Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Student Dispositions and Instructional Practices as Predictors of Science Achievement: A Multilevel Analysis of PISA 2022 Data from Kosovo

Fidelis Chinedu Onwunyili

fconwuny@iu.edu

Indiana University

Emmanuel Atiatorme

Indiana University

John Amoako Mensah

Indiana University

Adam Maltese

Indiana University

Research Article

Keywords: science achievement, PISA 2022, Kosovo, curiosity, creativity, adaptive instruction, multilevel modeling, WeMix, plausible values, multiple imputation

Posted Date: July 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10028310/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Science achievement in persistently low-performing education systems is shaped by a complex interplay of student-level motivational dispositions and school-level instructional practices whose relative contributions remain poorly understood. This study examined the direct effects of curiosity and creativity at the student level and adaptive instruction and feedback practices at the school level on science achievement among Kosovo students participating in PISA 2022. Grounded in Bronfenbrenner's Ecological Systems Theory a weighted two-level multilevel model was estimated using the WeMix package (Bailey et al., 2023) incorporating both student-level (W_FSTUWT) and school-level (W_SCHGRNRABWT) sampling weights. Plausible values were properly handled using Huang's (2024) mixPV wrapper function which estimates models separately across all ten plausible values and pools results using Rubin's rules. Missing data on student dispositional variables arising from PISA's rotated booklet design were addressed through multiple imputation using the mice package (Van Buuren & Groothuis-Oudshoorn, 2011) yielding a final analytic sample of 4,891 students nested within 100 schools. The intraclass correlation of .324 confirmed substantial between-school clustering justifying the multilevel approach. Model comparison using mean AICbar values identified Model 2 as the final model. At the student level curiosity was the strongest dispositional predictor of science achievement ($\beta = .084$, $p < .001$) followed by creativity ($\beta = .046$, $p = .005$). At the school level adaptive instruction was the strongest overall predictor in the model ($\beta = .192$, $p = .005$) explaining 19.1% of between-school variance and exceeding the contribution of all student-level predictors. Feedback practices was not significant ($\beta = .022$, $p = .755$) and cross-level interactions between student dispositions and instructional practices were not significant. The final model explained 2.7% of within-school variance, 19.1% of between-school variance, and 8.0% of total variance in science achievement. Findings establish adaptive instruction as the most powerful lever for reducing between-school science achievement inequality in Kosovo and highlight curiosity and creativity as meaningful student-level resources that support science performance across diverse school environments. Implications for teacher professional development and science education policy in Kosovo and similarly challenged systems are discussed.

Introduction

Scientific literacy has emerged as a cornerstone competency for informed citizenship and productive participation in knowledge-driven economies. Among the instruments used to assess scientific literacy at scale, the Programme for International Student Assessment (PISA) stands out for its capacity to generate internationally comparable evidence on how effectively education systems develop students' ability to engage with scientific concepts, interpret empirical evidence, and apply scientific reasoning to real-world problems (OECD, 2023). Results from the most recent PISA 2022 cycle continue to reveal substantial disparities in science performance across participating countries, drawing attention to the structural, instructional, and motivational factors that either support or impede students' scientific development (OECD, 2023; OECD, 2024).

A growing body of research situates science achievement within a broader ecological framework where outcomes reflect the interplay between learner characteristics and the instructional environments where learning unfolds. At the student level, motivational dispositions such as curiosity and creativity have attracted increasing attention as student dispositional resources that shape how students engage with scientific content, approach novel problems, and sustain effort through cognitive challenge (Jirout & Klahr, 2012; von Stumm et al., 2011; Zabelina & Robinson, 2010). Curiosity, in particular, has been conceptualized as a fundamental driver of exploratory learning behavior, enabling students to move beyond surface-level understanding toward deeper conceptual engagement with scientific ideas (Jirout & Klahr, 2012). Creativity, meanwhile, supports the flexible hypothesis generation and alternative reasoning that characterize productive scientific inquiry (OECD, 2024; Zabelina & Robinson, 2010).

At the school level, instructional practices represent a critical contextual influence on student outcomes. Research consistently demonstrates that the quality of classroom instruction, particularly practices centered on adapting teaching to individual student needs and providing informative feedback, is meaningfully associated with academic achievement across subjects and education systems (Hattie, 2009; Black & Wiliam, 1998). Adaptive instruction enables teachers to respond to the diverse learning needs present within classrooms, while effective feedback supports students' self-regulation and helps correct conceptual misunderstandings before they become entrenched. Multilevel analyses linking teacher practices to student outcomes confirm that schools in which these instructional approaches are more prevalent tend to demonstrate higher aggregate performance even after accounting for student background characteristics (Le Donné et al., 2016). Emerging evidence also highlights the growing role of technology-enhanced instructional approaches including AI-assisted tools in supporting STEM learning outcomes across diverse educational contexts (Uğraş et al., 2024a; Uğraş et al., 2024b).

Despite growing evidence for both student dispositions and instructional practices as contributors to science achievement, comparatively few studies have examined their joint influence within a unified analytical framework, and fewer still have done so in the context of persistently low-performing education systems. Kosovo represents one such context. Since its initial participation in PISA in 2015 Kosovo has consistently ranked among the lowest-performing education systems internationally. PISA 2022 results placed Kosovo's mean science score at 353.57, approximately 146 points below the OECD mean of 500, a gap equivalent to more than three years of schooling (OECD, 2023). This persistent underperformance reflects a constellation of systemic challenges that distinguish Kosovo from both high-performing Western systems and its regional neighbors. Kosovo's education system has been characterized by an entrenched reliance on teacher-centered pedagogical approaches that offer students limited opportunities for inquiry-based learning, autonomous scientific exploration, or creative engagement with content (Grajcevci & Shala, 2021). Teacher expertise and professional development opportunities remain constrained and significant disparities in instructional quality exist across schools particularly between urban and rural settings (KEC, 2018). Kosovo also carries the legacy of a post-conflict educational reconstruction process that has prioritized access and enrollment over pedagogical quality resulting in a system where structural reforms have outpaced improvements in the quality of classroom instruction (Grajcevci & Shala, 2021). Additionally, Kosovo operates a split-shift schooling

system in which the majority of students attend school for approximately half a day substantially limiting instructional time relative to OECD comparison countries and contributing to between-school variation in achievement outcomes. Given these systemic challenges understanding which student and school-level factors are most strongly associated with science achievement in Kosovo carries both theoretical and practical significance.

From a theoretical standpoint, examining whether dispositional resources such as curiosity and creativity retain predictive value in a resource-constrained, low-performing system contributes to broader debates about the universality versus context-dependence of motivational effects on learning. From a practical standpoint, identifying modifiable predictors at both the student and school levels offers actionable guidance for educators, school leaders, and policymakers seeking evidence-based leverage points for improvement.

The present study addresses these questions by applying a two-level hierarchical linear modeling framework to Kosovo's PISA 2022 data, incorporating both student and teacher questionnaire responses. Student-level predictors include curiosity and creativity as dispositional variables, alongside socioeconomic status, gender, and immigration background as demographic controls. School-level predictors comprise aggregated measures of adaptive instruction and feedback practices derived from the PISA Teacher Questionnaire. The analysis further examines whether school-level instructional practices moderate the relationships between student dispositions and science achievement, providing a theoretically grounded test of person-environment interaction within Kosovo's educational context.

Research Questions

The following questions guided the study:

RQ1: To what extent do student dispositions: specifically, curiosity and creativity predict science achievement after controlling for socioeconomic status, gender, and immigration background?

RQ2: Do school-level instructional practices: adaptive instruction and feedback explain additional variance in science achievement beyond student-level factors?

Conceptual Framework

Science achievement does not emerge from individual aptitude alone, nor is it determined exclusively by institutional conditions. Rather, it reflects structured relationships between learners and the educational environments where learning takes place. This study is theoretically grounded in Bronfenbrenner's ecological systems theory, which conceptualizes human development as the product of dynamic and reciprocal interactions between individuals and the nested contextual structures that surround them (Bronfenbrenner, 1979). This perspective is particularly well suited to the analysis of PISA data, where students are embedded within schools that differ systematically in instructional organization,

professional capacity, and the quality of learning opportunities they provide. Ecological theory thus offers both conceptual justification for attending to multiple levels of influence and methodological rationale for modeling student and school characteristics concurrently.

Ecological Foundations for Multilevel Inquiry

Two implications follow for the study of science achievement. First, individual dispositions such as curiosity and creativity cannot be fully understood independently of the instructional conditions within which students exercise them. Second, the strength and direction of individual-level associations may vary across institutional contexts, making cross-level examination both theoretically motivated and empirically necessary. These propositions render multilevel modeling not merely a statistical convenience but a conceptually appropriate analytical strategy for examining nested educational data (Raudenbush & Bryk, 2002).

Student Dispositions as Individual-Level Resources

At Level 1, this study focuses on two motivational dispositions: curiosity and creativity as individual resources that shape how students engage with scientific content and tasks. From an ecological standpoint, these characteristics influence how students respond to instructional demands, allocate cognitive effort, approach novel problems, and sustain engagement during challenging academic work.

