WP-2025-008

When Low Scores Don't Tell the Full Story: A Composite Indicator of Student Achievement

Japneet Kaur

Indira Gandhi Institute of Development Research, Mumbai

April 2025

When Low Scores Don't Tell the Full Story: A Composite Indicator of Student Achievement

Japneet Kaur

Email(corresponding author): japneet@igidr.ac.in

Abstract

This paper introduces a new composite indicator of student achievement, grounded in an axiomatic framework. Unlike conventional measures that assign equal weight to all subjects, our index applies student- and subject-specific weights, placing greater emphasis on areas where a student performs well. This allows for a more individualized assessment, recognizing strengths in non-core subjects like music, sports, or social sciences. Using test score data from 44,173 students studying in 117 private English medium schools in rural North India, we compare our indices with the traditional average score index. The results show that a substantial proportion of students initially ranked in the bottom quartile move up significantly under our metric, highlighting overlooked talent. The proposed indices CS1 and CS2 markedly increase mean scores from 0.696 (under the original index CS0) to 0.838 and nearly 1.0, respectively, while sharply reducing standard deviations from 1.88 to 0.129 and 0.017.

Keywords: Composite Indicators, Academic Achievement, Multiple Intelligences

JEL Code: D31; I31; P36

When Low Scores Don't Tell the Full Story: A Composite Indicator of Student Achievement

Japneet Kaur*

April 8, 2025

Abstract

This paper introduces a new composite indicator of student achievement, grounded in an axiomatic framework. Unlike conventional measures that assign equal weight to all subjects, our index applies student- and subject-specific weights, placing greater emphasis on areas where a student performs well. This allows for a more individualized assessment, recognizing strengths in non-core subjects like music, sports, or social sciences. Using test score data from 44,173 students studying in 117 private English medium schools in rural North India, we compare our indices with the traditional average score index. The results show that a substantial proportion of students initially ranked in the bottom quartile move up significantly under our metric, highlighting overlooked talent. The proposed indices CS_1 and CS_2 markedly increase mean scores from 0.696 (under the original index CS_0) to 0.838 and nearly 1.0, respectively, while sharply reducing standard deviations from 1.88 to 0.129 and 0.017.

1 Introduction

Can a handful of subjects provide a fair measure of a student's potential? World over, an overwhelming majority of scholars use test scores only in Math and English to evaluate overall student performance. However, an increasing body of evidence suggests that individuals are born with multiple abilities (Davis et al., 2011; Campbell

*I am deeply grateful to Prof. Satya Chakravarty and Prof. Rupayan Pal for their guidance, comments, and suggestions. I am especially thankful to Prof. S. Chandrasekhar for introducing me to this research idea.

and Campbell, 1999; Karaduman and Cihan, 2018; Gardner, 1993b). These include linguistic, cognitive, visual-spatial and kinaesthetic ability. A student might perform disastrously in Math and English but might do exceedingly well in say music, sports, public-speaking or the social sciences. Is it plausible to construct a composite indicator of student achievement that takes multiple areas into account and assigns higher weights to subjects in which a student excels? After all, a child labeled "low performance" might in reality possess exceptional talents that traditional measures do not recognize.

Globally, only a select few education systems use both student-specific and subject-specific weights to calculate performance, tailoring assessments to individual strengths. For example, Texas[1] uses a growth-based weighting approach in its Student Achievement Indicator, giving higher weights to subjects where students show significant progress. This method recognizes individual growth rather than just raw scores, providing a personalized assessment. In New Zealand's National Certificate of Educational Achievement[2] (NCEA), students accumulate credits in various subjects, with optional courses and levels of difficulty. Here, students can select courses that align with their strengths, and advanced courses carry higher weight, allowing both student-specific and subject-specific emphasis. Similarly, some competency-based systems (like those in Sweden) indirectly incorporate individual strengths by focusing on skill mastery across subjects, where weights may vary based on demonstrated competencies rather than just standardized subject marks. These systems aim to offer a more personalized view of student achievement by valuing diverse areas of expertise. Meanwhile, in India, the Central Board of Secondary Education[3] evaluates students by averaging their top five subject scores, ensuring a focus on their best-performing areas.

In this paper, we introduce a novel composite indicator of student performance that comprehensively evaluates students across multiple domains, assigning personalized, subject-specific weights that prioritize areas where each student excels. Using a unique data set that encompasses the performance of 44,173 students across six subjects - English, Science, Punjabi, Mathematics, Social Science, and Divinity (Behavior and reading scriptures) - we illustrate how our indicator significantly enhances existing

[1] Texas Education Agency. (2024). 2024 accountability manual: Chapters 2, 4, and 5. https://tea.texas.gov/texas-schools/accountability/academic-accountability/performance-reporting/2024-accountability-manual-full.pdf

[2] New Zealand Qualifications Authority. (n.d.). Understanding NCEA. https://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/

[3] https://www.cbse.gov.in

measures of student achievement.

Furthermore, our approach to building this indicator is grounded in an axiomatic framework, marking the first axiomatic characterization of a composite indicator of student achievement. In proposing these axioms, we draw inspiration from similar advancements in the literature on inequality and poverty. For example, the function mapping test scores to student performance, termed the achievement function, must adhere to some basic principles. These are as follows.

First, the function that transforms the scores to their corresponding achievement values must be bounded. A bounded function ensures that all transformed values fall within a specific predefined range (such as [0, 1]). This facilitates the comparison of scores by placing them on a common scale, which is particularly important when aggregating or analyzing data across different tests or assessments, as it allows for meaningful comparisons of achievement levels.

Second, mapping must ensure that higher scores correspond to higher achievement values. In addition, it is important that no two different scores yield the same achievement value. These conditions imply that the achievement function must be strictly increasing.

Third, the achievement function should be designed so that marginal changes in the student's score lead to only slight proportional changes in the corresponding achievement value, avoiding abrupt increases. For example, consider a function that converts test scores from 0 to 100 into a 0-to-1 achievement scale. If a student's score increases from 75 to 76, the achievement value should not leap from 0.5 to 0.99. Furthermore, the mapping should be robust to potential measurement errors, ensuring that small fluctuations in scores do not translate into disproportionately large changes in achievement values. These considerations underscore the importance of ensuring that the achievement function is continuous, providing smooth transitions without sudden spikes.

Fourth, the achievement function should prioritize early performance improvements over later refinements. For example, if a low-performing student improves his score from 10 to 20, the corresponding increase in achievement value should exceed that observed when a high-performing student raises his score from 85 to 95.

Fifth, the achievement mapping must be scale-invariant. A scale-invariant achievement function ensures that the relative achievement levels between students remain consistent, regardless of how the raw scores are scaled or transformed. This means the function depends on the relative positioning of scores, not their absolute values. For example, if two students have scores of 50 and 75, with corresponding achievement values of 0.5 and 0.75. After scaling the scores to 0.5 and 0.75, a scale-invariant function would still map these to the same achievement values of 0.5 and 0.75, preserving the proportional gaps and rankings.

