Center for Practice Engaged Education Research (C-PEER)

SCHOOL OF EDUCATION & HUMAN DEVELOPMENT

UNIVERSITY OF COLORADO **DENVER**

Research Brief: Relationships among Teaching and Learning Conditions Colorado (TLCC) Survey Results, School and Community Contextual Characteristics, and Student Academic Performance

Provided by:

Kent Anderson Seidel, PhD

Julie Oxenford O'Brian, PhD

Table of Contents

Executive Summary **1**

Introduction **2**

Part 1. Analytic Strategy and Modeling **2**

Purpose and Analytic Strategy *2*

Research Questions *3*

Data Sources and Variables *4*

Screening MAIHDA Variables *10*

MAIHDA Model Sequence and Interpretation *10*

Part 2: Results of Multi-Level Analysis **11**

Student Intersectional Identity *11*

School and Community Context *11*

School Academic Capacity *12*

School Capacity to Support Wellbeing *12*

Model-Based Student Intersectionality Results *12*

Model-Based School-Context Results *13*

Cross-Cutting Conclusions and Cautions *13*

Part 3. Implications for Educational Leadership and Policy **14**

Use Findings as Diagnostic Guidance, Not Causal Proof *14*

Build Coherent Instructional Capacity *14*

Directing Resources *15*

Invest in Staff Stability and Veteran Capacity *15*

Limitations **16**

References **17**

Executive Summary

This research brief summarizes analysis conducted by the Center for Practice Engaged Education Research (C-PEER) focused on the relationships among Teaching and Learning Conditions Colorado (TLCC) survey results, student academic performance, other school capacities and community contextual characteristics using responses from more than 21,000 Colorado educators.

For these analyses we focused on the following student academic performance outcome measures from Colorado Measures of Academic Success (CMAS) annual assessment: Grade 4 English Language Arts scale scores and Student Growth Percentiles (SGPs), and Grade 6 mathematics scale scores and SGPs. We identified potential student, TLCC, other school capacity, and community context predictors of student academic performance through review of extant research, and theoretical frameworks. We then screened and selected related variables available for Colorado schools to use in our analysis. We used correlation and simple and multiple linear and logistic regression analyses (as appropriate) to determine relevant predictors. We only retained variables with a meaningful statistically significant relationship to CMAS results to construct variables to use in Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA). As an analytic model, MAIHDA operationalizes analysis of intersectional influences by nesting individuals within strata defined by combinations of social identities.

The most consistent positive predictors of student academic performance were three TLCC Instructional Practices constructs: *Shared Responsibility for Learning, Meeting Diverse Student Needs,* and *Supportive Conditions*. These constructs formed the clearest organizational conditions associated with stronger school-level aggregate CMAS results. The TLCC constructs *Professional Development* and *Time* had positive bivariate correlations with CMAS results but negative adjusted coefficients in models incorporating other constructs also highly correlated with CMAS results. These reversals are consistent with suppression effects from shared predictor variance due to collinearity (relatedness) of the constructs as well as possibly compensatory allocation (e.g., schools with lower CMAS results may focus on providing more effective Professional Development in response).

Student-identity MAIHDA strata indicated recurring intersections of FRL and IEP status were among the lowest predicted scale-score strata, while non-FRL and non-IEP status was in the highest strata. Similar rankings appeared for both Grade 4 ELA and Grade 6 mathematics outcomes. Community-context MAIHDA strata indicated strong differences in student academic performance are associated with neighborhood affluence, linguistic access needs, and local educational capital (represented by the proportion of adults with a bachelor's degree or higher within the school's county). School-capacity strata associated with both high-performing and low-performing schools highlighted combinations of instructional conditions, teacher experience, school climate, classroom disruption, staffing ratios, and concentrated enrollment context. As a result, we recommend policy emphasize equity-weighted resources, coherent instructional capacity, support-staffing and disruption indicators, staff stability, structurally framed responses to segregation, and context-sensitive use of scale scores and growth measures.

Introduction

This research brief presents a consolidation of three connected strands of work conducted by the Center for Practice Engaged Education Research (C-PEER) to investigate the relationships among Teaching and Learning Conditions Colorado (TLCC) survey results, student academic performance, other school capacities and community contextual characteristics. It includes the screening and narrowing of variables in the development of analytic models to examine the relevance of TLCC constructs as predictors of student academic performance; interpretation of multi-level models including TLCC constructs, other school capacities, community contexts and intersectional student identities as predictors of student academic performance; and the implications of the findings for educational leadership and policy. The organization of the brief follows the analytic sequence from evidence development to interpretation.

