


GUIDE

A practical guide for considering and interpreting test score differences among student subgroups



Table of contents

- 3** Motivating scenario
 - 3** Introduction
 - 4** In what situations might leaders find this information useful?
 - 5** What do test scores measure?
 - 7** How can we understand subgroups and characteristics across students?
 - 9** Toward a practical understanding
 - 10** Actionable recommendations
 - 11** Conclusion
 - 12** References
- 

Motivating scenario

Principal Cortez sits down at his desk with his morning coffee to review this year’s state summative assessment results for his school, hopeful to see good news that his students have made progress in response to all the hard work of his teachers. After looking at the top-line numbers, he then reviews average scores through several lenses—such as gender, income, race/ethnicity, and language background—to understand how different students experienced instruction. He engages in this work to help identify which students might require more attention or to identify for whom his school’s efforts are working. But he still struggles with the best way to consume, think about, and act on these subgroup score differences.

Introduction

Principals and district leaders often analyze test score differences across student groups. This can be a sensitive topic requiring a nuanced understanding of what test scores actually measure (both intentionally and unintentionally), how students may belong to multiple subgroups simultaneously, how systemic issues impact students, and how different audiences might interpret differences in test scores. These understandings influence how schools, districts, and states act on assessment data.

This guidance document provides insights into how school and district leaders might consider and interpret test score differences across subgroups. In the following sections, we provide:

- Examples of relevant situations
- An overview of what test scores measure
- A discussion of how to analyze data on student subgroups and overlapping characteristics
- Practical considerations for using [MAP® Growth™](#) assessment data
- Four actionable recommendations for districts and schools

The purpose of this document is not to tell leaders whether they should focus on score differences between different groups or how they should narrow them. Instead, it is meant to offer recommendations for how to understand what differences may exist and give leaders some structure for how they might consider and interpret those differences. Such guidance is essential as many schools, districts, and charter authorizers set annual goals, or if your school or district receives points on state school report cards based on narrowing test score differences across subgroups.

In what situations might leaders find this information useful?

There are many situations in which leaders might find themselves wishing they had a better understanding of average test score differences across student subgroups. As a thought exercise to show how this information might be broadly used, we present three example situations where this information might prove useful.

Example 1

A principal sees that MAP Growth scores indicate that Hispanic students are, on average, scoring lower in math than their peers. In her school, many Hispanic students face systemic barriers, such as limited access to economic resources or instruction in their home language. She examines multiple dimensions—gender, income, race/ethnicity, and language—to understand what opportunity gaps in access to grade-level instruction, instructional supports, and learning resources may be contributing to these outcomes.

Example 2

A district is struggling with understanding how to effectively implement early literacy tutoring and does not think targeting students based solely on income level, without considering overlapping systemic factors such as language background, disability services, or inequitable access to high-quality instruction, is working. The district would like to take a deeper dive into its data to understand how student reading test scores vary not only by income level but also by other relevant subgroup categories such as multilingual learner status, students with disabilities, and race/ethnicity. By disaggregating the data across multiple subgroups, district leaders aim to identify patterns that can inform more targeted approaches to tutoring and prioritize the students most in need of additional support.

Example 3

A parent group reaches out to a school board member with concerns over score differences that appear to align with racial/ethnic groups. The group has not yet considered overlapping systemic factors—such as language access, opportunity to learn, or school-level inequities—that may contribute to these patterns. The group wants to know if such score differences could indicate a flaw in the test and what the district has done to ensure the test is measuring the content fairly for all students. The school board member plans to respond by acknowledging that assessments carry some degree of bias, while also emphasizing that score differences may reflect broader inequities in students' opportunities to learn.

What do test scores measure?

Leaders need to have a common understanding of what test scores measure before making data-driven decisions. Tests are designed with specific measurement goals that shape the data and insights they provide. This section explores the following:

- What tests are designed to measure
- How students' test scores can be influenced by things other than what they are supposed to measure
- The reality that not all assessed knowledge is gained in the classroom

First, tests of student achievement aim to measure the stock of accumulated knowledge of a student on a given topic or in a given subject (Jacob & Rothstein, 2016). Tests can vary in design and measurement when used for accountability or progress monitoring purposes. High-stakes summative assessments are typically longer tests designed to measure grade-level knowledge related to state instructional standards that a student has supposedly learned over an entire school year. In contrast, interim assessments are shorter tests that measure a broader range of student knowledge at multiple time points in the school year (Kuhfeld & Lewis, 2024). In the case of MAP Growth, the reading and math tests are adaptive and present each student with a unique set of questions to measure competencies above, at, or below grade-level in the fall, winter, and/or spring ([Rodríguez, 2026](#)).