Moreover, our characterization reveals that the members of the singularized family uniquely fulfill the axioms we have established. We show that these axioms are independent, underscoring the essential role each one plays in the characterization process. In essence, the set of axioms we utilize is minimal-each axiom is indispensable, and removing any one of them would undermine the integrity of the entire framework.

Our analysis shows that the proposed indicators, CS_1 and CS_2, diverge significantly from the original index CS_0 in both the average scores and the distributional characteristics. The mean scores increase from 0.696 under CS_0 to 0.838 with CS_1, and almost reach 1.0 with CS_2. At the same time, score dispersion drops sharply, with standard deviations falling from 1.88 (CS_0) to 0.129 (CS_1) and 0.017 (CS_2), suggesting a reduced performance inequality. Minimum scores also improve markedly, particularly under CS_2, where the floor rises from 0.019 to 0.825 (Table 1). Grade-wise trends (Table 3) further reinforce these differences. While average scores decline steeply across grades under CS_0, this pattern is far less pronounced under CS_1 and nearly absent under CS_2. Quartile-level analysis (Table 2) shows that the gains are largest for lower-performing students, especially under CS_2, where the average scores in Quartiles 2 to 4 approach 1.0, and even Quartile 1 reaches 0.89. Non-parametric analyses-transition probabilities and directional rank mobility-reveal widespread upward shifts across the performance distribution. Under CS_1, students exhibit moderate upward shifts, especially from the lower quartiles (Table 4). Under CS_2, nearly all students move to the top quartile, indicating a strong uplift in relative performance (Table 5). Rank mobility estimates show gains of 10-20 percentage points with CS_1, and 30–60 points with CS_2, particularly benefiting initially low performers (Table 6 and Table 7). In sum, CS_1 and CS_2 not only improve average outcomes but also compress the distribution, reduce inequality, and enhance students' relative standing.

The remainder of the paper is organized as follows. Section 2 outlines the formal framework for our proposed indicator along with the axioms that the indicator must satisfy. In Section 3, we characterize the general family of indices that capture student achievement. Section 4 presents a brief overview of existing indicators of student achievement. Finally, Section 5 offers an empirical illustration of the family of indices we propose, and in Section 6 we conclude.

2 Formal Framework

Let Y_{ij} denote student $i's$ standardised score in subject j where $i = 1, ..., n$ and $j = 1, ..., m$ and Y_{ij}'s are continuous variables. Our proposed indicator for student i is given as under:

$$I_i = \frac{1}{m} \sum_{j=1}^{m} \omega_{ij} f(Y_{ij})$$

where $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$.

Here f denotes the achievement function and ω_{ij} denotes the weight assigned to student i in subject j. The achievement function f satisfies the following assumptions.

2.1 Axioms For an Achievement Function

A.1. The achievement function $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is strictly increasing.

Formally, for any two scores Y_{ij} and Y'_{ij}

$$\text{if } Y_{ij} < Y'_{ij}, \text{ then } f(Y_{ij}) < f(Y'_{ij}).$$

Intuitively, a higher score must lead to higher achievement. The distinction between increasing and a strictly increasing function is crucial. To illustrate, consider a function that maps test scores from the range of 0 to 100 to an achievement scale between 0 and 1. Suppose we use an increasing function where, for scores between 0 and 50, the achievement value remains constant at 0.5; for scores from 51 to 80, the achievement value rises linearly up to 0.8; and for scores between 81 and 100, it increases linearly to reach 1.0. In this case, the function includes flat segments-intervals where the achievement value does not change despite variations in scores. For instance, all scores between 0 and 50 correspond to the same achievement value of 0.5. In contrast, a strictly increasing function eliminates flat segments and ensures that every score corresponds to a unique achievement value. For example, we could define the function as $f(x) = 0.01x$, where each score x maps directly to an achievement value. A score of 0 maps to an achievement of 0, 30 maps to 0.3, 50 maps to 0.5, 75 maps to 0.75, 90 maps to 0.9, and 100 maps to 1.0. Here, every increase in score results in a distinct increase in achievement value, ensuring no two scores share the same achievement. This strict one-to-one mapping highlights the key feature of a strictly increasing function: every point on the score scale uniquely contributes to the achievement value.

A.2. The achievement function $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is continuous.

Formally, $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is continuous if for every z-score $Y^0_{ij} > 0$ and every $\epsilon > 0$, there exists a $\delta > 0$, such that for all z-scores Y_{ij} in the domain:

$$\mid Y_{ij} - Y^0_{ij} \mid < \delta \Rightarrow \mid f(Y_{ij}) - f(Y^0_{ij}) \mid < \epsilon.$$

This means that there will not be any point at which two very close scores lead to vastly different achievement levels. Therefore, a measurement error in a score will not lead to an abrupt change in the achievement level.

A.3. The achievement function $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is strictly concave.

This implies that for any two distinct standardised scores say Y_{ij}, $Y^{'}_{ij}$ and for any $\lambda \in (0, 1)$

$$f(\lambda Y_{ij} + (1 - \lambda)Y^{'}_{ij}) > \lambda f(Y_{ij}) + (1 - \lambda) f(Y^{'}_{ij})$$

Thus the function places more weight on improvements at the lower end of the score distribution and progressively less weight on improvements at the higher end. This means that improving a score from 10 to 20 would lead to a larger jump in achievement than improving from 80 to 90.

The choice between a concave and a strictly concave function is critical. A strictly concave function ensures that every score increase contributes progressively less to the achievement value, with no flat or linear segments. In contrast, a function that is merely concave can have segments where the achievement value increases at a constant rate or even remains unchanged, reflecting lower sensitivity to score increases within specific ranges. To illustrate, consider two types of achievement functions. In a strictly concave function, the achievement value rises at a decreasing rate as scores increase. For instance, a score of 30 corresponds to an achievement value of 0.5, a score of 60 maps to 0.8, and a score of 90 maps to 0.95. Here, increasing the score from 30 to 60 (a 30-point increase) raises the achievement value by 0.3 (from 0.5 to 0.8). However, a similar 30-point increase from 60 to 90 results in a smaller gain of 0.15 (from 0.8 to 0.95). This diminishing return with higher scores demonstrates strict concavity. In contrast, a concave but not strictly concave function may include flat segments. For example, a score of 30 maps to an achievement value of 0.5, 60 to 0.8, 80 to 0.9, and 90 also to 0.9. While increasing the score from 30 to 60 raises the achievement value by 0.3 (from 0.5 to 0.8), moving from 60 to 80 adds only 0.1, and from 80 to 90 results in no gain at all, as the value remains flat at 0.9. This flat segment reflects the concave function's allowance for regions where achievement values stop increasing, distinguishing it from strict concavity.

A.4. The achievement function $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is scale invariant if for any positive scalar c and any score $Y_{ij} \geq 0$, $f(cY_{ij}) = c^\alpha f(Y_{ij})$ where α is a constant that characterizes the function's sensitivity to scaling. Scale invariance means that the student's performance remains unaffected by uniform scaling of scores. For instance, if one student's score is 80 and another student's is 40, doubling both scores to 160 and 80, respectively, should not change their relative performance according to a scale-invariant function. If the function ranks the student with 80 higher than the student with 40 originally, this ranking should remain the same after scaling.