We present a comprehensive record of the research underlying the *TLCC Student Academic Performance User Guide* (Oxenford O'Brian, Anderson Seidel & Mitchell, 2026) and a research manuscript focused on intersectional student identities, school contexts, and student academic outcomes (Anderson Seidel & Oxenford O'Brian, 2026). Material from those prior publications is incorporated to clarify the research design, data sources, methods, findings, or interpretation. Our findings are most useful as a diagnostic map of intersecting student needs, school capacities, and community conditions. They should not be treated as a causal ranking of programs or of leadership and teacher actions.

Part 1. Analytic Strategy and Modeling

Purpose and Analytic Strategy

We designed analyses to identify student-identity characteristics, school characteristics, TLCC constructs, and community-context variables associated with student academic outcomes (measured through CMAS). Our work proceeded in two related stages.

1. *Variable screening and narrowing.* We reviewed theory, existing conceptual frameworks, and the extant literature to identify candidate predictors of student academic performance. We used correlation and regression analysis to examine the relationship of each with CMAS results. Variables that did not demonstrate both an adequate relationship with the outcomes and statistical significance were not retained to construct variables for the MAIHDA model.
2. *Intersectional MAIHDA construction and examination*. We combined retained variables into one student-level intersectional identity variable and three school-level contextual or capacity variables. This analysis results in ranked strata predicted CMAS scores to identify recurring characteristics of the highest- and lowest-scoring combinations.

This approach is empirically informed rather than purely data driven. We used theory and prior research to nominate candidate variables, and analysis with CMAS results to determine which variables displayed sufficient empirical relevance to use in the MAIHDA model.

Conceptual and methodological foundations

Ecological understanding of student learning in which students and educators work within schools that are themselves embedded in district and community systems guided our analysis (Bronfenbrenner & Morris, 2006). This framing directs attention toward both student-level social locations and malleable organizational conditions that may shape opportunities to learn.

Intersectionality provided the conceptual basis for treating race or ethnicity, gender, language proficiency, ability, and socioeconomic status as simultaneous rather than independent dimensions of students' school experience. Identity categories are not simply additive, rather their combinations can be associated with distinctive patterns of belonging, expectations, access to rigorous learning, and support that broad subgroup averages may conceal (Crenshaw, 1989; Collins, 1990, 2009, 2019; OECD, 2023; Tefera, Powers, & Fischman, 2018; Zhang & Gao, 2024).

Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy, or MAIHDA, operationalizes this perspective by nesting individuals within intersectional strata defined by combinations of social identities. The MAIHDA model estimates both heterogeneity within strata and differences between strata while producing more stable and interpretable estimates than a flat regression model containing many interaction terms (Evans et al., 2018; Evans, Leckie, & Merlo, 2020; Keller et al., 2023; Merlo, 2018). In this project, we extended the same logic to schools as institutionally situated contexts shaped by interacting demographic, organizational, community, and resource conditions.

Our analysis was not intended to treat identities as deficits or schools as neutral backdrops. Rather we examined how patterned combinations of student characteristics and institutional conditions are associated with academic outcomes, with the purpose of locating opportunity structures that may reinforce or mitigate educational inequality.

Research Questions

We addressed four linked questions:

- How much variability in student academic achievement and growth is located at student, school, and district levels?
- What additional understanding of achievement differences is provided by modeling student demographic characteristics as intersectional rather than only as separate additive predictors?
- Which teacher-reported school climate and instructional conditions are associated with variation in CMAS outcomes?
- How do school, district, and community contextual variables contribute to intersectional differentiation and the discriminatory accuracy of the models?

Table 1

Focal Outcomes and Data

Outcome	N
Grade 4 ELA scale score	40,908
Grade 4 ELA SGP	36,807
Grade 6 mathematics scale score	34,786
Grade 6 mathematics SGP	32,038

We conducted school-level analyses using CMAS results from 757 public elementary and middle schools. We only used schools with available TLCC responses in modeling predictors of student academic performance, so the effective sample size varied by outcome and specification.

Data Sources and Variables

In this study we integrated 2024 statewide assessment records, student demographic records, the 2024 Teaching and Learning Conditions Colorado educator survey results, school and district administrative data, National Center for Education Statistics data, and American Community Survey indicators. We selected the 2024 assessment results because they aligned with the biennial 2024 TLCC administration, allowing us to link school learning-condition measures temporally with student outcomes.

Outcome measures

We analyzed CMAS scale scores and Student Growth Percentiles (SGPs) for both focal grade-subject combinations, Grade 4 English Language Arts (ELA) and Grade 6 mathematics, chosen because are recognized as key grade levels for student academic performance in the subject areas. Scale scores represent current achievement on the assessment scale. SGPs compare a student's growth with that of academically similar peers using the Colorado Growth Model's quantile-regression methods (Betebenner, 2009; Castellano & Ho, 2013; Colorado Department of Education, 2019). Because SGP was designed to reduce the influence of prior achievement and some outside-school factors, we expected smaller predictor relationships. We retained both outcomes to distinguish accumulated achievement from relative growth.