Curiosity supports exploratory thinking and drives students to seek explanations beyond what is immediately presented, fostering the kind of deep conceptual engagement that science learning requires. Students who are more curious are more likely to invest effort in understanding unfamiliar phenomena, generating questions, and pursuing answers through available resources and opportunities (Jirout & Klahr, 2012; von Stumm et al., 2011). Creativity contributes to flexible reasoning, hypothesis generation, and the integration of ideas across contexts, cognitive processes that are central to scientific inquiry and problem solving (Zabelina & Robinson, 2010; OECD, 2024). Together, these dispositions represent meaningful sources of heterogeneity among students exposed to broadly similar institutional conditions and thus account for variation in achievement within schools.

Instructional Practices as Contextual Structure

At Level 2, aggregated teacher instructional practices approximate the instructional culture of each school. Two dimensions of instructional practice are examined: adaptive instruction and feedback. Adaptive instruction refers to the extent to which teachers adjust their pedagogical approaches to accommodate the diverse learning needs, pacing, and prior knowledge levels of students within their classrooms. Feedback refers to the regularity and quality with which teachers communicate information about student understanding and progress in ways that support learning and self-correction.

From an opportunity-to-learn perspective, variation in these instructional practices reflects unequal access to cognitively supportive teaching environments (Floden, 2002). Schools where teachers more consistently adapt instruction and provide informative feedback create conditions under which students, particularly those with stronger motivational resources, can translate their dispositions into measurable learning gains. Even in systems where between-school variance in achievement is modest, systematic differences in instructional climate may contribute meaningfully to performance disparities across schools.

Cross-Level Processes and Theoretical Expectations

Ecological theory anticipates that individual dispositions and contextual features may interact in shaping developmental outcomes. In the present study, the central cross-level question concerns whether the associations between student curiosity and creativity and science achievement vary across schools that differ in the quality of their instructional practices.

Two competing theoretical expectations are plausible. An amplification perspective suggests that curious and creative students benefit disproportionately in schools characterized by stronger adaptive instruction and more informative feedback, as such environments provide the scaffolding and intellectual stimulation through which dispositional tendencies can be most productively expressed. A compensatory perspective, by contrast, posits that dispositions such as curiosity and creativity may be most consequential in weaker instructional environments, where students must rely more heavily on their own motivational resources to navigate limited external support.

Whether either pattern is detectable empirically depends on the degree of variability in instructional practices across schools and the statistical power available for detecting cross-level interaction effects. In contexts where instructional practices are relatively homogeneous, moderation effects may be attenuated regardless of their theoretical plausibility.

Analytical Boundaries

Several design features limit the extent to which ecological propositions can be fully evaluated in this study. The cross-sectional nature of PISA data precludes examination of the reciprocal and developmental processes that ecological theory emphasizes. School-level instructional measures are aggregated from teacher reports across subjects rather than derived from students' direct classroom experiences, which introduces some distance between the measured construct and the proximal learning environment. Additionally, the use of pre-computed composite scores for curiosity and creativity, while enhancing sample size and comparability, means that measurement error in these constructs cannot be explicitly modeled.

Grounded in ecological systems theory, this study examines how individual motivational resources and school-level instructional conditions jointly structure variation in science achievement among Kosovo

students. By modeling student and school influences simultaneously and evaluating their potential interaction, the analysis provides a theoretically coherent examination of nested influences within one of PISA's persistently lowest-performing educational contexts.

Literature Review

Understanding variation in students' science achievement requires situating learning within a broad body of work that examines how individual dispositions, classroom instruction, and school-level conditions shape performance on large-scale assessments. International studies such as PISA have consistently demonstrated that science achievement is influenced by both personal and contextual factors, though the relative strength of these influences varies across educational systems (OECD, 2023). This literature review synthesizes prior work on three interrelated domains central to the present study: science achievement in international assessment contexts, student-level dispositions associated with science learning, and teacher instructional practices as school-level predictors of achievement.

Science Achievement in International Assessment Contexts

Science achievement in PISA is conceptualized not merely as factual recall but as scientific literacy, the capacity to engage with science-related issues, interpret empirical data, evaluate competing claims, and apply scientific concepts in authentic and unfamiliar contexts (OECD, 2023). This emphasis on application and reasoning makes PISA particularly sensitive to both cognitive and dispositional influences on learning. International research consistently demonstrates that disparities in science achievement emerge from a combination of student background characteristics, motivational dispositions, and differences in instructional quality across schools and systems (OECD, 2023; OECD, 2024; Nilsen & Gustafsson, 2016).

Across OECD and partner countries, a substantial proportion of the variance in science performance lies within schools rather than between them, indicating that individual characteristics and classroom experiences shape outcomes as much as, and in many cases more than institutional resources alone. Nevertheless, school-level effects remain meaningful in systems where teacher preparation, instructional expectations, and assessment practices vary considerably (Le Donné et al., 2016). Research drawing on linked TALIS and PISA data consistently finds that teacher and instructional quality including clarity of explanation, provision of feedback, and the quality of classroom support carry meaningful associations with student achievement across academic domains, including science (Nilsen & Gustafsson, 2016).

Kosovo represents a particularly important and underrepresented context for examining the dispositional and instructional predictors of science achievement. Since its first participation in PISA in 2015 Kosovo has consistently ranked among the lowest-performing education systems internationally across successive assessment cycles. PISA 2022 results placed Kosovo's mean science score at 353.57, approximately 146 points below the OECD mean of 500, a gap equivalent to more than three

years of schooling (OECD, 2023). This persistent underperformance reflects a constellation of systemic challenges that distinguish Kosovo from both high-performing Western systems and its regional neighbors. Kosovo's education system has been characterized by an entrenched reliance on teacher-centered pedagogical approaches that offer students limited opportunities for inquiry-based learning, autonomous scientific exploration, or creative engagement with content (Grajcevci & Shala, 2021). Teacher expertise and professional development opportunities remain constrained, and significant disparities in instructional quality exist across schools, particularly between urban and rural settings (KEC, 2018). Kosovo also carries the legacy of a post-conflict educational reconstruction process that has prioritized access and enrollment over pedagogical quality, resulting in a system where structural reforms have outpaced improvements in the quality of classroom instruction (Grajcevci & Shala, 2021). Additionally, Kosovo operates a split-shift schooling system in which the majority of students attend school for approximately half a day, substantially limiting instructional time relative to OECD comparison countries and potentially contributing to between-school variation in achievement outcomes. As a post-conflict transitional society Kosovo is also characterized by collectivist social norms and strong cultural emphasis on educational attainment as a pathway to social mobility, contextual features that may shape how student motivational dispositions such as curiosity and creativity translate into academic performance in ways that differ meaningfully from Western educational contexts. The PISA 2022 cycle further expanded the range of student dispositions assessed including curiosity and creative thinking reflecting growing recognition that science learning is shaped not only by conceptual understanding but also by students' motivational and cognitive orientations toward learning and challenge (OECD, 2023; OECD, 2024). The present study builds on this body of work by investigating how curiosity and creativity function as predictors of science achievement within Kosovo's distinctive educational context and whether school-level instructional practices specifically adaptive instruction and feedback moderate or amplify the effects of these dispositions on science outcomes.

Student-Level Predictors of Science Achievement

Curiosity

Curiosity, broadly defined as students' intrinsic motivation to explore, ask questions, and seek explanations for phenomena they encounter, is conceptually aligned with the core epistemic practices of scientific inquiry (Jirout & Klahr, 2012). Students who are more curious tend to engage more actively with academic content, pursue understanding beyond what is formally required, and demonstrate greater persistence when confronted with intellectually demanding material. Research has increasingly positioned curiosity as an important motivational driver for engagement in science learning specifically, given that scientific reasoning fundamentally depends on the disposition to question, investigate, and revise understanding in light of new evidence (Sergeant-Perthuis et al., 2022).

Empirical studies provide consistent support for curiosity as a meaningful predictor of academic outcomes. In a comprehensive meta-analytic review, von Stumm et al. (2011) demonstrated that intellectual curiosity functions as a robust predictor of academic achievement, contributing unique

variance to performance even after accounting for cognitive ability and conscientiousness. Their analysis positioned curiosity as the third pillar of academic performance alongside intelligence and effort, underscoring its independent contribution to learning outcomes across domains. More recent work has elaborated the mechanisms through which curiosity influences learning, with evidence suggesting that curious students engage in more exploratory cognitive processing, make richer connections between ideas, and demonstrate greater retention of newly acquired knowledge (Sergeant-Perthuis et al., 2022). According to Onwunyili and Onwunyili (2020), this dispositional engagement is consistent with evidence showing that students who employ self-regulated learning strategies demonstrate stronger academic outcomes and lower test anxiety, suggesting that curiosity and self-regulation represent complementary motivational resources that jointly support academic performance.

The OECD's conceptual expansions in PISA 2022 further highlight curiosity as a contributor to students' capacity for creative and scientific reasoning, positioning it as a meaningful dispositional resource with direct relevance to science literacy outcomes (OECD, 2024). International assessment research has also shown that the academic benefits of curiosity tend to be stronger in instructional environments that provide students with opportunities to investigate phenomena, explore ideas, and engage in inquiry-oriented activities (OECD, 2023). Complementary evidence from Nigerian higher education contexts similarly demonstrates that self-regulatory strategies and motivational orientations, constructs closely related to curiosity, predict academic performance across diverse subject areas, suggesting that motivational resources operate as meaningful achievement determinants beyond Western educational contexts (Onwunyili & Ifenyinwa, 2020). In more traditional or didactic settings, the expression of curiosity may be constrained by limited opportunities for student-directed exploration, potentially attenuating its observable association with achievement. This interaction between curiosity and instructional context provides theoretical motivation for the cross-level examination pursued in the present study.