A.5. The achievement function $f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is differentiable.

$f : (\infty, \infty) \longrightarrow \mathbb{R}_+$ is said to be differentiable at point $c \in A$ if the following limit exists:

$$f'(c) = lim_{h \longrightarrow 0} \frac{f(c+h) - f(c)}{h}.$$

A differentiable function would ensure a smooth transition in the mapped values across the score range. This means there would be no abrupt jumps, kinks, or breaks in the mapping from test scores to the [0,1] scale.

3 The Characterization Theorem

Theorem 3.1. *The achievement function* $g : [0, 100] \longrightarrow [0, 1]$ *satisfies continuity, scale invariance, monotonicity, and strict concavity, normalizability, and differentiability if and only if* g *is strictly concave, scale invariant and differentiable over the interval [0,100].*

Proof. Step 1: We know that a concave function on a convex subset of $\mathbb{R}$ is continuous on the interior of its domain. If we prove right-continuity at $x = 100$ and left-continuity at $x = 0$, we are done.

Step 2: We first show that g is strictly increasing on [0,0]. For $x \in (0, 100)$, strict concavity implies

$$g(x) = g\left(\frac{x}{100}.100 + \left(1 - \frac{x}{100}\right).0\right) > \frac{x}{100}g(100) + \left(1 - \frac{x}{100}\right)g(0) = \frac{x}{100}.$$

Thus $g(x) > \frac{x}{100}$ for all $x \in (0, 100)$. Suppose g is not strictly increasing, then there exist $x_1 < x_2$ with $g(x_1) \geq g(x_2)$. By strict concavity, for $t \in (0, 1)$

$$g(tx_1 + (1-t)x_2) > tg(x_1) + (1-t)g(x_2) \geq tg(x_2) + (1-t)g(x_2) = g(x_2).$$

This contradicts the existence of a non-increasing segment. Therefore g is strictly increasing on $[0, 0]$.

Step 3: We now need to prove continuity at end points. We first show right continuity at $x = 0$. Since g is strictly increasing, $lim_{x \longrightarrow 0_+} g(x)$ exists and $lim_{x \longrightarrow 0^+} g(x) \geq g(0) = 0$. Since g is strictly increasing $g(x) > \frac{x}{100}$. As $x \longrightarrow 0_+$, $\frac{x}{100} \longrightarrow 0$. Therefore $\lim_{x \longrightarrow 0^+} g(x) = 0 = g(0)$. Next we show left continuity at $x = 100$. Since g is strictly increasing, $\lim_{x \longrightarrow 100^-} g(x)$ exists and $\lim_{x \longrightarrow 100^-} g(x) \leq g(100) = 1$. For $x \in (0, 100)$, strict concavity implies

$$g(x) = g\big(\frac{x}{100}.100 + \big(1 - \frac{x}{100}\big).0\big) > \frac{x}{100}.1 + \big(1 - \frac{x}{100}\big).0 = \frac{x}{100}.$$

As $x \longrightarrow 100^-$, $\frac{x}{100} \longrightarrow 1$. Therefore $\lim_{x \longrightarrow 100^-} g(x) = 1 = g(100)$. Since g is right continuous at $x = 0$ and left continuous at $x = 100$ therefore g is continuous at $[0, 100]$. □

Proposition 3.2. *Let* $g : [0, 100] \longrightarrow [0, 1]$ *with* $g(0) = 0$ *and* $g(100) = 1$ *be strictly concave. Then* g *is strictly increasing.*

Proof. Given $g(0) = 0$ and $g(1) = 1$, by strict concavity of g, $g(\theta 0 + (1 - \theta)) > \theta g(0) + (1 - \theta) g(1)$. Hence $g(1 - \theta) > (1 - \theta)$. Thus for any $0 < u < 1$, $g(u) > u$, which implies that $g(u)$ is increasing at 0. □

Claim 1: Strict concavity of g implies that $g(u) > 0$ for all $0 < u < 100$.

Proof. Let $x_1 = 0$, $x_2 = 100$. Consider $x = (1 - \lambda)(0) + \lambda(100) = 100\lambda$. By strict concavity, $g(x) = g(100\lambda) > (1 - \lambda) g(0) + \lambda g(100)$. Substituting the known values, we get $g(x) > \lambda$. Since $0 < \lambda < 1$, it follows that $g(x) > 0$. □

Claim 2: g is strictly increasing.

Proof. Let g not be strictly increasing. This implies that there exist at least 3 points $0 < t_1 < t_2 < t_3 < 1$ such that $g(t_1) < g(t_2)$ and $g(t_2) > g(t_3)$. By construction we can find $0 < \delta < 1$ such that $t_3 = \delta t_2 + (1 - \delta)1$. Then by strict concavity of g,

$$g(t_3 > \delta g(t_2) + (1 - \delta) g(1)$$

$$\Rightarrow g(t_3) > \delta g(t_2) + (1 - \delta).$$

Subtract $\delta g(t_2)$ from both sides. We get

$$g(t_3) - \delta g(t_2) > (1 - \delta) \tag{3.1}$$

Since we assume $g(t_2) > g(t_3)$, thus $g(t_2) - g(t_3) > 0$. Substitute $g(t_3) < g(t_2)$ in (3.1) we get $g(t_2) > 1$. Hence we arrive at a contradiction because $g(t_2) \in [0, 1]$. □

Theorem 3.3. *The assumptions of scale invariance, concavity, differentiability, and boundary conditions are independent.*

Proof. By Theorem 1, strict concavity and monotonicity imply that the function is strictly increasing and continuous between $[0, 100]$. Hence, we can do away with the assumptions that the function is strictly increasing and continuous. A general class of functions which are strictly increasing, strictly concave, normalized and differentiable from $[0, 100] \longrightarrow [0, 1]$ would look something like follows

$$g(x) = \frac{\int_0^x f(t)d(t)}{\int_0^{100} f(t)d(t)},$$

where $f(t) > 0$, and $f'(t) < 0$ and $f(x) : [0, 100] \longrightarrow \mathbb{R}^+$ is a positive, decreasing (i.e., concave) function. Examples of valid choices for $f(x)$ are a linear decreasing function with $f(x) = a - bx$, for $0 < b < \frac{a}{100}$, an exponential decay function with $f(x) = e^{-kx}$, $k > 0$, a power-based function with $f(x) = \frac{1}{(x+c)^r}$, $c > 0, r > 1$ and a logarithmic decay function, with $f(x) = \frac{1}{\log(bx+c)}$, $b, c > 0$. Scale invariance implies that if we scale the upper bound, the form of the function stays the same after rescaling. Formally, if we define a rescaled version of g over an interval $[0, M]$, say

$$g_M(x) = \frac{\int_0^x f_M(t)d(t)}{\int_0^M f_M(t)d(t)}$$

then for scale invariance we require $g_M(x) = g\big(\frac{x}{M}.100\big)$. That is, the function looks identical after rescaling, just on a different domain. Among all the functional forms given above, we find that only the power functional form where $f(t) = t^\beta$ satisfies the scale invariance property. This leaves us with $g(x) = \big(\frac{x}{100}\big)^\alpha$ as the only functional form.