For the focal analytic data sample, the Grade 4 ELA scale-score mean was 742.2 with a standard deviation of 37.9, and the Grade 6 mathematics scale-score mean was 735.3 with a standard deviation of 31.6. Means of Student Growth Percentiles, SGP, were approximately 50.4 with standard deviations of approximately 28.9 for both outcomes as expected for percentile-ranked data.

TLCC survey sample and measurement evidence

The statewide 2024 TLCC administration resulted in 30,424 responses from educators across school types and grade spans. We used 21,108 teacher responses from traditional public schools serving elementary and middle grades with CMAS results. We aggregated teacher TLCC results to the school level after a diagnostic review of response rates and representation.

The TLCC items grouped into constructs as expected in diagnostic factor analyses. We reported principal component analysis with varimax rotation, a Kaiser-Meyer-Olkin adequacy statistic above .90, Bartlett's sphericity test at $p < .002$, and Cronbach's alpha of at least .80 for each focal construct. These diagnostic results supported use of the school-level construct scores as measures of teaching and learning conditions. We tested the following TLCC constructs: Professional Climate, Staff Evaluation, Staff Leadership, Student Behavior, Student Supports, Shared Responsibility for Learning, Meeting Diverse Student Needs, Supportive Conditions, Professional Development, Time, and Support for Educators' Own Wellbeing.

CMAS outcomes varied meaningfully across schools, based on variance decomposition, and to a smaller degree across districts. Contextual clustering was greater for Grade 4 ELA scale scores than for Grade 4 ELA growth. School-level variation remained substantial for Grade 6 mathematics growth. These findings supported using multilevel and MAIHDA approaches rather than treating the outcomes as exclusively student-level phenomena.

Table 2
Evidence supporting multilevel and MAIHDA analyses

Outcome	Between-school variance	Between-district variance
Grade 4 ELA scale score	16.6%	6.3%
Grade 4 ELA SGP	8.7%	2.3%
Grade 6 mathematics scale score	16.5%	6.1%
Grade 6 mathematics SGP	14.0%	5.5%

Variation in the TLCC results also supported a multilevel approach. Across the teacher survey constructs and summary items, school-level VPCs ranged from approximately .04 to .23 and district-level VPCs from approximately .02 to .06. School-level variation was notable for Student Behavior, Professional Climate, Staff Leadership, and the three Instructional Practice constructs. These results indicated that the teaching and learning conditions measured by TLCC constructs were not solely individual perceptions but also reflected shared organizational contexts.

For the TLCC outcome item "I would recommend this school as a good place for students to learn," differences among schools accounted for approximately 22.7% of variance and differences among

districts accounted for 3.7% of the variance. Differences among schools accounted for approximately 23.2% of variance in the Student Behavior construct, while differences among districts accounted for 3.6%. Similarly, differences among schools accounted for 15.2% of the variance in Shared Responsibility for Learning and 13.8% in Supportive Conditions, compared with 2.3% and 5.7% among districts, respectively. For Meeting Diverse Student Needs, differences among schools accounted for 12.9% of the variance, while differences among districts accounted for 5.4%.

Screening and narrowing the TLCC constructs

Very large student samples made it possible for small associations to reach statistical significance. Therefore, our narrowing procedure needed to emphasize recurring direction, relative magnitude, consistency across outcomes, conceptual relevance, and incremental explanatory value rather than statistical significance alone.

Correlational screening

At the school-aggregate level, three TLCC instructional practice constructs were most positively correlated (.36 to .42) with mean CMAS scale scores: Meeting Diverse Student Needs, Shared Responsibility for Learning, and Supportive Conditions. The TLCC Student Behavior and Support for Staff Wellbeing constructs showed positive bivariate relationships, while Professional Development and Time had smaller positive zero-order correlations.

TLCC construct correlations with SGP were generally smaller than correlations with scale scores. This distinction is substantively important because scale scores reflect accumulated achievement and prior opportunity to learn, while SGP is a relative growth measure. The weaker construct-SGP associations are therefore consistent with the broader finding that contextual variables accounted for less variation in some growth outcomes, especially Grade 4 ELA growth.

Table 3

Multivariable TLCC models

ELA construct	Standardized beta	Mathematics construct	Standardized beta
Supportive Conditions	.451	Shared Responsibility for Learning	.400
Meeting Diverse Student Needs	.335	Supportive Conditions	.340
Shared Responsibility for Learning	.327	Meeting Diverse Student Needs	.313
Time	-.319	Professional Development	-.314
Professional Development	-.238	Time	-.294

The summaries in Table 3 indicate adjusted *R*-squared values of approximately .320 for mean ELA scale scores and .335 for mean mathematics scale scores. All coefficients were statistically significant at $p < .001$. The three instructional-practice constructs represented the strongest stable positive indicators. The negative partial coefficients for Time and Professional Development necessitated examining potential "suppression effects" in interpreting TLCC construct-level findings in the regression models.