Creativity

Creativity has gained increasing recognition as an educationally significant construct, particularly in contexts that value flexible reasoning, novel problem formulation, and the integration of ideas across disciplinary boundaries (Zabelina & Robinson, 2010). In science learning, creativity supports hypothesis generation, the evaluation of alternative explanations, and the kind of divergent thinking that underlies productive scientific inquiry. The formalization of creative thinking as a measured domain in PISA 2022 reflects growing consensus that creativity is not peripheral to academic achievement but rather integral to the higher-order reasoning that science education seeks to develop (OECD, 2024).

Empirical research supports the relevance of creativity to science-related performance. International analyses from PISA 2022 reveal positive associations between creative thinking scores and performance across academic domains, including science, though effect sizes tend to be modest (OECD, 2024). Research in applied educational contexts similarly finds that students with stronger creative orientations demonstrate advantages in tasks requiring flexible reasoning, hypothesis evaluation, and the interpretation of novel or ambiguous information (Schoevers et al., 2022). Liu et al.

(2025) further demonstrated that creative self-efficacy significantly predicted problem-solving performance among engineering students, with this relationship mediated by critical thinking and metacognitive strategy use, suggesting that creativity supports not only idea generation but also the strategic cognitive processes underlying effective scientific reasoning.

The present study treats creativity as an observed predictor of science achievement, drawing on PISA 2022's pre-computed composite score for creative school and classroom engagement. While this measure captures students' perceptions of creative opportunities within their school environment rather than a pure cognitive trait, it remains theoretically relevant as an indicator of the extent to which students experience and engage with creativity-promoting conditions that may transfer to science performance.

Teacher-Level Predictors: Adaptive Instruction and Feedback Practices

While student dispositions shape readiness to learn, the instructional environments teachers create determine the extent to which students can translate that readiness into measurable achievement gains. Two dimensions of instructional practice are examined in the present study: adaptive instruction and feedback. Emerging evidence also underscores the growing relevance of technology-enhanced and AI-supported instructional tools as complements to traditional adaptive and feedback practices in STEM education contexts (Uğraş et al., 2024a; Uğraş et al., 2024b).

Adaptive instruction refers to the degree to which teachers adjust their pedagogical approaches, including the pacing, content, and complexity of instruction to accommodate variation in students' prior knowledge, learning needs, and rates of progress. High-quality adaptive instruction ensures that all students, regardless of their starting point, have access to appropriately challenging and supportive learning experiences. Research consistently links adaptive instructional practices to stronger student outcomes, particularly in diverse classrooms where student heterogeneity is pronounced (Hattie, 2009). When teachers are responsive to the varied needs present within their classrooms, students are better positioned to engage productively with academic content and to experience the kind of incremental success that sustains motivation and builds competence over time.

Feedback represents a second critical dimension of effective instruction. Formative feedback, information provided to students about the quality of their current understanding and the steps required to improve, has been identified as one of the most powerful influences on student learning across educational contexts (Black & Wiliam, 1998; Hattie, 2009). Evidence from classroom-based research further corroborates this view, demonstrating that the frequency and quality of teacher feedback in formative assessment contexts is positively associated with students' academic engagement and performance outcomes (Onwunyili & Onwunyili, 2019). Effective feedback makes student thinking visible to both teachers and learners, enables the identification and correction of misconceptions, and supports the development of self-regulatory skills that are particularly important for science learning. When feedback is specific, timely, and focused on the processes of learning rather than merely on performance

outcomes, it creates conditions under which students can monitor their own understanding and adjust their cognitive strategies accordingly.

International studies, including research drawing on linked TALIS and PISA data, confirm that teacher practices aggregated at the school level predict variation in student achievement beyond the contributions of individual student characteristics (Le Donné et al., 2016). Schools in which teachers collectively employ more adaptive, monitoring-oriented, and feedback-rich instructional approaches tend to demonstrate stronger achievement profiles. In Kosovo specifically, documented reliance on teacher-centered pedagogical traditions and limited adoption of formative assessment practices suggest that between-school variation in adaptive instruction and feedback may carry particular explanatory relevance for understanding achievement disparities (Grajcevci & Shala, 2021; KEC, 2018).

Method

Research Design

This study employed a cross-sectional multilevel modeling design to examine the direct effects of student-level dispositional characteristics and school-level instructional practices on science achievement among Kosovo students participating in PISA 2022. The multilevel analytical approach was selected to account for the nested structure of the data; students clustered within schools and to appropriately partition variance in science achievement between student and school levels. All models were estimated using the WeMix package (Bailey et al., 2023) in R version 4.4.1, which incorporates both student-level and school-level sampling weights simultaneously to produce population-representative parameter estimates that account for the complex two-level survey design of PISA. To properly handle the plausible value structure of the PISA science achievement outcome, all models were estimated using Huang's (2024) mixPV wrapper function extending WeMix to allow automatic pooling of results across multiple plausible values using Rubin's rules.

Participants and Sample

The target population consisted of 15-year-old students enrolled in Grade 7 or higher in Kosovo during the 2022 PISA assessment cycle. PISA employed a two-stage stratified probability sampling design in which schools were first selected with probability proportional to size, followed by random selection of approximately 35 students per sampled school. Kosovo's full PISA 2022 student sample comprised 6,027 students drawn from 229 schools. The teacher sample comprised 1,290 teachers drawn from 100 schools.

The analytic sample was constructed through a multi-step process. First, teacher data were aggregated to the school level by computing mean scores on adaptive instruction and feedback practices for each school, retaining only schools with three or more participating teachers to ensure reliable school-level estimates. This yielded 100 schools with valid school-level instructional data. Second, student data were merged with school-level teacher aggregates using the common school identifier (CNTSCHID). Third,

missing data on student dispositional variables, specifically curiosity and creativity arose from PISA's rotated booklet design in which these composites were administered to subsets of students by design, a planned missingness mechanism. To address this missingness multiple imputation was implemented using the mice package (Van Buuren & Groothuis-Oudshoorn, 2011) in R, generating ten imputed datasets using predictive mean matching applied to curiosity, creativity, and ESCS using all available student-level and school-level variables as predictors. The final analytic sample comprised N = 4,891 students nested within 100 schools, representing 81.2% of Kosovo's full PISA 2022 student cohort, with an average of 48.9 students per school.

Sampling Weights

To account for the complex two-level sampling design of PISA, sampling weights were incorporated at both levels of the model following the recommendations of Rutkowski et al. (2010) and Huang (2024). At Level 1, final student sampling weights (W_FSTUWT) were scaled within each school by dividing each student's weight by the mean weight of students in their school, producing within-school scaled weights with a mean of 1.0 as required by WeMix (Bailey et al., 2023). At Level 2, school sampling weights (W_SCHGRNRABWT) were extracted from the PISA 2022 teacher questionnaire file, the only file available in the current dataset containing school-level weight information and averaged within schools to produce a single school-level weight. The mean Level 2 weight across schools was 1.106. This two-level weighting approach ensures that parameter estimates reflect the population of Kosovo 15-year-old students and their schools rather than the unweighted analytic sample. All school-level variables required for the present study: adaptive instruction, feedback practices, and school-level sampling weights were available in the teacher questionnaire file, making the school questionnaire file unnecessary for the specific aims of this study.

Measures

Science Achievement

Science literacy was assessed through PISA's cognitive test battery measuring students' capacity to explain phenomena scientifically, evaluate and design scientific investigations, and interpret empirical data. Achievement scores are reported as ten plausible values generated through item response theory and latent regression modeling. For descriptive purposes a composite science score was computed as the unweighted mean of the ten plausible values (M = 353.57, SD = 64.85). Following Huang (2024) all models were estimated using the mixPV wrapper function which runs the model separately for each of the ten plausible values and pools results using Rubin's rules, appropriately accounting for the additional measurement uncertainty introduced by plausible value methodology.

Student-Level Dispositional Variables

Three student dispositional constructs were derived from PISA's pre-computed weighted likelihood estimate composite indices. All composites were grand-mean centered prior to analysis to facilitate

interpretation of school-level coefficients as effects at the average student profile (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002).

Curiosity was measured using the PISA 2022 Curiosity composite (CURIOAGR), a WLE scale derived from seven questionnaire items assessing students' intrinsic motivation to explore ideas and seek explanations (M = 0.27, SD = 1.15).

Creativity was measured using the PISA 2022 Creative School and Classroom Environment composite (CREATSCH), a WLE scale derived from five questionnaire items assessing students' perceptions of creative opportunities within their school environment (M = 0.22, SD = 1.16). Table 1 presents the student disposition scales with reliability estimates and descriptive statistics.