□

3.1 Weighing Scheme

We begin by converting raw scores into standardized z-scores, which indicate how far above or below the subject average a student performs, relative to the spread of scores in that subject. This adjustment accounts for differences in subject difficulty and score distribution, ensuring fair comparisons. As a result, a high score in a subject where most students struggle is given due credit. In addition, by standardizing the scores within each subject, this method avoids the problem of certain subjects having inflated or deflated raw scores. For instance, if scores in Social Science tend to be higher across the board than in Math, raw scores would be misleadingly skewed. Standardization corrects for this. The conversion is done using the following formula:

$$z_{ij} = \frac{Y_{ij} - \mu_j}{\sigma_j}$$

Here Y_{ij} denotes the raw score of the student i in the subject j, μ_j is the mean score for the subject j among all students, and σ_j is the standard deviation of the scores in the subject j. To make sure all weights are positive, we can shift each z_{ij} by adding a constant k so that all adjusted z scores z'_{ij} are non-negative.

$$z'_{ij} = z_{ij} + k$$

where k is chosen as $k =\mid min(z) + \epsilon \mid$. Here, $min(z)$ represents the minimum z-score across all students and subjects, and ϵ is a small positive number (for example, 0.01) to ensure that all adjusted z-scores z'_{ij} are strictly positive. Using the adjusted z-scores, the weight for a student i in subject j can be calculated as follows.

$$w_{ij} = \frac{z'_{ij}}{\sum_{j=1}^{m} z'_{ij}}$$

Note that students who perform consistently well across all subjects will have more evenly distributed weights, whereas those who perform outstandingly in a few subjects will receive higher weights for those subjects. This balances the recognition of generalists and specialists. Note that the weights sum up to 1.

$$\sum_{j=1}^{m} w_{ij} = \sum_{j=1}^{m} \frac{z'_{ij}}{\sum_{k=1}^{m} z'_{ik}} = \frac{\sum_{j=1}^{m} z'_{ij}}{\sum_{k=1}^{m} z'_{ik}} = 1$$

The weights are normalized to sum to 1 in all subjects for each student. This allows the weights w_{ij} to be interpreted as the relative contribution or importance of the subject j to student $i's$ overall achievement. In effect, students are allowed to "specialize" in subjects where they perform relatively better. Thus, the proposed weighing scheme attempts to integrate relative performance (within-subject comparison) and individual strengths (across-subject emphasis) into a coherent and mathematically consistent framework.

3.2 Alternative Indicators of Student Achievement

Building on the above, we propose the following alternative composite indicators.

$$CS_{i1} = \sum_{j=1}^{J} w_{ij} \left(\frac{y_{ij}}{100} \right)^{\alpha}$$

where y_{ij} is the score of student i in subject j, w_{ij} is the weight for student i in subject j and $\alpha \in (0, 1)$ determines the senstivity of the score transformation. As we move from $\alpha = 0$ to $\alpha = 1$, the transformed scores shift from being highly compressed to closely resembling the original raw scores. With α near 0, all non-zero scores are mapped close to 1, minimizing differences and emphasizing inclusivity. As α increases, the function becomes less concave, and score differences begin to reflect raw performance more sharply. At $\alpha = 1$, the transformation is linear, fully preserving the original scale and maximizing distinctions between high and low performers. Thus, higher α values emphasize achievement gaps, while lower values compress them.

We also introduce another indicator that satisfies all the desired axioms except scale invariance. It is defined as follows.

$$CS_{i2} = \sum_{j=1}^{J} w_{ij}(1 - e^{-Y_{ij}})$$

where $Y_{ij} = \frac{y_{ij}}{100}$. A key advantage of this indicator over the previous one is that it does not depend on selecting a specific value of α, making it less parameter dependent and more robust by design.

4 A Brief Review of the Existing Indices

Globally, only a few education systems tailor assessments by incorporating student-specific and subject-specific weights, aligning evaluation with individual strengths and subject relevance. For example, in the Texas accountability system[4], the Student Achievement Indicator incorporates a growth-based weighting approach that rewards academic improvement, not just proficiency. Rather than merely evaluating proficiency levels, this method rewards schools where students demonstrate substantial year-over-year academic improvement. For instance, if a student underperforms in reading but

[4]Texas Education Agency. (2024). 2024 accountability manual: Chapters 2, 4, and 5. https://tea.texas.gov/texas-schools/accountability/academic-accountability/performance-reporting/2024-accountability-manual-full.pdf

shows marked progress compared to the prior year, while their math score remains constant, the system assigns greater weight to the reading score in the composite achievement index. This incentivizes schools to focus not just on high-achieving students but also on supporting those making significant academic gains, especially in under-served areas.

Similarly, New Zealand's National Certificate of Educational Achievement[5] (NCEA) enables students to accumulate credits across subjects and levels of difficulty, with the flexibility to choose internal or external standards that reflect their interests and capabilities. Higher-level or more advanced courses contribute more significantly to overall certification, allowing both subject-specific and student-specific emphasis. For example, a student excelling in sciences may opt for advanced standards in chemistry and physics, thereby earning higher credits, while another may tailor their course load around the arts or vocational training. This structure recognizes varied learner trajectories and supports diverse strengths.

In Sweden's competency-based education system[6], assessment is focused on mastery of clearly defined competencies rather than traditional subject marks. Students are evaluated on transferable skills—such as reasoning, analysis, and problem-solving—within and across subjects. For instance, a student who excels in mathematical problem-solving can be credited with high achievement in that domain, even if their broader mathematics performance is average. This model permits individual strengths to be emphasized, and promotes a holistic understanding of learning outcomes.

Ontario's secondary education system in Canada adopts a more individualized approach through the use of Individualized Education Plans (IEPs)[7]. While core curriculum standards remain intact, students with exceptional needs receive tailored accommodations that can influence how specific subjects or competencies are weighted in evaluations. For example, a student with language difficulties may have oral assessments prioritized over written tasks, while another showing strength in creative writing may receive greater emphasis on that domain in grading. Although not a formal system of weighted scores, the IEP framework allows for adaptive and growth-oriented evaluations.