What does suppression mean in these analyses

A zero-order correlation asks whether schools with higher values on a construct tend to have higher CMAS scores. Multiple regression asks whether the portion of that construct that is unique after controlling for all other predictors is related to CMAS. These are different estimated parameters.

A common suppression warning sign is a positive zero-order correlation paired with a negative adjusted coefficient. The coefficient then describes the residual portion of the predictor, not the original construct in its ordinary, system-level form.

Table 5

Where suppression appeared

Construct	Zero-order relationship	Adjusted ELA beta	Adjusted mathematics beta
Professional Development	Positive, about .20	-.238	-.314
Time	Positive, about .11 to .12	-.319	-.294

Professional Development and Time were positively associated with CMAS overall but received negative coefficients after Shared Responsibility for Learning, Meeting Diverse Student Needs, Supportive Conditions, and other overlapping constructs were added to the model. Similar reversals appeared for Student Supports, Professional Climate, and selected survey items in expanded specifications.

Overlapping school-organization constructs

The TLCC constructs are conceptually distinct but interconnected in practice. Schools with strong instructional conditions often also report better professional learning, collaboration, climate, student support, and use of staff time. Once entered together, each coefficient is based on a narrow residual component. The positive relationship may reside largely in shared school-capacity variance, while the remaining unique variance acquires a different coefficient representing its additional amount of explanatory power.

Compensatory allocation and underlying need

Schools experiencing academic or organizational challenge may add coaching, professional development, intervention programs, and student supports. Conditional on similar instructional-practice ratings, the school reporting more support activity may also be the school facing greater unmeasured

need. The negative coefficient can therefore reflect compensatory response or confounding by indication rather than program harm.

Reciprocal suppression

A negatively signed predictor can remove variance from another predictor that is unrelated to CMAS, thereby increasing the positive coefficient of the other predictor. Suppression may therefore partly sharpen the positive coefficients for the instructional-practice constructs. Their strength should still be interpreted considering shared variance rather than as fully independent effects.

Multicollinearity and coefficient sensitivity

Several TLCC predictors had Variance Inflation Factor (VIF) values in the approximate range of 3 to 5, with some item-level models reaching or exceeding 5. Values of 1 to 5 are generally considered acceptable in practice, while larger values recommend further analysis of collinearity. VIF measures how much the variance of a variable's estimated regression coefficient is inflated because of correlation with other independent variables in the model. This does not automatically invalidate the models, but it indicates enough overlap to make partial coefficients sensitive to model composition. Standard errors can increase, signs may reverse, and small changes in the predictor set can redistribute shared explanatory variance.

Table 6

What the findings support and do not support

Supported interpretation	Unsupported interpretation
The instructional-practice constructs had stable positive bivariate and multivariable relationships with scale scores.	More professional learning causes lower CMAS performance.
Professional Development and Time were positive in bivariate correlations, but negative in conditional regression modeling.	Giving teachers more time harms students.
Shared predictor variance is central to the model.	Support for multilingual learners or family communication lowers achievement.
The TLCC constructs operate as an interconnected organizational system.	Negative partial coefficients justify reducing programs or supports.

Implications of Suppression for variable selection and reporting

Several TLCC constructs, particularly Professional Development and Time, had positive bivariate associations with mean CMAS performance but negative coefficients in multivariable models. These sign reversals are consistent with suppression arising from substantial shared variance among the TLCC constructs. The negative coefficients should therefore be interpreted as relationships involving the

predictors' residual components after adjustment for overlapping dimensions of instructional practice and school climate, rather than as evidence that greater professional learning or time adversely affects achievement. This pattern may also reflect compensatory allocation, whereby schools facing greater academic or organizational challenges provide additional professional development or supports.

Suppression effects reinforce the need to not use the sign of one adjusted coefficient as the sole retention criterion. Selection and interpretation should consider theoretical relevance, zero-order relationships, consistency across outcomes, incremental contribution to explanation of variance, coefficient stability, collinearity of variables, and whether a variable represents a resource, an underlying need, or a response to that need.

The suppression findings rule out a simplistic policy response in which resources with positive coefficients are expanded and resources with negative coefficients are reduced. Professional Learning and Time support, as represented in the TLCC construct items, should instead be evaluated for coherence, implementation, and alignment.

- Alignment with identified student and teacher needs
- Connection to curriculum and instructional routines
- Reinforcement through usable coaching and feedback
- Protected implementation time
- Providing new approaches with sufficient duration to produce change
- Evaluation of implementation quality and student outcomes

The appropriate response is greater coherence and stronger evaluation, not automatic disinvestment.