Table 1
Student Disposition Scales: Reliability and Descriptive Statistics

Construct	Items (N)	Example Item	α	M	SD	Range
Curiosity	7	"I like to find out how things work"	.84	0.27	1.15	-3.64–4.08
Creativity	5	"I come up with original solutions"	.86	0.22	1.16	-2.62–2.81

Note. Scale scores are PISA pre-computed weighted likelihood estimates (WLE) aggregated across available questionnaire booklets. Reliability estimates (Cronbach's α) were calculated using item-level responses from the analytic sample. M and SD are pooled across 10 multiply imputed datasets (N = 4,891).

Demographic Covariates

Socioeconomic status was operationalized using PISA's Economic, Social, and Cultural Status index (ESCS), a composite measure of parental education, occupational status, and home educational resources standardized across OECD countries (M = -0.34, SD = 0.89). Gender was coded as a binary variable with 0 = female and 1 = male, with 51% of the analytic sample identified as male. Immigrant status was coded as a binary variable with 0 = native student and 1 = first- or second-generation immigrant, with 3% of the analytic sample identified as immigrant students as presented in Table 2.

Table 2
Demographic Covariates: Descriptive Statistics (N = 4,891)

Variable	Operationalization	M	SD	%
ESCS	Continuous composite of parental education, occupational status, and home resources; standardized across OECD countries	-0.34	0.89	—
Gender	Binary: 0 = female, 1 = male	—	0.50	51% male
Immigrant Status	Binary: 0 = native, 1 = first/second generation immigrant	—	0.18	3% immigrant

Note. ESCS = Economic, Social, and Cultural Status index. Gender coded as dummy variable with female as reference category. Immigrant status coded as dummy variable with native students as reference category. Descriptive statistics pooled across 10 multiply imputed datasets.

School-Level Instructional Variables

Two school-level instructional practice constructs were derived by aggregating teacher-reported PISA 2022 WLE composite scores to the school level. For each school the mean of all participating teachers' scores was computed with a minimum of three teachers required per school to ensure reliability of school-level estimates. Both school-level variables were grand-mean centered prior to analysis (Enders & Tofighi, 2007).

Adaptive Instruction was measured using the PISA 2022 Adaptation of Instruction composite (ADAPTINSTR), a WLE scale derived from teacher questionnaire items assessing the extent to which teachers adapt their instructional approaches to meet diverse student learning needs (M = 0.27, SD = 0.38 at school level).

Feedback Practices was measured using the PISA 2022 Feedback composite (FEEDBINSTR), a WLE scale derived from teacher questionnaire items assessing the frequency and quality of feedback provided to students (M = 0.26, SD = 0.36 at school level). Table 3 presents the teacher instructional practice scales with school-level descriptive statistics.

Table 3

Teacher Instructional Practice Scales: Description and School-Level Descriptives

Construct	PISA Scale	Description	M	SD	Range
Adaptive Instruction	ADAPTINSTR	Frequency with which teachers adapt instruction to diverse student learning needs	0.27	0.38	-0.75–1.28
Feedback Practices	FEEDBINSTR	Frequency with which teachers provide informative feedback on student learning progress	0.26	0.36	-0.57–1.32

Note. School-level values represent means of teacher-level WLE composite scores aggregated across all participating teachers within each school. Only schools with a minimum of three teacher respondents were included. M and SD reported at school level (100 schools).

Data Preparation

Missing Data

Missing data on student dispositional variables arose primarily from PISA's rotated booklet design in which curiosity and creativity composites were administered to subsets of students by design, a planned missingness mechanism rather than random attrition. To address this missingness, multiple imputation

was implemented using the mice package (Van Buuren & Groothuis-Oudshoorn, 2011) in R, generating ten imputed datasets using predictive mean matching. Curiosity, creativity, and ESCS were imputed using all available student-level and school-level variables as predictors. After merging the imputed student data with school-level teacher aggregates the final analytic sample comprised N = 4,891 students nested within 100 schools representing 81.2% of Kosovo's full PISA 2022 student cohort. Table 4 presents the missingness rates for all analytic variables prior to imputation.

Table 4

Missing Data Rates for Analytic Variables Prior to Imputation

Variable	% Missing	N Available
Science Plausible Values	0.0%	4,891
Curiosity	20.6%	3,884
Creativity	27.4%	3,550
ESCS	3.4%	4,726
Gender	0.0%	4,891
Immigrant Status	0.0%	4,891
Adaptive Instruction (school-level)	0.0%	100 schools
Feedback Practices (school-level)	0.0%	100 schools

Note. Missingness rates reflect the merged analytic sample of 4,891 students prior to multiple imputation. Missingness on curiosity and creativity arose from PISA's rotated questionnaire booklet design a planned missingness mechanism. All variables were fully observed after multiple imputation using predictive mean matching via the mice package.

Variable Centering

All continuous student-level predictors were grand-mean centered prior to model estimation. Grand-mean centering renders the model intercept interpretable as the expected science score for a student with average values on all predictors enrolled in an average school and reduces nonessential multicollinearity between main effects and interaction terms in the cross-level interaction model (Enders & Tofighi, 2007; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). School-level instructional predictors were similarly grand-mean centered. Categorical predictors: gender and immigrant status retained their original coding with explicitly defined reference categories.

Outlier Screening

Univariate outliers were identified using a ± 3 SD criterion and multivariate outliers were assessed using Mahalanobis distance. Observations exceeding conventional thresholds were retained because they represented plausible values within a nationally representative probability sample. Sensitivity analyses excluding potential outliers yielded substantively equivalent results.

Analytic Approach

Multilevel Modeling Framework

Multilevel modeling also referred to as hierarchical linear modeling is the appropriate analytical framework when data have a nested structure in which lower-level units such as students are clustered within higher-level units such as schools (Raudenbush & Bryk, 2002). Ignoring this nesting violates the assumption of independence of observations that underlies ordinary regression and produces artificially deflated standard errors leading to inflated Type I error rates. The ICC from Model 0 confirmed that a meaningful proportion of variance in science achievement was attributable to between-school differences, justifying the multilevel approach. All models specified random intercepts for schools, allowing each school to have its own baseline science achievement score while fixing slopes across schools given insufficient between-school variability in student-level slopes to justify random slope specification.

Handling of Plausible Values

PISA science achievement is represented by ten plausible values rather than a single point estimate, each representing a draw from the posterior distribution of student proficiency conditional on item responses and background characteristics. As demonstrated by Huang (2024), taking the average of plausible values or using only one value results in underestimated standard errors that do not account for the variability resulting from the differing values across analyses. Following Huang (2024) all models were therefore estimated using the mixPV wrapper function for WeMix, which runs the model separately for each of the ten plausible values and automatically pools the resulting estimates and standard errors using Rubin's rules producing a single set of results that appropriately accounts for the additional measurement uncertainty introduced by plausible value methodology. Within each imputed dataset, ten separate models were estimated one per plausible value and results were pooled using Rubin's rules. A second application of Rubin's rules then pooled estimates across the ten imputed datasets, producing final standard errors that simultaneously reflect both plausible value measurement uncertainty and imputation uncertainty. In total 100 models were estimated per model specification. Barnard-Rubin adjusted degrees of freedom were computed for each pooled parameter estimate to account for the uncertainty introduced by imputation (Barnard & Rubin, 1999).

Model Fit Evaluation

Model fit was evaluated using mean AICbar values, the average Akaike Information Criterion across plausible values as reported by the mixPV glance function (Huang, 2024). Lower AICbar values indicate better model fit with improvements assessed by comparing mean AICbar across successive models. Huang (2024) cautions against the use of likelihood ratio tests with plausible values due to the complexity of appropriate pooling methods for nested model comparisons noting that while the D2 statistic exists for this purpose its computation is not straightforward. AIC-based comparison was therefore adopted as the primary model selection criterion. In addition to AICbar the following criteria

were examined: the magnitude and statistical significance of fixed effect coefficients, the proportion of variance explained at each level relative to the unconditional model, and the reduction in ICC across models reflecting the degree to which predictors accounted for between-school variance.

Model Specification

A sequential hierarchical modeling approach was employed using the WeMix package (Bailey et al., 2023) and Huang's (2024) mixPV wrapper function in R version 4.4.1. Four models were estimated sequentially to address the research questions. The general two-level model can be expressed as:

Level 1 — Student equation:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Curiosity})_{ij} + \beta_{2j}(\text{Creativity})_{ij} + \beta_{3j}(\text{ESCS})_{ij} + \beta_{4j}(\text{Gender})_{ij} + \beta_{5j}(\text{Immigrant})_{ij} + r_{ij}$$

Level 2 — School equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{ADAPTINSTR})_j + \gamma_{02}(\text{FEEDBINSTR})_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{ADAPTINSTR})_j$$

$$\beta_{2j} = \gamma_{20} + \gamma_{22}(\text{FEEDBINSTR})_j$$

Combined equation:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(\text{Curiosity})_{ij} + \gamma_{20}(\text{Creativity})_{ij} + \gamma_{30}(\text{ESCS})_{ij} + \gamma_{40}(\text{Gender})_{ij} + \gamma_{50}(\text{Immigrant})_{ij} + \gamma_{01}(\text{ADAPTINSTR})_j + \gamma_{02}(\text{FEEDBINSTR})_j + \gamma_{11}(\text{Curiosity} \times \text{ADAPTINSTR})_{ij} + \gamma_{22}(\text{Creativity} \times \text{FEEDBINSTR})_{ij} + u_{0j} + r_{ij}$$

where Y_{ij} is science achievement for student i in school j, γ_{00} is the grand mean intercept, γ coefficients represent fixed effects, u_{0j} is the school-level random intercept representing school deviation from the grand mean, and r_{ij} is the student-level residual.