[5] New Zealand Qualifications Authority. (n.d.). Understanding NCEA. https://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/

[6] European Commission/EACEA/Eurydice. (2023). Assessment in single-structure education: Sweden. Eurydice – National Education Systems. https://eurydice.eacea.ec.europa.eu/national-education-systems/sweden/assessment-single-structure-education

[7] Ontario Ministry of Education. (2022). Special Education in Ontario: Kindergarten to Grade 12 – Policy and Resource Guide. https://www.ontario.ca/document/special-education-policy-and-resource-guide

In a similar vein, the Central Board of Secondary Education[8] (CBSE), a national education system in India governed by the Union of India, calculates students' aggregate scores based on their top five performing subjects. This approach mirrors similar practices aimed at prioritizing students' strengths and ensuring a fair representation of their academic performance. For example, assume that a student appears in 6 subjects, each graded out of 100 marks. Let the marks in these subjects be $M_{i1}, M_{i2}, M_{i3}, M_{i4}, M_{i5}, M_{i6} \in [0, 100]$. Let these marks be ordered such that $M_{i1} > M_{i2} > M_{i3} > M_{i4} > M_{i5}$. According to the best of 5 scheme, the composite score would be given as follows.

$$CS_i = \frac{1}{5}\sum_{j=1}^{5} M_{ij}$$

.

Here, j denotes the subject, while i denotes the student. Therefore, a composite score under the best of 5 scheme is calculated by sorting the 6 subject scores in descending order, selecting the top 5 scores and taking their average.

In addition, the cumulative grade point average[9] (CGPA), a measure commonly used in many educational systems to summarize a student's overall academic performance across all semesters or terms, is defined as the weighted average of the grade points obtained in all courses over multiple terms. Rather than treating all courses equally, the CGPA accounts for both the grade achieved and the weight (or credit hours) of each course, offering a more accurate reflection of academic achievement. It is typically calculated on a scale—commonly out of 10 or 4, depending on the institution or country.

$$CGPA = \frac{\sum(\text{Credit Hours} \times \text{Grade Points})}{\sum \text{Credit Hours}}$$

where Grade Points refer to the numeric values assigned to the letter grades (e.g., A = 4, B = 3), and credit hours refer to the weight or importance of the course.

For example, assume that a student takes three subjects with varying credit weights and earns different grades for each. In Subject 1, which carries 3 credits, the student earns an A, corresponding to 4.0 grade points. In Subject 2, worth 4 credits, they receive a B, corresponding to 3.0 grade points. In Subject 3, a 3-credit course, the

[8] https://www.cbse.gov.in

[9] Collegedunia. (n.d.). CGPA Calculator: How to calculate CGPA. https://collegedunia.com/exams/cgpa-calculator-how-to-calculate-cgpa-articleid-2403

student earns a C, which translates into 2.0 grade points. To calculate the CGPA, we multiply each grade point by its respective course credit and take the weighted average:

$$CGPA = \frac{(3 \times 4.0) + (4 \times 3.0) + (3 \times 2.0)}{3 + 4 + 3} = 3.$$

Finally, the Intelligence Quotient[10] (IQ) is also measured using some kind of weighing technique. IQ is commonly measured using standardized tests such as the Wechsler Adult Intelligence Scale (WAIS) or the Wechsler Intelligence Scale for Children (WISC). These tests assess different domains of cognitive ability, such as verbal comprehension, perceptual reasoning, working memory, and processing speed. For example, a child taking the WISC might be asked to define vocabulary words (verbal comprehension), solve visual puzzles (perceptual reasoning), recall digit sequences (working memory), and quickly match symbols (processing speed). Each domain yields a scaled score that is then combined into a full scale IQ (FSIQ). To meaningfully combine these scores, factor analysis is used. Factor analysis is a statistical method that identifies patterns of correlation among test items and estimates how much each test (or subtest) contributes to an underlying ability, such as general intelligence or g. The factor loadings from this analysis reflect the strength of association between a subtest and the general intelligence factor and serve as weights in computing the FSIQ. For example, consider a student who scores 14 in Verbal Comprehension, 12 in Perceptual Reasoning, 10 in Working Memory, and 8 in Processing Speed, each measured on a scale of 20. Based on factor analysis conducted on a large sample, the corresponding weights (or factor loadings) for these subtests are 0.40, 0.30, 0.20 and 0.10, respectively, reflecting the relative contribution of each ability to overall intelligence. To compute the composite IQ score (before standardization), multiply each score by its weight:

$$\text{Raw Composite IQ Score} = (14 \times 0.40) + (12 \times 0.30) + (10 \times 0.20) + (8 \times 0.10) = 12.$$

This raw score is then standardized to have a mean of 100 and standard deviation of 15, typically using population norms, to produce the final Full-Scale IQ score.

5 An Empirical Illustration

5.1 Dataset

Using unique data on test scores obtained by 44,173 students (spanning grades 1 to 10) studying in a chain of 117 rural schools in five states of North India, we compare the

[10] Encyclopædia Britannica. (n.d.). IQ. In Britannica.com. https://www.britannica.com/science/IQ

three indicators of achievement. We use student scores across six subjects: English, Math, Punjabi (mother tongue), Science, Social Science, and Divinity.

5.1.1 About the Schools

We collect data from Akal Academies[11], a chain of 117 private schools located exclusively in rural areas of North India. These are English medium schools that adhere to the National Curriculum Framework and are affiliated with the Central Board of Secondary Education (CBSE), a national board of education in India for public and private schools, managed by the Union Government of India. Akal Academies follow a uniform curriculum and examination system. All students take a common exam, and the answer scripts are graded by teachers from a randomly chosen school (belonging to the same chain) other than the one in which a child studies.

5.2 Empirical Methodology

We compare three indicators: the first employs a simple averaging scheme, while the other two use our proposed alternative weighting schemes. Details of all three are given below.

$$CS_{i0} = \frac{1}{6}\left(\frac{y_{i,Math}}{100} + \frac{y_{i,English}}{100} + \frac{y_{i,Punjabi}}{100} + \frac{y_{i,Science}}{100} + \frac{y_{i,SocialScience}}{100} + \frac{y_{i,Divinity}}{100}\right) \quad (5.1)$$

$$CS_{i1} = w_{i,Math}\left(\frac{y_{i,Math}}{100}\right)^{0.5} + w_{i,English}\left(\frac{y_{i,English}}{100}\right)^{0.5} + w_{i,Punjabi}\left(\frac{y_{i,Punjabi}}{100}\right)^{0.5} + w_{i,Science}\left(\frac{y_{i,Science}}{100}\right)^{0.5} + w_{i,SocialScience}\left(\frac{y_{i,SocialScience}}{100}\right)^{0.5} + w_{i,Divinity}\left(\frac{y_{i,Divinity}}{100}\right)^{0.5} \quad (5.2)$$

$$CS_{i2} = w_{i,Math}(1 - e^{-Y_{(i,Math)}}) + w_{i,English}(1 - e^{-Y_{(i,English)}}) + w_{i,Punjabi}(1 - e^{-Y_{(i,Punjabi)}}) + w_{i,Science}(1 - e^{-Y_{(i,Science)}}) + w_{i,SocialScience}(1 - e^{-Y_{(i,SocialScience)}}) + w_{i,Divinity}(1 - e^{-Y_{(i,Divinity)}}) \quad (5.3)$$

[11] Akal Academies are managed by the Kalgidhar Trust - Baru Sahib, a non-profit organization that is committed to providing quality education to rural areas of North India. The schools offer a rare blend of modern scientific education and spiritual values. Apart from the focus of the NGO in the field of education, it also works in areas of women empowerment, health and medical sciences, disaster management and treatment of persons afflicted with substance abuse.