Item-level narrowing

We used item analyses to identify survey elements that carried the most information within the broader constructs. Note that closely related items compete for shared variance, as a result construct-level interpretations are more stable than individual item-level partial coefficients. Items from the three instructional practices TLCC constructs, in order of CMAS explanatory power are:

- *Shared Responsibility for Learning*: opportunities to learn from other teachers; staff accountability for every child's academic growth; individualized family engagement; interactive strategies for families; and continuous improvement of instruction.
- *Meeting Diverse Student Needs*: social and emotional supports; curriculum alignment with diverse academic needs; support for multilingual learners; gifted education; and support for students with disabilities.
- *Supportive Conditions*: support for trying new instructional strategies; substitute availability; reasonable class size or caseload; instructional resources; adequate workspace; and autonomy within coherent expectations.

The TLCC school-quality general reflection item: "I would recommend this school as a good place for students to learn" emerged as a consistent predictor of student academic performance, but it should be understood as a broad summary perception rather than a discrete mechanism.

Screening MAIHDA Variables

Student-level demographics

Student-level regressions supported retaining the following student demographic variables for the MAIHDA model: gender, race/ethnicity, FRL status, English-language proficiency or status, and IEP status. Our analysis identified IEP and FRL status as the strongest predictors of adjusted scale-score differences, with race or ethnicity and English-language proficiency also demonstrating meaningful effects. The effects of Gender differed across outcomes and specifications, supporting our treatment of it as part of an intersection rather than as a uniform marginal effect.

School capacity and staffing

We considered the following school capacity variables for the model: overall school effectiveness categories, Colorado Teacher Quality Standard 2 and Quality Standard 3 school-wide mean proficiency, mean teacher experience, pupil-teacher ratio, students per counselor, students per psychologist, students per social worker, and the proportion of serious classroom disruptions. Stronger effectiveness ratings and greater teacher experience were generally associated with better academic performance. More disruptions and larger student-to-support-staff ratios were generally associated with lower performance. Pupil-teacher ratio was less stable across specifications.

Community context

We considered the following community context variables: school-neighborhood Income-to-Poverty Ratio, county bachelor's degree attainment, the percentage of homes in which English was not the primary language, county income and poverty categories, and related demographic descriptors. Income-to-Poverty Ratio was among the most consistent positive predictors of CMAS scale scores. County educational attainment was also a positive predictor. Linguistic-context indicators were generally negative in our initial scale-score screening but less stable in some fully adjusted SGP models.

MAIHDA Model Sequence and Interpretation

Analysis using the student-level MAIHDA first assigns students to intersectional strata based on real-world combinations of identity characteristics. This baseline random-intercept model estimates between-stratum heterogeneity. An adjusted model then adds the individual identity variables used to define the strata. Comparing baseline and adjusted models makes it possible to quantify how much between-stratum differentiation is captured by additive individual characteristics and how much residual intersectional variation remains. See Anderson Seidel & Oxenford O'Brian (2026) for detailed descriptions and findings related to MAIHDA analyses of student intersectionality and CMAS outcomes.

We followed a parallel process for school-level analyses. We screened candidate school, district, and community variables for their relationship to student academic performance, independence, conceptual fit, and collinearity. We categorized retained indicators using research precedents, policy-relevant thresholds, professional standards, or distribution-based cut point, combining them into *Community Context*, *Capacity to Support Academic Success*, and *Capacity to Support Wellbeing* strata.

MAIHDA generated summaries use the Variance Partition Coefficient, or VPC, to express the share of total variance attributable to stratum membership, and the Proportional Change in Variance, or PCV, to express the reduction in between-stratum variance after measured component variables enter the adjusted model. These metrics describe discriminatory accuracy and the degree to which measured variables account for patterned stratum differences (Evans et al., 2018; Evans et al., 2024; Keller et al., 2023).

Table 4
Final MAIHDA variable compositions

MAIHDA variable	Core dimensions
MV1: Student identity	Gender; race or ethnicity; FRL status; English-language proficiency or status; IEP status
School wellbeing support capacity	Meeting diverse student needs; Quality Standard 2 capacity; classroom disruption; counselor and other support-staffing ratios
School academic capacity	Instructional-practice effectiveness; Quality Standard 3; teacher experience; school climate or learning-quality rating; instructional size; concentrated enrollment context
School and community context	Neighborhood Income-to-Poverty Ratio; linguistic isolation or access needs; local adult educational capital

Part 2: Results of Multi-Level Analysis

Student Intersectional Identity

The student-identity strata emerged as one of the clearest and most consistent predictors of Grade 4 ELA and Grade 6 mathematics scale scores. The lowest predicted strata included disproportionately high numbers of students eligible for FRL and identified with an IEP, often in combination with Black, Native, Latino, or multiracial identity. In contrast, the highest predicted strata consisted primarily of non-FRL, no-IEP combinations and often included Asian, White, or multiracial students.