Model Building Sequence

Four models were estimated sequentially across all ten plausible values and ten imputed datasets using the mixPV function:

Model 0 — Unconditional model: Science achievement modeled as a function of school random intercepts only with no predictors at either level. Used to estimate the intraclass correlation coefficient (ICC) quantifying the proportion of total variance attributable to between-school differences and to establish baseline variance components for subsequent comparisons.

Model 1 — Student-level predictors: Added all five student-level predictors: curiosity, creativity, ESCS, gender, and immigrant status to examine their direct effects on science achievement while accounting for school clustering.

Model 2 — School-level predictors: Added the two school-level instructional practice predictors: adaptive instruction and feedback practices to examine their contributions to between-school variance after accounting for student-level characteristics.

Model 3 — Cross-level interactions: Added two cross-level interaction terms: Curiosity × Adaptive Instruction and Creativity × Feedback Practices to test whether school-level instructional practices moderated the relationships between student dispositions and science achievement.

Model Selection

Model selection was guided by mean AICbar values: the average Akaike Information Criterion across the ten plausible values as computed by the mixPV glance function (Huang, 2024), averaged further across the ten imputed datasets. Lower mean AICbar values indicate better model fit and improvements in fit were assessed by comparing mean AICbar across successive models. Following Huang (2024) likelihood ratio tests were not used for model comparison given the complexity of appropriately pooling test statistics across both plausible values and imputed datasets. The final model was selected as the most parsimonious model demonstrating meaningful improvement in mean AICbar over simpler alternatives. All continuous predictors were grand-mean centered prior to analysis to facilitate interpretation and reduce multicollinearity between main effects and interaction terms (Enders & Tofighi, 2007; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002).

Intraclass Correlation and Variance Explained

The ICC from Model 0 was used to confirm the appropriateness of multilevel modeling. A meaningful ICC indicates that students within the same school share more similar science achievement scores than students across schools, justifying the nested analytical approach. ICC values were computed from variance components pooled across all ten plausible values and ten imputed datasets using Rubin's rules. Proportions of variance explained at each level were computed as the proportional reduction in pooled variance components relative to the unconditional model, Level 1 R^2 reflecting the reduction in within-school residual variance and Level 2 R^2 reflecting the reduction in between-school variance.

Results

Descriptive Statistics

Descriptive statistics for all analytic variables are presented in Table 5. Science achievement averaged 355.24 (SD = 59.73) computed as the unweighted mean of ten plausible values for descriptive purposes only, consistent with Kosovo's persistent underperformance relative to the OECD mean of 500. At the student level curiosity averaged 0.27 (SD = 1.15) and creativity averaged 0.22 (SD = 1.16), both expressed

as WLE composites centered approximately around zero and pooled across ten imputed datasets. Socioeconomic status averaged − 0.34 standard deviations below the OECD mean (SD = 0.89), reflecting Kosovo's socioeconomic disadvantage relative to OECD countries. Approximately 51% of the analytic sample was male and 3% were identified as immigrant students. At the school level adaptive instruction averaged 0.27 (SD = 0.38) and feedback practices averaged 0.26 (SD = 0.36), reflecting moderate levels of instructional quality across Kosovo schools.

Table 5
Descriptive Statistics for All Analytic Variables

Variable	N	M	SD	Min	Max
Outcome					
Science Achievement	4891	355.24	59.73	—	—
Student Level					
Curiosity	4891	0.27	1.15	-3.64	4.08
Creativity	4891	0.22	1.16	-2.62	2.81
ESCS	4891	-0.34	0.89	-3.48	4.26
Gender (% male)	4891	51%	0.50	0.00	1.00
Immigrant (% immigrant)	4891	3%	0.18	0.00	1.00
School Level (100 schools)					
Adaptive Instruction	100	0.27	0.38	—	—
Feedback Practices	100	0.26	0.36	—	—

*Science achievement M and SD reported as the unweighted mean of ten plausible values for descriptive purposes only. All inferential models were estimated across all ten plausible values using mixPV (*Huang, 2024*)*

Correlations and Intercorrelations

Bivariate correlations among all student-level analytic variables computed on the first imputed dataset are presented in Table 6. Science achievement was positively and significantly correlated with curiosity (r = .19, p < .001), creativity (r = .12, p < .001), and ESCS (r = .26, p < .001). Gender showed a significant negative correlation with science achievement at the bivariate level (r = − .06, p < .001) a pattern that reverses in the multilevel model once school-level clustering and other student characteristics are properly accounted for, as discussed in the Results section. Immigrant status showed a small negative correlation with science achievement (r = − .04, p = .003). Among the dispositional variables curiosity and creativity were moderately positively correlated (r = .25, p < .001) indicating that more curious

students also tended to report stronger creative orientations. ESCS showed a small but significant positive correlation with curiosity (r = .04, p = .007) and creativity (r = .04, p = .007). At the school level adaptive instruction and feedback practices were moderately positively correlated (r = .65, p < .001) indicating that schools with higher adaptive instruction also tended to report stronger feedback practices, a pattern that introduces potential multicollinearity at Level 2 discussed in the Limitations section.

Table 6
Intercorrelations Among All Student-Level Analytic Variables (N = 4,891, Imputed Dataset 1)

Variable	1	2	3	4	5	6
1. Science Achievement	—					
2. Curiosity	.19***	—				
3. Creativity	.12***	.25***	—			
4. ESCS	.26***	.04*	.04*	—		
5. Gender (male)	− .06***	− .14***	− .08***	.01	—	
6. Immigrant	− .04**	− .04**	.00	.09***	.06***	—

*Note. Correlations computed on first imputed dataset. *p < .05, **p < .01, ***p < .001 (two-tailed). ESCS = Economic, Social, and Cultural Status index. Gender coded as 0 = female, 1 = male. Immigrant coded as 0 = native, 1 = first/second generation immigrant.*

Intraclass Correlation and Justification for Multilevel Modeling

The unconditional model (Model 0) yielded an intraclass correlation coefficient of ICC = .324, indicating that 32.4% of the total variance in science achievement was attributable to between-school differences after applying proper two-level sampling weights and correctly handling all ten plausible values using the mixPV function (Huang, 2024; Bailey et al., 2023). ICC values were pooled across all ten plausible values and ten imputed datasets using Rubin's rules. The magnitude of the ICC provides strong justification for the multilevel analytical approach, students within the same school shared considerably more similar science achievement scores than students across different schools, confirming that ignoring school-level clustering would produce substantially biased standard errors and misleading inferences about predictor effects. The pooled between-school variance from Model 0 was τ_{00} = 1,378.0 and the within-school residual variance was σ^2 = 2,879.6 yielding a total variance of 4,257.6 points squared.

Model Comparison and Selection

Table 7 presents the model comparison results using mean AICbar values, averaged across ten plausible values and ten imputed datasets. Model 1 adding student-level predictors produced a meaningful improvement in fit over the unconditional model (ΔAICbar = -128.87), confirming that curiosity, creativity, ESCS, gender, and immigrant status collectively explained meaningful variance in science achievement beyond school-level clustering. Model 2 adding school-level instructional predictors produced a further improvement over Model 1 (ΔAICbar = -9.01), indicating that adaptive instruction and feedback practices contributed meaningfully to explaining between-school variance after accounting for student-level characteristics. Model 3 adding cross-level interaction terms produced negligible improvement over Model 2 (ΔAICbar = -0.40), indicating that school-level instructional practices did not significantly moderate the relationships between student dispositions and science achievement. Model 2 was therefore retained as the final model.

Table 7
Model Comparison Using Mean AICbar

Model	Mean AICbar	ΔAICbar	Decision
Model 0 — Unconditional	53184.42	—	Baseline
Model 1 — Student level	53055.55	-128.87	Retain
Model 2 — School level	53046.54	-9.01	Retain
Model 3 — Interactions	53046.14	-0.40	Drop

*Note. Mean AICbar = average Akaike Information Criterion across ten plausible values computed by mixPV (*Huang, 2024*) and averaged across ten imputed datasets. Lower values indicate better fit. Model 2 retained as the most parsimonious model demonstrating meaningful improvement over simpler alternatives.*

Multilevel Model Results

Model 1: Student-Level Predictors

Table 8 presents pooled fixed effect estimates from Model 1 across ten plausible values and ten imputed datasets. All student-level predictors except immigrant status demonstrated statistically significant associations with science achievement. Among student-level dispositional predictors curiosity demonstrated the strongest association with science achievement (b = 4.716, SE = 1.093, t = 4.313, df = 564.4, p < .001, β = .084), followed by creativity (b = 2.585, SE = 0.911, t = 2.838, df = 1042.5, p = .005, β = .046). Among demographic covariates gender demonstrated a significant positive association (b = 8.113, SE = 1.735, t = 4.676, df = 4706.9, p < .001, β = .058) with male students scoring higher than female students after accounting for school clustering, notably this positive association emerged in the multilevel model despite a negative bivariate correlation (r = − .06) reflecting the importance of properly accounting for school-level clustering. ESCS was also significant (b = 5.705, SE = 0.921, t = 6.194, df = 3200.7, p < .001, β = .078). Immigrant status was not significant (b = -10.781, SE = 4.622, t = -2.333, df =

4469.1, p = .197, β = − .030). The ICC decreased from .324 in Model 0 to .306 in Model 1 indicating that student-level predictors explained a modest proportion of between-school variance.