We start by comparing the average composite scores of our proposed indices (CS_1 and CS_2) with the average composite score of the original indicator (CS_0). These averages are calculated separately for each grade. We use a paired t test to assess whether the mean differences between CS_1 and CS_0, as well as CS_2 and CS_0, are significantly different from zero for each grade.

5.2.1 Non-parametric methods

To better understand how our proposed indicators compare with the original indicator, we complement our analysis by constructing transition probabilities and rank mobilities. These metrics are widely used to study intergenerational income or occupational mobility within households (Shorrocks, 1978, Bhattacharya and Mazumder, 2011). However, only a handful of studies that analyze academic achievement make use of these metrics[12]. These measures estimate the proportion of individuals who shift from their original position when scores are recalculated using our proposed indices, conditional on their rank in the original index. The following example illustrates. Let us assume that there exist 10 individuals in the bottom 25 percent of the score distribution when the score is calculated using CS_0. Now when the score is calculated using CS_1, 3 among those lying in the bottom quartile experience positive upward movement, that is, they now lie within the top 75 percent of the score distribution. Therefore, this would imply that the upward mobility for students is 30 percent.

More formally, transition probability matrices are constructed as follows. Corresponding to each composite indicator, the scores are partitioned into 4 quartiles - less than or equal to 0.25 (quartile 1) greater than 0.25 and less or equal to 0.5 (quartile 2), greater than 0.5 and less than or equal to 0.75 (quartile 3), and greater than 0.75 and less than or equal to 1 (quartile 4). The total number of students corresponding to each quartile is then calculated. Next, each element of the transition probability matrix is constructed by calculating different ratios. For example, if we wish to know the probability that individuals will transition from quartile 1 to quartile 2 when composite scores are calculated using CS_0 versus when scores are calculated using CS_1, we will have to divide the number of individuals in quartile 2 using CS_1 conditional on them being in quartile 1 using CS_0 with the number of individuals in quartile 1 using CS_0. Movements from a lower quartile towards a higher quartile are termed upward transition probabilities, while movements from higher quartile towards lower quartiles are termed downward transition probabilities. Movements within the same quartile are

[12]Examples include McDonough, 2015; Ellison and Swanson, 2023.

termed staying probabilities. Assuming that the scores are divided into k partitions, the transition matrix corresponding to any two indices, CS_0 and CS_1 is calculated as follows.

$$\pi_{CS_0,CS_1} = \begin{bmatrix} \pi_{11}^{CS_0,CS_1} & ... & ... & \pi_{1k}^{CS_0,CS_1} \\ ... & ... & ... & ... \\ \pi_{1k}^{CS_0,CS_1} & ... & ... & \pi_{kk}^{CS_0,CS_1} \end{bmatrix}$$

Each element $\pi_{kl}^{CS_0,CS_1}$ of the matrix $(K \times K)$ denotes the fraction of children in partition k using CS_0 who are in partition l using CS_1. $\pi_{kl}^{CS_0,CS_1}$ denotes the staying probability if $k = l$, the upward transition probability if $k > l$ and the downward transition probability if $k < l$.

$$\pi_{kl}^{CS_0,CS_1} = \frac{Pr(y^{CS_0} \in k, y^{CS_1} \in l)}{Pr(y^{CS_0} \in k)} \quad k, l = 1, ...K.$$

Despite being instructive, transition probabilities do not take into account movements within a partition or a quartile. That is, to be qualified as mobile, individuals only need to cover the distance between their position and the upper bound of their quartile. These upper and lower bounds corresponding to each partition are arbitrary (or non-unique). It is possible that an individual's performance, as measured by CS_0, lies close to the lower bound of a given partition, and although it improves significantly when assessed using CS_1, it still remains within the same partition. In contrast, another individual near the upper bound of the same quartile might gain less than the first person but still cross the threshold and move into the next quartile. Thus, the latter individual, despite experiencing lesser gains, is said to made an upward movement, while the former individual is said to have stayed in the same quartile – despite making larger gains. This implies that individuals close to the lower bound have a lesser likelihood of upgrading their performance than those close to the upper bound. This suggests that the performance gains experienced by individuals are only comparable if their initial position is taken into account.

Therefore, we employ an improved version of transition probability matrix that takes into account an individual's position in the dataset - directional mobility. Here, a student is said to have made 'upward movement' (downward movement) if the difference in performance between the two indicators $(CS_1 - CS_0)$ lies beyond certain thresholds γ. Since γ is the same for all, therefore, everyone is equally likely to be upwardly or downwardly mobile. Formally, the following expression is used to calculate the proportion of individuals who experience upward mobility relative to a given value of γ.

$$\theta_{k,\gamma}^{CS_0,CS_1} = \frac{Pr(CS_0 \in k, (CS_1 - CS_0) \geq \gamma)}{Pr(CS_0 \in k)}$$

Similarly, the proportion of people who experience downward mobility is calculated as follows.

$$\theta_{k,\gamma}^{CS_0,CS_1} = \frac{Pr(CS_0 \in k, (CS_1 - CS_0) \leq \gamma)}{Pr(CS_0 \in k)}$$

Note that $\gamma \in (-\infty, \infty)$. For positive values of γ, we focus exclusively on upward mobility, while for negative values, we examine downward mobility. For $\gamma = 0$, we consider both upward and downward mobility[13].

5.3 Results

The comparison between our proposed indicators-CS_1 and CS_2-and the original index CS_0 reveals considerable differences in both central tendencies and the spread of student performance scores (Table 1). On average, student scores are markedly higher when evaluated using the new indicators: the mean increases from 0.696 under CS_0 to 0.838 for CS_1, and nearly reaches its upper bound under CS_2 at 0.998. At the same time, the standard deviation shrinks dramatically: from 1.88 under the original index to just 0.129 with CS_1 and further down to 0.017 with CS_2, highlighting a substantial compression of the distribution and suggesting reduced disparities in performance. This pattern is also reflected in the minimum scores: While CS_0 records a minimum of 0.019, the floor rises significantly to 0.135 under CS_1 and 0.825 under CS_2..

In Table 3 we present grade wise average scores across the three indices. It should be noted that while CS_2 significantly improves the student's performance, CS_1 only yields a modest improvement compared to the original indicator. This pattern is evident in columns 4 and 5, which show the differences in mean scores between CS_2 and CS_0, and between CS_1 and CS_2, respectively. The gap between CS_2 and CS_0 increases noticeably with each successive grade, while the difference between CS_0 and CS_1 increases only marginally. Importantly, these results do not suggest a worrying trend in student performance. Although CS_0 shows a steep decrease in average scores, from 0.845 in Grade 1 to 0.527 in Grade 10, the decline is considerably less severe when performance is evaluated using CS_1 and CS_2. Specifically, the average score based on CS_1 declines from 0.923 in Grade 1 to 0.7595 in Grade 10, while scores based on CS_2

[13]For $\gamma = 0$ only the inequality signs $<$ and $>$ are considered.

remain consistently high, staying within the range of 0.9 to 1 across all grades.