Any use of assessment data must begin with an understanding of what the assessment was designed to measure. Just as one cannot draw valid inferences about height from a scale, schools should not try to make instructional decisions based on assessments that don't measure student learning of instructional content.

Second, students' test scores are often influenced by things other than skill or content mastery. What a student can accomplish on a particular day of testing is a factor of their current and prior educational inputs, as well as extraneous factors to education and testing such as student health, economic security, living conditions, and out-of-school learning experiences ([Ladson-Billings, 2006](#); [Downey & Condrón, 2016](#)).

Test scores can serve as indicators of the lived educational opportunities that students have encountered throughout their academic journey. As a student progresses through their education, they are exposed to various resources, teaching methods, and learning environments. These factors collectively contribute to the development of the skills and knowledge assessed by standardized tests. For example, students with greater access to advanced coursework, experienced educators, and enrichment opportunities—access that is unevenly distributed due to systemic inequities—tend to demonstrate higher test performance. Differences in lived educational experiences—and the systemic inequities that shape them—underscore the need to interpret test scores within the broader context of educational equity and access. Lower scores can just as easily represent fewer past learning opportunities as they can ineffective instruction.

Third, the skills measured by a test are not solely developed within classrooms or formal educational settings. Students acquire knowledge and skills through a variety of experiences, including family interactions, community involvement, and extracurricular activities ([Downey et al., 2004](#)). These informal learning environments can significantly influence a student's academic performance. For instance, a student who engages in reading and problem-solving activities at home may develop stronger literacy and analytical skills, which can positively impact their test scores.

Importantly, test scores do not measure anything inherent to who students are. Test scores measure what a student knows at a specific moment in time and not a student's innate ability ([Jacob & Rothstein, 2016](#)). There is no such thing as a "pure" measure of student ability that is not influenced by educational opportunity ([Corno et al., 2001](#)). Standardized tests are designed to evaluate specific academic skills and knowledge, but they cannot capture the full complexity of a student's abilities, potential, or character.

When analyzing test scores, it is crucial to recognize that subgroup differences are often a reflection of varying opportunities to develop the skills being measured. Students from different socioeconomic backgrounds and geographic locations may have unequal access to quality education and resources. Additionally, some racial and ethnic groups may experience limited access to education and resources due to inequitable systems and social biases. This inequality can manifest in significant differences in test performance among subgroups. For example, research has shown that students from low-income families often face greater challenges in accessing high-quality educational opportunities, leading to lower test scores than students who have had such opportunities ([Reardon, 2013](#)). Addressing these opportunity differences requires targeted interventions to ensure that all students have access to the resources and support they need to succeed.



How can we understand subgroups and characteristics across students?

Examining student-level data can be tricky. Although students ultimately get translated into binary data points in datasets as members of a specific subgroup (or not), we intuitively know that's a simplistic view. Students' subgroup membership is not limited to a single characteristic. Thus, examining subgroup data is complex because aggregate data about test score growth, for instance, only tells us about the association between growth and that singular subgroup characteristic. It ignores that the students within those subgroups simultaneously occupy multiple positions within and across subgroups. In this section, we first discuss this important issue from the perspective of individual students and then from the perspective of variation within subgroup categories.

All students belong to multiple different subgroups, making their experiences and identities uniquely complex. For example, students may simultaneously navigate racial identity, income level, gender identity, disability status, and language background—all of which shape their worldview, educational experiences, social interactions, and achievement. This often means that students navigate a dynamic and sometimes challenging landscape, where they must reconcile multiple aspects of their identity, face diverse expectations, and cope with varying degrees of privilege and marginalization. Understanding this complexity is crucial for leaders to create supportive educational environments that cater to the diverse needs and perspectives of all students.

Here are two examples that highlight how student characteristics can be seen across subgroups:

Example 2.1

A multiracial student, with parents from two different countries, may encounter conflicting cultural expectations or assumptions from peers and adults due to how schools and communities interpret their identity. These dynamics—rooted in societal and institutional norms—can shape a student's sense of belonging and access to supportive learning environments and can ultimately impact their academic performance and test scores.