The same Female / Native / FRL / English proficient / IEP combination was the lowest-ranked stratum for both Grade 4 ELA and Grade 6 mathematics scale scores. The same Male / Asian / Non-FRL / Not fully English proficient / No IEP combination was the highest-ranked stratum for both outcomes, with predicted means near 769. The near replication across subjects and grades suggests accumulated educational advantage and disadvantage rather than a subject-specific pattern.

School and Community Context

The lowest community-context strata combined neighborhoods near or below the poverty line with high linguistic-access needs and limited or moderate local educational capital. The highest strata generally combined affluence with lower linguistic isolation and moderate to high educational capital. The

approximate lowest-to-highest predicted-mean differences were 41.6 points for Grade 4 ELA and 38.4 points for Grade 6 mathematics.

Economic resources were the clearest recurring predictor of CMAS variability. Linguistic isolation and educational capital further differentiated strata, but neither appeared to fully offset the broader association with neighborhood affluence. Exact ranks should be interpreted cautiously because some extreme strata had very small n sizes.

School Academic Capacity

The lowest academic-capacity strata commonly combined low or moderate effectiveness on at least one school-capacity dimension, experienced or mid-career rather than veteran teacher composition, and agreement rather than strong agreement on the positive TLCC school reflection item. The highest strata commonly combined veteran teacher composition, strongly positive school ratings, and at least moderate instructional effectiveness.

The approximate lowest-to-highest predicted-mean differences were 60.2 points for Grade 4 ELA and 57.9 points for Grade 6 mathematics. No single indicator uniformly defined the highest strata. The pattern is intersectional -- experience, school climate, instructional conditions, and segregation context operated in combination.

School Capacity to Support Wellbeing

For Grade 6 mathematics, the lowest strata commonly combined low or moderate effectiveness in meeting diverse student needs, “Agree” rather than “Strongly agree” on the TLCC school reflection item, severe shortages of counselors and psychologists in district, and high cumulative incidences of classroom disruptions. The highest strata more often included high or moderate effectiveness, strong agreement, low disruption, and better counselor staffing, although some shortages in other support roles remained. The predicted range was approximately 54.4 points.

Model-Based Student Intersectionality Results

In the baseline student-identity MAIHDA models, 36% of Grade 4 ELA scale-score variance and 32% of Grade 6 mathematics scale-score variance were attributable to the intersectional strata. After the additive identity variables were included, the Variance Partition Coefficients (VPCs) declined to approximately 10% for ELA and 11% for mathematics. The corresponding Proportional Change in Variance (PCV) values were 82% and 73%. Thus, measured individual characteristics accounted for most between-stratum differentiation, but nontrivial residual intersectional variance remained beyond the additive factors.

Translated to the assessment scale, the baseline Grade 4 ELA stratum variance corresponded to a stratum-level standard deviation of about 23 points, or approximately 0.60 of the test standard deviation. After adjustment, residual stratum differentiation was approximately 10 points. For Grade 6 mathematics, the baseline stratum-level standard deviation was about 18 points, or approximately 0.58 of the test standard deviation, with adjusted residual differentiation of roughly 9 to 10 points.

These estimates complement but should not be confused with the ranked high-low strata. The variance components summarize average differentiation across all modeled strata, while the ranked tables identify the compositions at the extremes. Both are descriptive and correlational.

Model-Based School-Context Results

Community Context strata showed substantial baseline differentiation, with VPCs of 39% for Grade 4 ELA and 36% for Grade 6 mathematics. After neighborhood Income-to-Poverty Ratio, linguistic context, and adult educational attainment were entered, VPCs fell to 1% and 2%, with PCVs of 99% and 97%. The observed between-stratum differences were therefore largely captured by the measured socioeconomic and linguistic indicators.

Academic Capacity strata accounted for 23% of baseline Grade 4 ELA variance and 20% of baseline Grade 6 mathematics variance. After adjustment for the component organizational and instructional measures, the VPCs fell to 2% and 3%, with PCVs of 92% and 87%. Most differentiation was captured by the independently measured capacity indicators, including instructional conditions and teacher experience.

Wellbeing Capacity retained the strongest residual intersectional signal. Baseline VPCs were 37% for Grade 4 ELA and 39% for Grade 6 mathematics. After adjustment, VPCs remained approximately 14% and 18%, with PCVs of 75% and 70%. The persistence of stratum-level variance suggests that the combination of inclusive learning environment, support access, and behavioral load carried information not fully reduced to the separate component variables.

Taken together, these results distinguish domains in which intersectional differentiation was almost entirely explained by measured component indicators from the wellbeing domain, in which meaningful contextual differentiation remained. They support understanding schools as institutionally situated intersectional contexts, while also showing that many of the operative conditions are measurable and potentially monitorable.