Next, we compare mean scores across indices by quartile (Table 2). To do this, we divide the distribution of scores under CS_0 into four quartiles: Quartile 1 (0-0.25), Quartile 2 (0.25-0.50), Quartile 3 (0.50-0.75) and Quartile 4 (0.75-1). For each quartile, we then compute the average score of individuals based on our proposed indicators. Using CS_0, the mean scores for Quartiles 1 through 4 are 0.070, 0.419, 0.637, and 0.860, respectively. With CS_1, these means increase to 0.252, 0.680, 0.817, and 0.931, showing substantial improvements, especially in the lower quartiles. The uplift is most pronounced for lower-performing students, while the differences become smaller in the higher quartiles. Under CS_2, the transformation is even more striking: the mean score for Quartile 1 jumps to 0.89, and for Quartiles 2, 3, and 4, the averages are all close to 1. Among the three indices, CS_1 is the most conservative in adjusting scores, offering moderate improvements while maintaining greater differentiation across performance levels.

We then supplement our analysis with non-parametric techniques that examine the proportion of individuals located at various points in the performance distribution, based on our proposed indices, conditional on their placement within a given percentile range according to the original indicator. To this end we construct transition probabilities and directional rank mobilities. The results of the transition probability matrices are presented in Table 4 and Table 5. We start by comparing CS_0 with CS_1 (Table 4). We find that 15.68 percent of students who originally fell into quartile 1 based on CS_0 move to quartile 2 when their score is calculated using CS_1. Similarly, 16.46 percent of students who were in quartile 1 based on CS_0 move to quartile 3 under CS_1. Interestingly, about 95 percent of the students who were in quartile 2 transition to quartile 3. Furthermore, 92.69 percent of the individuals initially in quartile 3 move to quartile 4, while 7.31 percent remain in quartile 3. Next, we compare CS_0 with CS_2. The results are presented in Table 5. Interestingly, using CS_2, we find that all students shift toward the top quartile (Q_4), regardless of their initial position in the distribution. This paints a highly encouraging picture of overall performance.

Tables 6 and 7 present our estimates of the directional rank mobility matrices, comparing CS_0 with CS_1 and CS_0 with CS_2 respectively. We begin by comparing CS_0 with CS_1. The results indicate that all students originally in Quartiles 1 and 2 under CS_0 experience a uniform upward shift of 20 percentage points in their performance rankings when assessed using CS_1. Among those initially in Quartile 3, all show a gain of 10 percentage points, while 30 percent achieve a more substantial increase of

20 percentage points. For students originally placed in Quartile 4, 26 percent register a modest gain of 10 percentage points. Next, we turn to the comparison between CS_2 and CS_2. Here, the upward mobility is even more pronounced. All students in Quartile 1 under CS_0 experience a gain of 30 percentage points when recalculated using CS_2. Likewise, every student in Quartile 2 records a gain of 40 percentage points, with 31 percent of them achieving an even greater improvement of 60 percentage points.

6 Conclusion

This paper presents an axiomatic construction of a novel composite indicator to measure student achievement. Unlike conventional metrics that aggregate individual scores by assigning equal weights to all subjects, our index introduces student-specific and subject-specific weights, giving greater emphasis to subjects in which a student excels. This approach acknowledges that a student who underperforms in subjects like Mathematics and English may still demonstrate exceptional abilities in areas such as music, sports, public speaking, or social sciences. We propose a composite measure that accounts for these diverse strengths, offering a more holistic assessment of student achievement. To evaluate the effectiveness of our proposed index, we employ a unique dataset comprising 44,173 students from a network of English-medium private schools in rural North India. We compare the traditional aggregate score with our proposed index by analyzing mean scores across various percentile groups.

Our analysis reveals that the proposed indicators, CS_1 and CS_2 , substantially differ from the original index CS_1 in both average performance levels and distributional characteristics. Overall, mean scores increase notably under the proposed indicators: from 0.696 in CS_0 to 0.838 in CS_1, and nearly reaching 1 under CS_2. Alongside this, score dispersion decreases significantly, indicating a reduction in performance inequality. The standard deviations drop from 1.88 in CS_0 to 0.129 and 0.017 in CS_1 and CS_2, respectively. Minimum scores also rise sharply, particularly under CS_2, where the lowest score is 0.825 compared to 0.019 in the original index. Non-parametric analysis of transition probabilities and directional rank mobility matrices further confirms these patterns. A notable proportion of students shift to higher performance quartiles when evaluated using the proposed indicators. Under CS_1, students from lower quartiles show modest but widespread upward movement. In contrast, under CS_2, virtually all students transition into the top quartile, signaling a dramatic boost in performance rankings. Directional rank mobility estimates show consistent gains of 10-20 percentage points under CS_1, while CS_2 generates even larger shifts of 30 to

60 percentage points, especially for initially low-performing students. Together, these findings indicate that the proposed indicators not only improve average performance levels but also compress the distribution, thus reducing inequality and improving student relative rankings. Crucially, they reveal that students traditionally labeled as low-performing may not, in fact, be underachievers when their individual strengths and subject-specific proficiencies are appropriately recognized. This underscores the need to revisit our assessment frameworks.

Table 1: Summary Statistics of the Proposed Indices

Index	**Mean**	**SD**	**Min**	**Max**	**N**
CS_0	0.6955	0.1884	0.0185	1	44,173
CS_1	0.8379	0.1291	0.1347	1	44,173
CS_2	0.9976	0.0166	0.8246	1	44,173

Notes: Column 2 reports the mean score for each indicator, Column 3 presents the corresponding standard deviation, while Columns 4 and 5 show the minimum and maximum values of the indicator, respectively.

Table 2: Quartile-wise Mean Scores Across All Indicators

Quartile	CS_0	CS_1	CS_2	$Diff_1$	$Diff_2$
1	0.070	0.252	0.899	0.182***	0.829***
2	0.419	0.680	0.999	0.261***	0.580***
3	0.637	0.817	1.000	0.180***	0.363***
4	0.860	0.931	1.000	0.072***	0.140***

Notes: Columns 2, 3, and 4 report the mean scores based on CS_0, CS_1, and CS_2,respectively, while Columns 5 and 6 present differences between CS_2 and CS_0 and CS_1 and CS_0 respectively.

Table 3: Grade-wise Mean Scores Across All Indicators

Grade	CS_0	CS_1	CS_2	$Diff_2(CS_2 - CS_0)$	$Diff_1(CS_1 - CS_0)$
1	0.8545	0.9230	0.9978	0.1433***	0.0685***
2	0.8369	0.9131	0.9975	0.1606***	0.0762***
3	0.6998	0.8400	0.9976	0.2977***	0.1402***
4	0.6861	0.8296	0.9965	0.3104***	0.1435***
5	0.7161	0.8508	0.9984	0.2823***	0.1347***
6	0.6658	0.8169	0.9972	0.3314***	0.1511***
7	0.6938	0.8351	0.9975	0.3037***	0.1414***
8	0.7173	0.8481	0.9973	0.2801***	0.1308***
9	0.5722	0.7724	0.9985	0.4263***	0.2002***
10	0.5271	0.7595	0.9970	0.4699***	0.2324***

Notes: Columns 2, 3, and 4 report the mean scores based on CS_0, CS_1, and CS_2, respectively, while Columns 5 and 6 present the differences between CS_2 and CS_0 and CS_1 and CS_0 respectively.