Example 2.2

A student who uses a wheelchair and comes from a low-income family might encounter accessibility issues and a lack of appropriate accommodations in their school. These challenges can result in decreased opportunities for academic engagement and success, ultimately affecting their test scores.

Additionally, when examining subgroup categories, it is important to consider that there is sizable variation within each category. Although a singular category—such as students eligible for free- or reduced-price lunch (FRL) or multilingual learners—may seem straightforward, that category is comprised of students who occupy a spectrum of values within that category.

Here are two examples that highlight variation within subgroup categories:

Example 3.1

FRL eligible students: Although for data purposes we often combine these two related categories, that choice creates a category that captures a wide range of family incomes. For example, the federal guidelines for 2024–2025 indicate that, for most states, annual household incomes could range from < \$26,572 (for a single-parent household with only one child) to \$97,532 (for a two-parent household with six children). Additionally, the actual adjusted purchasing power of these incomes might vary significantly within states, across districts, and within districts due to differences in cost of living (e.g., housing, transportation, food, and utilities).

Example 3.2

Multilingual learners: Students who are classified as multilingual learners do not all share the same levels of English fluency. Some will have zero to very limited experience with the English language, while others will have greater proficiency nearing the top boundary of this classification. This is only one type of variation we might find within a singular category. Another example is that these same students do not all share a native language. Some may be native Spanish speakers and others native Mandarin Chinese speakers. The speed and proficiency at which students may graduate out of receiving English as a second language (ESL) services will depend on a multitude of factors including, but not limited to, their native language.



Toward a practical understanding

Given this information about students and subgroups, leaders may need context for the data they look at. When examining aggregate test scores and growth, they should consider how one subgroup might be divided into ways that account for other important subgroup characteristics. Here are two examples using hypothetical MAP Growth data:

Example 4.1

In a large urban district, 52% of Black students but only 21% of white students are classified as FRL eligible. When examining aggregate numbers, Black students' test scores and growth in both math and reading are lower than white students' by a significant margin. However, when comparing Black students who are FRL eligible to white students who are similarly eligible, the gaps close for test scores and for growth in both math and reading. In other words, when disaggregated by both race and income, most of the overall Black-white difference reflects the disproportionately higher share of white students with access to economic advantages. These structural advantages raise aggregate white averages—not inherent group differences.

Takeaway from Example 4.1: A district leader might be tempted to focus additional resources on Black students to reduce outcome differences in this district. However, in this case, those resources would be better targeted specifically toward Black students who are also classified as FRL eligible or simply to all FRL eligible students. In that case, the students with the lowest test scores and growth would receive additional needed attention.

Example 4.2

A small suburban school has a sizeable population of Asian, Hispanic, and white students. When examining aggregate numbers, Hispanic students perform similar to white students in terms of test scores and growth in both math and reading. However, Asian students lag behind white and Hispanic students in growth for reading. These results are explained by the fact that the school serves a community that has a longstanding Hispanic population and a relatively new—but rapidly growing—Asian population. The local Hispanic families have experience with the norms, expectations, and testing procedures at the school, and very few Hispanic students are classified as receiving ESL services. However, the school's Asian population is newer to the community. Many families are still navigating district processes that assume familiarity with US school norms and English-dominant systems. As a result, Asian students are more likely to receive ESL services, which shapes their assessed growth trajectories. These patterns illustrate how examining data through an intersectional lens can reveal dynamics that may be obscured in aggregate comparisons.

Actionable recommendations

How might school and district leaders put this knowledge about subgroup score differences into practice? Below, we provide four actionable recommendations, depending on the scenario and audience.

Recommendation 1: Examine data across subgroups and diverse student experiences

To effectively interpret test-score differences across groups, leaders should first focus on examining data across subgroups. Where possible and relevant, leaders should also analyze data to account for diversity of student experiences within subgroup categories, such as across different income brackets within the FRL category or across native languages of multilingual learner students. This approach allows districts to gain a clearer understanding of their student populations, identify specific needs for support, and uncover common factors associated with score differences. By examining data through the lens of various subgroup characteristics, leaders can pinpoint areas where targeted interventions are necessary and develop strategies to address the unique challenges faced by different student groups.

Recommendation 2: Look at subgroup overlap

To have more effective and fruitful discussions, it is crucial to highlight the importance of subgroup overlap. Leaders can demonstrate that, while all assessments contain some degree of bias, test scores also reflect the lived educational opportunities available to different groups—opportunities shaped not only by school practices but by broader systemic inequities. By showcasing how membership in multiple subgroups overlaps and its influence on academic performance, leaders can better understand the complexities of disparities and work collaboratively to create more equitable solutions. This approach fosters a more nuanced perspective on test score data, helping to build consensus and support for targeted initiatives aimed at narrowing score differences across groups.