Cross-Cutting Conclusions and Cautions

The following findings emerge across the analysis described above:

- Scale scores were more context sensitive than SGP for Grade 4 ELA, while Grade 6 mathematics SGP retained substantial school-level effects.
- The three instructional-practice TLCC constructs were the most stable positive predictors of Student Academic Performance (CMAS scale scores in Grade 4 ELA and Grade 6 mathematics).
- FRL and IEP status were central to the lowest student-identity intersections across both Grade 4 ELA and Grade 6 math scale-score outcomes.
- Community affluence was a dominant contextual factor, with linguistic access and adult educational capital further differentiating strata across student academic performance outcomes.
- Positive school climate, low aggregate levels of classroom disruptions, more veteran staffing, and stronger instructional capacity mattered in combination for explaining CMAS data variability.

- Some extreme strata had n values from 1 to 8. Their predicted means are model-based and should not be treated as precise observed subgroup averages.
- Rank ordering does not establish that each adjacent stratum is statistically distinguishable. Overlapping confidence intervals require caution.
- Large samples make small effects statistically significant. Practical magnitude, incremental R-squared, theoretical relevance, and stability should guide interpretation.
- The analyses are observational and intersectional. They identify patterned differences in predicted outcomes but do not establish causality.

Part 3. Implications for Educational Leadership and Policy

Use Findings as Diagnostic Guidance, Not Causal Proof

Policy implications from these analyses relate to targeting, sequencing, and evaluating educational supports. Our analyses identified promising levers for change as well as structural inequities but were not designed or able to establish causal effects. Identifying potential causal factors requires longitudinal analysis and analysis across a broader representation of grade level outcomes and school/community contexts. As a result, leaders should use these results to frame questions, identify combinations of need, and design stronger tests of specific interventions.

Table 7

Use scale scores and Student Growth Percentile (SGP) outcome data for different purposes

Pattern	Leadership and policy question
Low Scale Scores, high SGP	Is the school accelerating learning while overcoming large, accumulated learning gaps?
Low Scale Scores, low SGP	What intensive instructional and/or contextual supports are needed?
High Scale Scores, low SGP	Are initially advantaged students progressing adequately?
High Scale Scores, high SGP	Which conditions may be transferable after accounting for context?

Scale scores provide data about current achievement and accumulated opportunity. SGPs inform relative growth but do not eliminate contextual influence. Neither should be used alone for high-stakes decisions.

Build Coherent Instructional Capacity

The strongest recurring TLCC findings support schools engaging in a coordinated instructional strategy built around collective responsibility, capacity to meet diverse needs, and supportive instructional conditions.

This interpretation is consistent with research showing that academic performance reflects interacting school, classroom, family, and student factors, and that school learning climate, supportive

relationships, coherent instruction, and organizational conditions represent meaningful areas for improvement (Berkowitz et al., 2017; Roorda et al., 2017; Tan, Gao, & Shi, 2020; Tan, Dimmock, & Walker, 2024). Our evidence does not imply that a single initiative will produce uniform gains. Rather, effective schools organize teaching, learning, and student support through connected systems.

Schools appear to have stronger outcomes when supporting instructional practices, staff capacity, positive learning conditions, and community resources operate together. Policy should build those systems, direct additional resources toward intersecting need, and evaluate specific interventions before drawing causal conclusions.

Directing Resources

Our findings from MAIHDA analyses of student and school/community intersectional variables argue against relying on a single need indicator to establish priorities for additional resources. Both policy makers and educational leaders should acknowledge that needs can intensify at intersections, such as poverty plus disability, poverty plus multilingual-learning needs, or high student need combined with limited school support capacity. This must inform improvement efforts. Resources should be allocated according to the concentration and intersection of educational needs, not simply the presence of one demographic category. Student identities should be framed as locations within unequal opportunity structures, not as causes of lower achievement.

Neighborhood and community resource contexts are relevant to educational practice and policy. Strong community-context differences imply that schools do not operate under equivalent conditions. Accountability systems based only on unadjusted scale scores can conflate student learning with accumulated opportunity, neighborhood resources, residential segregation, and local access to services. This means policy makers and education leaders should:

- Use multiple indicators, including scale scores, growth, opportunity-to-learn measures, staffing capacity, school climate, access to advanced coursework, and contextual diagnostics.
- Make adjustments that consider contextual differences to improve inferences and target support, not to lower expectations for students.
- Coordinate education policy with housing, transportation, broadband, health and mental-health services, extended learning, early childhood, and family language access.