Table 8
Fixed Effects for Model 1: Student-Level Predictors (N = 4,891 students, 100 schools)

Predictor	b	SE	t	df	p	β
Intercept	352.168	3.930	89.615	4880.7	< .001	—
Curiosity	4.716	1.093	4.313	564.4	< .001	.084
Creativity	2.585	0.911	2.838	1042.5	.005	.046
ESCS	5.705	0.921	6.194	3200.7	< .001	.078
Gender (male)	8.113	1.735	4.676	4706.9	< .001	.058
Immigrant	-10.781	4.622	-2.333	4469.1	.197	− .030

Note. Pooled estimates across ten plausible values and ten imputed datasets using two-stage Rubin's rules. b = unstandardized pooled coefficient; SE = pooled standard error incorporating within and between imputation variance; df = Barnard-Rubin adjusted degrees of freedom; β = standardized coefficient. Gender coded as 0 = female, 1 = male. Immigrant coded as 0 = native, 1 = first/second generation immigrant. ESCS = Economic, Social, and Cultural Status index.

Model 2: Adding School-Level Instructional Predictors

Table 9 presents pooled fixed effect estimates from the final Model 2. Student-level findings remained substantially unchanged from Model 1 after adding school-level predictors confirming the robustness of student-level effects. Among dispositional predictors curiosity remained the strongest student-level predictor (β = .084) followed by creativity (β = .046). At the school level adaptive instruction demonstrated a significant positive association with science achievement (b = 32.665, SE = 11.594, t = 2.817, df = 4879.1, p = .005, β = .192) emerging as the strongest predictor in the entire model with a standardized coefficient exceeding all student-level predictors. A one standard deviation increase in school-level adaptive instruction was associated with approximately 32.67 points higher science achievement, a practically meaningful difference. Feedback practices was not significant (b = -3.951, SE = 12.678, t = -0.312, df = 4879.2, p = .755, β = − .022). The robust non-significance of feedback practices was confirmed across all ten plausible values and ten imputed datasets suggesting this reflects a genuine absence of association rather than insufficient statistical power. The ICC decreased further to .285 in Model 2 reflecting the additional between-school variance explained by the school-level instructional predictors.

Table 9

Fixed Effects for Model 2: Student and School-Level Predictors — Final Model (N = 4,891 students, 100 schools)

Predictor	b	SE	t	df	p	β
Intercept	352.673	3.679	95.848	4871.2	< .001	—
Student Level						
Curiosity	4.716	1.093	4.313	564.4	< .001	.084
Creativity	2.585	0.911	2.838	1042.5	.005	.046
ESCS	5.626	1.216	4.626	4262.0	< .001	.078
Gender (male)	7.567	2.340	3.233	4439.7	.001	.058
Immigrant	-11.023	6.316	-1.745	4806.4	.081	− .030
School Level						
Adaptive Instruction	32.665	11.594	2.817	4879.1	.005	.192
Feedback Practices	-3.951	12.678	-0.312	4879.2	.755	− .022

Note. Pooled estimates across ten plausible values and ten imputed datasets using two-stage Rubin's rules. b = unstandardized pooled coefficient; SE = pooled standard error; df = Barnard-Rubin adjusted degrees of freedom; β = standardized coefficient. Gender coded as 0 = female, 1 = male. Immigrant coded as 0 = native, 1 = first/second generation immigrant. ESCS = Economic, Social, and Cultural Status index.

Table 10

Cross-Level Interaction Effects — Model 3 (N = 4,891 students, 100 schools)

Predictor	b	SE	t	df	p	β
Intercept	352.441	3.533	99.770	4875.7	< .001	—
Student Level						
Curiosity	5.827	0.741	7.861	423.9	< .001	.104
Creativity	3.458	0.937	3.692	29.0	< .001	.062
ESCS	5.654	0.911	6.204	3076.5	< .001	.078
Gender (male)	8.164	1.733	4.711	4684.8	< .001	.063
Immigrant	-10.746	4.612	-2.330	4439.2	.198	− .029
School Level						
Adaptive Instruction	33.064	11.512	2.872	4876.3	.004	.194
Feedback Practices	-4.010	12.316	-0.326	4874.5	.745	− .022
Cross-Level Interactions						
Curiosity × Adaptive Instruction	1.127	1.784	0.632	833.1	.528	.007
Creativity × Feedback Practices	0.333	1.737	0.192	731.1	.848	.002

Note. Model 3 not retained as final model due to negligible improvement over Model 2 (ΔAICbar = -0.40). Pooled estimates across ten plausible values and ten imputed datasets using two-stage Rubin's rules. b = unstandardized pooled coefficient; SE = pooled standard error; df = Barnard-Rubin adjusted degrees of freedom; β = standardized coefficient.

Variance Decomposition Summary

The proportion of variance explained at each level was computed as the proportional reduction in pooled variance components relative to the unconditional model across all ten plausible values and ten imputed datasets. At Level 1 student-level predictors explained 2.7% of within-school variance in science achievement (R^2_{L1} = .027). At Level 2 school-level instructional predictors explained 19.1% of between-school variance (R^2_{L2} = .191). The total proportion of variance explained across both levels was 8.0% (R^2_{total} = .080). The substantially higher Level 2 R^2 relative to Level 1 R^2 reflects the dominant role of school-level instructional conditions in accounting for between-school differences in science achievement, a finding with important implications for educational policy in Kosovo.

Table 11

Variance Partitioning Across Models (Pooled Across Ten Plausible Values and Ten Imputed Datasets)

Model	τ_{00}	SE	σ^2	SE	ICC	R^2L1	R^2L2	R^2total
Model 0 — Unconditional	1378.0	203.97	2879.6	85.62	.324	—	—	—
Model 1 — Student level	1233.8	—	2804.1	—	.306	.027	—	—
Model 2 — School level (Final)	1115.0	162.0	2802.8	108.6	.285	.027	.191	.080
Model 3 — Interactions	1090.1	—	2191.4	—	.285	.027	.191	.080

Note. τ_{00} *= between-school variance pooled across ten plausible values and ten imputed datasets; SE = pooled standard error of variance estimate where available;* σ^2 *= within-school residual variance; ICC = intraclass correlation coefficient;* R^2L1 *= proportion of within-school variance explained relative to Model 0;* R^2L2 *= proportion of between-school variance explained relative to Model 0;* R^2total *= proportion of total variance explained relative to Model 0. Model 2 is the final retained model.*

Summary of Key Findings

RQ1: Curiosity (β = .084, p < .001) and Creativity (β = .046, p = .005) both demonstrated significant positive direct effects on science achievement after controlling for socioeconomic status, gender, immigrant status, and school-level instructional practices supporting H1 and H2 respectively.

RQ2: Adaptive Instruction demonstrated a significant positive effect on science achievement at the school level (β = .192, p = .005) supporting H3. Feedback Practices was not significant (β = − .022, p = .755) not supporting H4.

RQ3: The cross-level interaction between Curiosity and Adaptive Instruction was not significant (b = 1.127, p = .528) not supporting H5.

RQ4: The cross-level interaction between Creativity and Feedback Practices was not significant (b = 0.333, p = .848) not supporting H6.

RQ5: The final model explained 2.7% of within-school variance (R^2L1 = .027), 19.1% of between-school variance (R^2L2 = .191), and 8.0% of total variance in science achievement (R^2total = .080). The ICC of .324 confirmed substantial between-school clustering justifying the multilevel analytical approach.

Discussion

This study examined the direct effects of student-level dispositional characteristics specifically curiosity and creativity, and school-level instructional practices specifically adaptive instruction and feedback practices on science achievement among Kosovo students participating in PISA 2022, using a weighted two-level multilevel modeling approach that incorporated both student-level and school-level sampling

weights through the WeMix package (Bailey et al., 2023) and properly handled all ten plausible values using the mixPV wrapper function (Huang, 2024). The findings provided support for several hypothesized relationships while also revealing patterns that warrant careful theoretical interpretation. The discussion addresses each set of findings in turn, situating them within the broader theoretical and empirical literature before drawing implications for educational policy and practice in Kosovo.

Intraclass Correlation and Between-School Variance

The unconditional model yielded an ICC of .324, indicating that 32.4% of the total variance in science achievement was attributable to between-school differences after proper two-level weighting and correct plausible value handling using mixPV (Huang, 2024). This ICC reflects the true degree of school-level clustering in Kosovo's nationally representative student population and is substantially higher than estimates typically reported in unweighted analyses that do not properly account for plausible value measurement uncertainty. The magnitude of the ICC confirms that schools in Kosovo vary considerably in their science achievement outcomes, a pattern consistent with documented inequalities in school quality, resource availability, and instructional capacity across Kosovo's education system (KEC, 2018; Grajcevci & Shala, 2021). Beyond confirming these structural inequalities this finding underscores the importance of examining both student-level and school-level predictors simultaneously and confirms that multilevel modeling was not merely a methodological preference but a substantive necessity for accurately characterizing the determinants of science achievement in Kosovo.