Table 4: Transition Probabilities (CS_0 & CS_1)

	Comparison of CS_0 and CS_1				
Quartile	Q1	Q2	Q3	Q4	Total
Q1	67.86	15.68	16.46	0.00	100.00
Q2	0.00	0.00	95.81	4.19	100.00
Q3	0.00	0.00	7.31	92.69	100.00
Q4	0.00	0.00	0.00	100.00	100.00
Total	1.40	0.32	14.87	83.41	100.00

Notes: Each element $\pi_{kl}^{CS_0,CS_1}$ of the matrix represents the fraction of students in partition k using CS_0 who are in partition l using CS_1.

Table 5: Transition Probabilities (CS_0 & CS_2)

	Comparison of CS_0 and CS_2				
Quartile	Q1	Q2	Q3	Q4	Total
Q1	0.00	0.00	0.00	100.00	100.00
Q2	0.00	0.00	0.00	100.00	100.00
Q3	0.00	0.00	0.00	100.00	100.00
Q4	0.00	0.00	0.00	100.00	100.00
Total	0.00	0.00	0.00	100.00	100.00

Notes: Each element $\pi_{kl}^{CS_0,CS_2}$ of the matrix represents the fraction of students in partition k using CS_0 who are in partition l using CS_2.

Table 6: Directional Rank Mobilities (CS_0 & CS_1)

	Proportion of Students with Score Gains			
Quartile	$\gamma \geq 0.1$	$\gamma \geq 0.2$	$\gamma \geq 0.3$	$\gamma \geq 0.4$
Q1 (N=964)	1.00	0.30	0.14	0.06
Q2 (N=5294)	1.00	1.00	0.08	0.01
Q3 (N=18663)	1.00	0.29	0.00	0.00
Q4 (N=19252)	0.25	0.00	0.00	0.00

Notes: Each element $\theta_{k,\gamma}^{CS_0,CS_1}$ of the matrix represents the proportion of individuals who experience upward mobility relative to the given value of γ and conditional on being in the quartile k of the composite score distribution based on CS_0.

Table 7: Directional Rank Mobilities: (CS_0 & CS_2)

	Proportion of Students with Score Gains					
Quartile	$\gamma \geq 0.2$	$\gamma \geq 0.4$	$\gamma \geq 0.6$	$\gamma \geq 0.8$	$\gamma \geq 0.850$	$\gamma \geq 0.9$
Q1 (N=964)	1.00	1.00	1.00	0.80	0.06	0.01
Q2 (N=5294)	1.00	1.00	0.34	0.00	0.00	0.00
Q3 (N=18663)	1.00	0.33	0.00	0.00	0.00	0.00
Q4 (N=19252)	0.24	0.00	0.00	0.00	0.00	0.00

Notes: Each element $\theta_{k,\gamma}^{CS_0,CS_2}$ of the matrix represents the proportion of individuals who experience upward mobility relative to the given value of γ and conditional on being in the quartile k of the composite score distribution based on CS_0.

References

Ahmad, A. R., Seman, A. A., Awang, M. M., & Sulaiman, F. (2015). Application of multiple intelligence theory to increase student motivation in learning history. *Asian Culture and History*, *7*(1), 210.

Aydin, H. (2019). The effect of multiple intelligence (s) on academic success: A systematic review and meta-analysis. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(12), em1777.

Bhattacharya, D., & Mazumder, B. (2011). A nonparametric analysis of black–white differences in intergenerational income mobility in the united states. *Quantitative Economics*, *2*(3), 335–379.

Buros, O. (1999). The thirteenth mental measurements yearbook: Supplement.

Camanho, A. S., Stumbriene, D., Barbosa, F., & Jakaitiene, A. (2023). The assessment of performance trends and convergence in education and training systems of european countries. *European Journal of Operational Research*, *305*(1), 356–372.

Campbell, L., & Campbell, B. (1999). *Multiple intelligences and student achievement: Success stories from six schools*. ASCD.

Centre, J. R. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD publishing.

Chen, J.-Q., Krechevsky, M., Viens, J., & Isberg, E. (1998). *Building on children's strengths: The experience of project spectrum. project zero frameworks for early childhood education, volume 1.* ERIC.

Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. V. (2007). An introduction to 'benefit of the doubt'composite indicators. *Social indicators research*, *82*, 111–145.

Davis, K., Christodoulou, J., Seider, S., & Gardner, H. E. (2011). The theory of multiple intelligences. *Davis, K., Christodoulou, J., Seider, S., & Gardner, H.(2011). The theory of multiple intelligences. In RJ Sternberg & SB Kaufman (Eds.), Cambridge Handbook of Intelligence*, 485–503.

Douglas, O., Burton, K. S., & Reese-Durham, N. (2008). The effects of the multiple intelligence teaching strategy on the academic achievement of eighth grade math students. *Journal of instructional psychology*, *35*(2).

Ellison, G., & Swanson, A. (2023). Dynamics of the gender gap in high math achievement. *Journal of Human Resources*, *58*(5), 1679–1711.

Forrest-Pressley, D. L., & Waller, T. G. (1984). *Cognition, metacognition, and reading.* Springer-Verlag.

Gardner, H. (1993a). *Frames of mind*. Basic Books.
Gardner, H. (1993b). *Multiple intelligences: The theory in practice*. Basic Books/Hachette Book Group.
Gardner, H. (1995). Reflections on multiple intelligences. *Phi Delta Kappan*, *77*(3), 200–208.
Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources.
Karaduman, G. B., & Cihan, H. (2018). The effect of multiple intelligence theory on students' academic success in the subject of geometric shapes in elementary school. *International Journal of Higher Education*, *7*(2), 227–233.
McDonough, I. K. (2015). Dynamics of the black–white gap in academic achievement. *Economics of Education Review*, *47*, 17–33.
Segovia-González, M. M., & Contreras, I. (2023). A composite indicator to compare the performance of male and female students in educational systems. *Social Indicators Research*, *165*(1), 181–212.
Shearer, B. (1994). *The midas: A professional manual*. MI Research; Consulting.
Shearer, C. B., & Jones, J. A. (1994). The validation of the hillside assessment of perceived intelligences (hapi): A measure of howard gardner's theory of multiple intelligences.
Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 1013–1024.
Snyder, R. F. (1999). The relationship between learning styles/multiple intelligences and academic achievement of high school students. *The high school journal*, *83*(2), 11–20.
Verbunt, P., & Rogge, N. (2018). Geometric composite indicators with compromise benefit-of-the-doubt weights. *European Journal of Operational Research*, *264*(1), 388–401.