Recommendation 3: Identify themes and areas for improvement across subgroups

When making changes within schools and districts to address differences across groups, it is essential to look across subgroups to identify common themes and areas for improvement. By framing the discussion around access and opportunity, leaders can develop comprehensive strategies that address the root causes of achievement gaps. For example, if data reveals that several subgroups lack access to advanced coursework or extracurricular programs, districts can implement policies to expand these opportunities for all students. Additionally, some strategies may need to be specific to certain subgroups based on their unique needs and challenges. Research on successful interventions in other schools can also provide valuable insights into what works, allowing leaders to adopt evidence-based practices that have been proven to reduce specific subgroup differences. Beyond analyzing data, leaders also need a deep and nuanced

understanding of the communities they serve to identify underlying themes and areas of need that numbers alone may not reveal. By combining broad, inclusive policies with targeted interventions, schools can create a more equitable educational environment that supports the success of all students.

Recommendation 4: Share data with us

Many districts using MAP Growth do not share enough subgroup data to allow for data disaggregation within MAP Growth reports. This is a barrier to many of the recommendations described above. Sharing these data provides significant benefits: it enables deeper analysis across multiple subgroups simultaneously (e.g., Asian students who are also low-income) and supports a clearer understanding of achievement patterns and equity gaps. With this information, districts can receive tailored insights that inform more effective strategies for improving student outcomes.

Conclusion

It is essential to understand that test scores reflect the lived educational opportunities and experiences of students. Factors such as systemic bias and inequity, differential access to quality education, socioeconomic status, and availability of community resources play a significant role in shaping academic performance. By considering these contextual factors, we can better interpret test score differences and avoid attributing differences in test scores solely to inherent abilities or to presumed cultural or behavioral traits of student groups.

We encourage you to focus on the following key takeaways:

1. No test score is a pure measure of a student's potential or what the student has learned. Every test is influenced by a range of multiple, interrelated, and systemic factors.
2. Test-score differences reflect students' opportunities to learn—not inherent group traits. Leaders should avoid deficit explanations or assumptions based on gender, race, income, language background, or disability status.
3. Disaggregating test scores or achievement growth by student subgroup is a great first step to mitigating any differences, but this often requires an understanding of how subgroup categories mask wide variation and overlap. Interpreting test scores responsibly requires a multidimensional and context specific approach.

References

- Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. E. (2001). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Routledge.
- Downey, D. B., & Condrón, D. J. (2016). Fifty years since the Coleman Report: Rethinking the relationship between schools and inequality. *Sociology of education, 89*(3), 207-220.
- Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives, 30*(3), 85-108.
- Kuhfeld, M. & Lewis, K. (2024). Assessing evidence of academic recovery: Slight progress in math, hardly any in ELA. *Brookings*. Retrieved from <https://www.brookings.edu/articles/assessing-evidence-of-academic-recovery-slight-progress-in-math-hardly-any-in-ela/>
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational researcher, 35*(7), 3-12.
- Rodríguez, M. (2026). 12 common questions parents ask about MAP Growth. *NWEA: Teach. Learn. Grow. The education blog*. Retrieved from <https://www.nwea.org/blog/2026/12-common-questions-parents-ask-map-growth-assessment/>
- Reardon, S. F. (2013). The widening income achievement gap. *Structure of Schooling: Readings in the Sociology of Education*.



NWEA®, a division of HMH, supports educators worldwide by providing responsive, evidence-based assessment solutions that illuminate learning needs and fuel student growth. For more than 40 years, NWEA has developed innovative pre-K-12 assessments, including its flagship assessment—MAP Growth, and professional learning that helps educators strengthen their practice and improve student outcomes. As part of its commitment to bring valuable insights to the education community, NWEA engages in research examining issues that shine a light on inequities and other barriers to academic opportunities. Visit [NWEA.org](https://www.nwea.org) to find out how NWEA partners to help all kids learn.

© 2026 HMH Education Company. NWEA and MAP are registered trademarks, and MAP Growth is a trademark, of HMH Education Company in the US and in other countries. All rights reserved. The names of other companies and their products mentioned are the trademarks of their respective owners.

APR26 | NWEA_WF2687700