Furthermore, Kosovo's split-shift schooling system in which the majority of students attend school for approximately half a day represents an important structural feature not captured in PISA data that may contribute to the substantial between-school variance observed here and warrants consideration in future research.

Curiosity as a Predictor of Science Achievement

Consistent with H1 curiosity demonstrated a significant positive association with science achievement ($\beta = .084$, $p < .001$), confirming that more curious students achieved higher science scores after controlling for creativity, socioeconomic status, gender, immigrant status, and school-level instructional practices. This finding is consistent with prior evidence positioning intellectual curiosity as a robust and consequential predictor of academic performance across domains and educational contexts (von Stumm et al., 2011). That curiosity retained a significant positive association with science achievement even within Kosovo's predominantly teacher-centered instructional environment where opportunities for student-directed exploration are constrained (Grajcevci & Shala, 2021) suggests that curiosity functions as a relatively robust dispositional resource whose academic benefits are not entirely contingent on the availability of inquiry-supportive instruction. This extends prior evidence suggesting that curiosity yields stronger benefits in inquiry-supportive environments (OECD, 2023) by demonstrating that its positive association with achievement persists even where such environments are limited. This finding is also

consistent with evidence from Nigerian educational contexts demonstrating that motivational orientations including self-regulatory strategies function as meaningful predictors of academic performance across diverse learning settings (Onwunyili & Ifenyinwa, 2020).

Creativity as a Predictor of Science Achievement

Consistent with H2 creativity demonstrated a significant positive association with science achievement ($\beta = .046$, $p = .005$), confirming that students with stronger creative orientations achieved higher science scores after controlling for all other predictors. Although the standardized effect of creativity was considerably smaller than that of curiosity the finding is theoretically meaningful, it establishes that students' perceptions of creative opportunities within their school environment are independently associated with science performance beyond the contribution of curiosity and socioeconomic factors. This is consistent with research linking creative thinking to flexible reasoning, divergent problem solving, and the evaluative thinking that science learning demands (Zabelina & Robinson, 2010; OECD, 2024). The relatively modest effect size may reflect the specific operationalization of creativity through the CREATSCH composite which captures students' perceptions of the creative climate of their school environment rather than individual creative ability or creative self-efficacy. Schools that foster a more creative learning environment appear to support science achievement modestly but meaningfully even after accounting for the quality of instructional practices at the school level.

Socioeconomic Status and Demographic Covariates

Socioeconomic status demonstrated a significant positive association with science achievement ($\beta = .078$, $p < .001$), confirming the persistent influence of socioeconomic background on science performance in Kosovo. This finding is consistent with broader PISA literature documenting strong socioeconomic gradients in academic achievement across participating countries and is particularly pronounced in lower-income systems where material resource disparities translate directly into unequal access to quality learning experiences (OECD, 2023). Gender was a significant positive predictor of science achievement in the multilevel model ($\beta = .058$, $p = .001$) with male students scoring approximately 7.57 points higher than female students on average after controlling for all other predictors. Notably gender showed a negative correlation with science achievement at the bivariate level ($r = .06$) but emerged as a positive significant predictor once school-level clustering and other student characteristics were properly accounted for in the multilevel framework (Raudenbush & Bryk, 2002), illustrating the importance of appropriate analytical methods for revealing relationships that may be obscured in simpler analyses. Immigrant status was not a significant predictor of science achievement ($\beta = .030$, $p = .081$), indicating that immigrant and native students did not differ significantly in science performance after accounting for other student and school characteristics. This finding is consistent with the relatively small proportion of immigrant students in Kosovo's student population (3%) and the consequent limited statistical power to detect immigrant-native achievement gaps. Importantly this non-

significance was robust across all ten plausible values and ten imputed datasets confirming it reflects a genuine pattern rather than a methodological artifact.

Adaptive Instruction as a School-Level Predictor

Consistent with H4 adaptive instruction demonstrated a significant positive association with science achievement at the school level (β = .192, p = .005), emerging as the strongest predictor in the entire model, stronger than any individual student-level predictor including curiosity. Schools where teachers more consistently adapted their instructional approaches to meet diverse student learning needs showed substantially higher science achievement after accounting for student-level characteristics. A one standard deviation increase in school-level adaptive instruction was associated with approximately 32.67 points higher science achievement, a practically meaningful difference. This finding is consistent with research demonstrating that adaptive or differentiated instruction which involves adjusting the pace, content, and pedagogical approach to accommodate diverse student needs is among the most powerful school-level influences on student learning outcomes (OECD, 2023). Within Bronfenbrenner's ecological framework the school microsystem, particularly the quality of instructional practices within classrooms, represents a proximal environmental influence on student learning that operates above and beyond the dispositional characteristics students bring to the learning environment. The finding that adaptive instruction accounted for 19.1% of between-school variance in science achievement underscores the critical role of instructional quality in explaining why some Kosovo schools produce substantially better science outcomes than others, even when serving students with similar curiosity, creativity, and socioeconomic backgrounds.

This finding carries particular significance in the Kosovo context where documented challenges in teacher preparation, professional development, and instructional quality have been identified as central contributors to persistent underperformance in international assessments (KEC, 2018; Grajcevci & Shala, 2021). Evidence from classroom assessment research similarly highlights the critical role of responsive instructional practices in shaping students' academic engagement and performance outcomes (Onwunyili & Onwunyili, 2019; Chinedu & Chinwendu, 2020), reinforcing the view that targeted investment in teachers' adaptive instructional capacity represents one of the most promising levers for improving science achievement in Kosovo.

Feedback Practices as a School-Level Predictor

Contrary to H5 feedback practices was not a significant predictor of science achievement at the school level (β = .022, p = .755). The non-significant finding for feedback practices was robust across all ten plausible values and ten imputed datasets, confirmed by both the magnitude of the coefficient (b = 3.951) and its negligible standardized effect (β = .022), suggesting that this result reflects a genuine absence of association between school-level feedback practices and science achievement in Kosovo after accounting for adaptive instruction rather than insufficient statistical power. Several explanations

may account for this finding. First, the school-level aggregation of teacher feedback scores, while methodologically justified, may obscure within-school variation in feedback quality and frequency that is more consequential for student learning than the average school-level feedback score. Second, the moderate positive correlation between adaptive instruction and feedback practices at the school level (r = .65) suggests that these two instructional constructs are substantially overlapping in Kosovo's schools, making it difficult to detect the independent contribution of feedback practices after adaptive instruction is already in the model. Third, it is possible that in Kosovo's predominantly teacher-centered instructional environment feedback is delivered in ways that do not align with the formative assessment practices most strongly linked to student learning in the international research literature (Onwunyili & Onwunyili, 2019). Future research should examine specific dimensions of feedback quality rather than overall feedback frequency to better understand the conditions under which teacher feedback supports science achievement in Kosovo.

Cross-Level Interactions

Contrary to H6 and H7 neither cross-level interaction was statistically significant. The interaction between curiosity and adaptive instruction (b = 1.127, p = .528) indicated that the positive association between curiosity and science achievement did not vary significantly across schools with different levels of adaptive instruction. The interaction between creativity and feedback practices (b = 0.333, p = .848) similarly indicated that the relationship between creativity and science achievement did not differ significantly across schools with different feedback practice levels. These non-significant interactions suggest that the benefits of curiosity and creativity for science achievement are relatively uniform across Kosovo schools regardless of the instructional environment, a finding that diverges from theoretical predictions derived from person-environment fit models suggesting that student dispositions yield stronger academic benefits in environments that support their expression. Two interpretations merit consideration. First, Kosovo's relatively homogeneous instructional environment characterized by predominantly teacher-centered approaches across most schools may produce insufficient between-school variation in adaptive instruction and feedback practices to detect meaningful cross-level moderation effects. Second, the range of adaptive instruction scores across Kosovo schools may be too narrow to produce detectable interaction effects even where theoretical moderation exists. Future research in contexts with greater between-school variation in instructional quality would provide a stronger test of the cross-level moderation hypotheses.

Variance Explained and Model Adequacy

The final model explained 2.7% of within-school variance, 19.1% of between-school variance, and 8.0% of total variance in science achievement. The substantially higher proportion of between-school variance explained relative to within-school variance reflects the dominant contribution of school-level adaptive instruction in accounting for between-school differences in science outcomes. The relatively modest within-school R^2 of 2.7% is consistent with findings from large-scale assessment research where

unmeasured factors including prior achievement, cognitive ability, and specific classroom experiences account for substantial portions of within-school achievement variation (Murayama et al., 2013). The unexplained variance points to the need for more comprehensive models that incorporate prior achievement, school-level resource indicators, and longitudinal data capable of capturing the developmental dynamics among motivational dispositions, instructional quality, and science performance over time.

Implications for Policy and Practice

The findings carry several provisional implications for educational policy and practice in Kosovo. The robust positive association between curiosity and science achievement ($\beta = .084$, $p < .001$) suggests that learning environments designed to stimulate students' intrinsic motivation to explore questions and investigate scientific phenomena may support science performance even within teacher-centered instructional traditions. Integrating inquiry-oriented tasks, open-ended questions, and opportunities for student-directed exploration even within structured instructional formats may help cultivate the curiosity that this study identifies as the strongest student-level dispositional predictor of science achievement.