Invest in Staff Stability and Veteran Capacity

Our analysis linked better student academic outcomes to more experienced teachers, positive educator ratings of the school as a good place to learn, and school conditions that support instruction, facilitate shared responsibility for learning, and meet the diverse needs of students. Policy and leadership responses can include retention incentives, improved working conditions, teacher-leader roles, mentoring, strong induction, stable principal support, and career pathways that allow accomplished teachers to remain close to instruction. See the C-PEER *Research Brief: Relationships among Teaching and Learning Conditions Colorado (TLCC) Survey Results, Teacher Retention and School Contextual Characteristics* for details on our in-depth analyses related to teacher retention and turnover. The goal is not a zero-sum movement of experienced educators. It is to make high-need schools places where

accomplished teachers can remain, lead, and succeed. This emphasis is also consistent with research linking working conditions and organizational context to teacher satisfaction, turnover, and student achievement (Johnson, Kraft, & Papay, 2012; Kraft, Marinell, & Yee, 2016; Ladd, 2009). In this report, those studies provide interpretive context for the observed associations; they do not convert the present observational findings into causal estimates.

Limitations

Our findings are based on data aligned across assessment, survey, administrative, and community sources in 2024. Temporal alignment strengthens interpretation but does not establish causal ordering, nor are causal conclusions appropriate.

TLCC measures are educator reports aggregated to the school level. C-PEER analysis supports the use of TLCC constructs, but these data remain perceptions of conditions. School-level sample sizes vary across models because of differences in data availability and in some cases, the requirement of complete case data (i.e., some analysis steps required that survey respondents answer all questions within the group being examined).

The categorization required for MAIHDA makes complex continuous data interpretable as strata but can reduce within-category information. Rarer strata containing the highest and lowest outcomes are stabilized through the multilevel model; however, exact rank positions remain uncertain. Confidence intervals, variance components, and recurring high-low combinations of the variables comprising the MAIHDA intersectional representation should receive more emphasis than an isolated rank.

The analyses reported here focused on two grade-subject combinations selected to represent elementary ELA and middle-grade mathematics. Replication across grades, subjects, years, and post-pandemic cohorts over time is necessary before treating any pattern as broadly stable.

The TLCC User Guides and research articles are translations of portions of this work for practitioner and scholarly audiences, respectively. Recommendations and theoretical emphases in those publications should not be read as additional independent evidence beyond the analyses documented here.

References

Anderson Seidel, K., & Oxenford O'Brian, J. (in submission, 2026). *How Intersectional Student Identities and School Contexts Shape Academic Performance.* Contact authors for pre-publication copy until published.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. Educational Measurement: Issues and Practice, 28(4), 42–51. https://doi.org/10.1111/j.1745-3992.2009.00161.x

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In Handbook of Child Psychology, Vol. 1. Castellano, K. E., & Ho, A. D. (2013). A practitioner's guide to growth models. Council of Chief State School Officers.

Collins, P. H. (1990). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Unwin Hyman.

Collins, P. H. (2009). Black feminist thought. Routledge.

Collins, P. H. (2019). Intersectionality as critical social theory. Duke University Press.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex. University of Chicago Legal Forum, 139–168.

Evans, C. R., Williams, D. R., Onnela, J.-P., & Subramanian, S. V. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. Social Science & Medicine, 203, 64–73. https://doi.org/10.1016/j.socscimed.2017.11.007

Evans, C. R., Leckie, G., & Merlo, J. (2020). Multilevel versus single-level regression for the analysis of multilevel information: The case of quantitative intersectional analysis. Social Science & Medicine, 245, 112499. https://doi.org/10.1016/j.socscimed.2019.112499

Evans, C. R., Leckie, G., Subramanian, S. V., Bell, A., & Merlo, J. (2024). A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy. SSM - Population Health, 26, 101664. https://doi.org/10.1016/j.ssmph.2024.101664

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools. Teachers College Record, 114(10), 1–39.

Keller, L., Lüdtke, O., Preckel, F., & Brunner, M. (2023). Educational inequalities at the intersection of multiple social categories. Educational Psychology Review, 35, 31. https://doi.org/10.1007/s10648-023-09733-5

Kraft, M. A., Marinell, W. H., & Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement. American Educational Research Journal, 53(5), 1411–1499. https://doi.org/10.3102/0002831216667478

Ladd, H. F. (2009). Teachers' perceptions of their working conditions: How predictive of policy-relevant outcomes? CALDER Working Paper No. 33.

Merlo, J. (2018). Multilevel analysis of individual heterogeneity and discriminatory accuracy within an intersectional framework. Social Science & Medicine, 203, 74–80. https://doi.org/10.1016/j.socscimed.2017.12.026

Organisation for Economic Cooperation and Development. (2023). Intersectionality in education: Rationale and practices to address the needs of students' intersecting identities. OECD Education Working Paper No. 302. https://doi.org/10.1787/dbb1e821-en

Tefera, A. A., Powers, J. M., & Fischman, G. E. (2018). Intersectionality in education: A conceptual aspiration and research imperative. Review of Research in Education, 42(1), vii–xvii. https://doi.org/10.3102/0091732X18768504

Zhang, A. Y., & Gao, G. (2024). Intersectionality in educational psychology and curriculum studies: A critical review. Extensive Reviews, 4(1), 1–14. https://doi.org/10.21467/exr.4.1.5387