it is to 1, the more reliable the test is and the less measurement error the test contains.

### WHAT IS ITEM DISCRIMINATION?

According to classical test theory, **item discrimination** is the relationship between student performance on a single assessment item and the criterion score. Therefore, the item discrimination index is also referred to as **point-biserial correlation**.

Theoretically, the item discrimination index ranges from -1 to +1. The larger the index is, the better the item differentiates test takers with different ability levels. In practice, an important criterion of the discrimination index is that it is not negative nor too close to zero. Negative **point-biserial correlation** is often an indicator of a mis-keyed item. Such items require careful review.

Second, the project has demonstrated its value as a source of meaningful professional development. Teacher feedback and discussions during the assessment development trainings indicate that the knowledge gained around test design and data-informed decision-making has enhanced teachers’ assessment literacy and the application of the new knowledge to classroom instruction. One Dance instructor who teachers at multiple LAUSD schools said:

*“The trainings and writing of the assessments assisted me in further clarifying 5th grade dance curriculum and standards.”*

On an end-of-year survey, all of the teacher participants reported that the trainings yielded a significant and positive impact on their instructional planning.

Finally, the pilot produced what could actually be the most significant benefit: meaningful and collaborative conversations between teachers about standards, curriculum, and instruction. Teachers indicated that the collective process of developing assessments allowed them to foster a shared understanding of their teaching practices. Another LAUSD teacher reported:

*“Working on assessment questions as a theatre group brought consistency and alignment to the LAUSD theatre and dance teachers. This opportunity was so enlightening for me.”*

### NEXT STEPS

The success of the LAUSD pilot suggests several next steps, including future assessment development across new subjects and grade levels and expansion to other assessment types, such as constructed response and performance tasks. Findings from the first year of the process also underscore the need to develop and scale comprehensive and meaningful professional development that enables teachers to make connections between assessment data, learning standards, and professional practice.

For more information about partnering with Education Analytics around assessment literacy and professional development work, please contact **Selina Eadie** at [seadie@edanalytics.org](mailto:seadie@edanalytics.org)