The dominant role of adaptive instruction as the strongest overall predictor of science achievement with a standardized effect ($\beta = .192$) exceeding all student-level predictors carries particularly strong policy implications. Investment in professional development programs that equip Kosovo teachers with the skills, knowledge, and pedagogical repertoire to adapt instruction to diverse student learning needs represents one of the most promising systemic levers for improving science outcomes at scale. Evidence from classroom research similarly demonstrates that responsive instructional practices and quality feedback are central to student academic development (Onwunyili & Onwunyili, 2019; Chinedu & Chinwendu, 2020; Onwunyili & Ifenyinwa, 2020). Given that adaptive instruction explained 19.1% of between-school variance in science achievement even modest improvements in the quality of differentiated instruction across Kosovo schools could translate into meaningful reductions in between-school achievement inequality.

The persistent significance of socioeconomic status ($\beta = .078$, $p < .001$) underscores that dispositional and instructional interventions are unlikely to produce equitable improvements in science achievement without concurrent attention to the structural inequalities that constrain learning opportunities for students from socioeconomically disadvantaged backgrounds. Equity-focused investments in school resources including access to laboratory equipment, digital learning tools, and qualified science teachers are necessary complements to instructional quality improvements.

Limitations

Several limitations warrant consideration when interpreting the findings of this study.

First, the cross-sectional design precludes causal inference, the associations reported here are correlational in nature and should not be interpreted as evidence of causal influence. Students' science achievement may shape their dispositional characteristics, dispositional characteristics may shape achievement, or both may reflect the influence of unobserved third variables such as prior academic performance, general cognitive ability, or family learning environment. Similarly, the association between adaptive instruction and science achievement may reflect selection effects whereby more effective schools both attract better teachers and produce higher achievement rather than a direct instructional effect. Future research employing longitudinal or experimental designs would provide stronger grounds for causal inference about the relationships examined here.

Second, several measurement limitations qualify the school-level findings. The school-level sampling weight (W_SCHGRNRABWT) was extracted from the PISA 2022 teacher questionnaire file rather than the school questionnaire file, which was unavailable in the current dataset. While W_SCHGRNRABWT is a valid school-level weight variable and all school-level variables required for the present study were available in the teacher file, future analyses should incorporate weights directly from the school questionnaire file where possible. Additionally, the moderate positive correlation between adaptive instruction and feedback practices at the school level ($r = .65$) indicates substantial overlap between these two constructs, raising questions about multicollinearity and the ability to cleanly separate their independent contributions to science achievement. Future research should employ confirmatory factor analysis at the school level to establish the discriminant validity of these constructs before including them simultaneously as predictors.

Third, the final model explained 8.0% of total variance in science achievement leaving the majority of variance unexplained by the included predictors. Important unmeasured variables include prior academic achievement, general cognitive ability, parental involvement, curriculum coverage, access to laboratory and digital learning resources, and specific classroom-level instructional practices. The absence of prior achievement is particularly consequential as longitudinal research consistently shows that controlling for baseline performance substantially reduces the apparent contributions of motivational and instructional predictors to achievement outcomes (Murayama et al., 2013). Additionally, Kosovo's split-shift schooling system in which most students attend school for approximately half a day represents an important structural feature not captured in PISA data that may account for a meaningful portion of the between-school variance observed in this study. The effect sizes reported here should therefore be understood as upper-bound estimates and future research should incorporate baseline achievement measures and structural school context variables to obtain more precise and comprehensive estimates of the dispositional and instructional pathways examined here.

Conclusion

This study applied a weighted two-level multilevel modeling approach to examine the direct effects of student dispositional characteristics and school-level instructional practices on science achievement among Kosovo students participating in PISA 2022. By incorporating both student-level and school-level

sampling weights through the WeMix package (Bailey et al., 2023) and properly handling all ten plausible values using Huang's (2024) mixPV wrapper function the study produced population-representative parameter estimates with appropriately estimated standard errors that reflect both sampling uncertainty and plausible value measurement uncertainty simultaneously.

The intraclass correlation of .324 confirmed that 32.4% of total variance in science achievement was attributable to between-school differences, providing strong empirical justification for the multilevel approach and demonstrating that the degree of school-level clustering in Kosovo's science achievement is substantially higher than estimates from analyses that do not properly account for plausible value measurement uncertainty.

At the student level curiosity emerged as the strongest dispositional predictor of science achievement ($\beta = .084$, $p < .001$) with a robust positive association that persisted after controlling for all other predictors, confirming curiosity as a relatively context-invariant motivational resource whose academic benefits are not entirely contingent on inquiry-supportive instruction. Creativity also demonstrated a significant positive association with science achievement ($\beta = .046$, $p = .005$) establishing that students' perceptions of creative opportunities within their school environment contribute independently to science performance beyond the contribution of curiosity and socioeconomic factors.

At the school level adaptive instruction emerged as the strongest predictor in the entire model ($\beta = .192$, $p = .005$) stronger than any individual student-level predictor, explaining 19.1% of between-school variance in science achievement. This finding positions the quality of teachers differentiated instructional practices as the most powerful lever available to Kosovo's educational system for reducing between-school achievement inequality and improving science outcomes at scale. Feedback practices did not demonstrate a significant independent association after accounting for adaptive instruction and cross-level interactions between student dispositions and school-level instructional practices were not significant suggesting that the academic benefits of curiosity and creativity are relatively uniform across Kosovo schools regardless of instructional environment.

This study contributes to international literature in three ways. First it establishes curiosity and creativity as significant student-level predictors of science achievement in a persistently low-performing non-Western educational context extending prior research conducted predominantly in high-performing Western systems. Second it demonstrates that adaptive instruction is a powerful school-level predictor that accounts for substantially more between-school variance than student-level dispositional characteristics. Third it demonstrates the methodological importance of proper plausible value handling and two-level weighting in multilevel PISA analyses showing that ICC estimates and standard errors differ substantially between correct and incorrect analytical approaches with important implications for the accuracy of policy-relevant inferences.

For Kosovo's educational stakeholders the findings point to two complementary priorities: strengthening teachers' adaptive instructional capacity through sustained professional development and cultivating students' curiosity through instructional practices that reward intellectual exploration. Evidence from

classroom research consistently demonstrates that responsive instructional practices and student motivational engagement are mutually reinforcing conditions for academic success (Onwunyili & Onwunyili, 2019; Chinedu & Chinwendu, 2020; Onwunyili & Ifenyinwa, 2020). Advancing science education in Kosovo will require sustained commitment to both teacher development and student motivational development guided by theoretically grounded empirical evidence that takes seriously the cultural and structural contexts in which learning unfolds. Several limitations qualify these conclusions including the cross-sectional design precluding causal inference, the school-level weight source, and the presence of unmeasured confounders including Kosovo's split-shift schooling system. Future research should employ longitudinal designs, replicate weight-based approaches for fully design-consistent inference, and incorporate structural school context variables to more comprehensively account for the sources of science achievement variation in Kosovo.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Clinical trial number

Not applicable.

Ethics approval and consent to participate

This study utilized publicly available deidentified secondary data from the Organisation for Economic Cooperation and Development PISA 2022 international database. No primary data collection involving human participants was conducted. All data were collected by the Organisation for Economic Cooperation and Development under their established ethical protocols and participant consent procedures. Formal institutional ethics approval was not required for this study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are publicly available from the Organisation for Economic Cooperation and Development PISA 2022 database. The student questionnaire dataset (CY08MSP_STU_QQQ.SAV) and teacher questionnaire dataset (CY08MSP_TCH_QQQ.SAV) used in this study can be freely accessed at https://www.oecd.org/pisa/data/2022database/. No additional datasets were generated during the current study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

F. O., E. A., and J. M. conceptualized the study, conducted data curation, formal analysis, and methodology, and wrote the original draft, as well as contributed to reviewing and editing the manuscript. Adam Maltese provided supervisory oversight and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to express sincere gratitude to Professor Adam Maltese of Indiana University for his invaluable guidance, constructive feedback, and scholarly mentorship throughout the development of this manuscript. The authors also wish to thank Professor Leslie Rutkowski of Indiana University for her expert methodological guidance on the application of weighted multilevel modeling using the WeMix package and the proper handling of plausible values using the mixPV function, which substantially strengthened the analytical rigor of this work.

Authors' information

Fidelis Onwunyili is a PhD student in the Department of Curriculum and Instruction with specialization in Science Education at Indiana University, Bloomington. Email: fconwuny@iu.edu. ORCID: 0009-0006-3154-0597

Emmanuel Atiatorme is a PhD student in the Department of Curriculum and Instruction with specialization in Science Education at Indiana University, Bloomington. Email: ematia@iu.edu. ORCID: 0000-0002-4655-8750

John Mensa is a PhD student in the Department of Curriculum and Instruction with specialization in Science Education at Indiana University, Bloomington. Email: jamoako@iu.edu. ORCID: 0009-0006-2221-3631

Adam Maltese is a Professor in the Department of Curriculum and Instruction at Indiana University, Bloomington. Email: amaltese@iu.edu. ORCID: 0000-0001-8422-9395

References

1. Bailey P, Webb B, Kelley C, Nguyen T, Huo H. *WeMix: Weighted mixed-effects models using multilevel pseudo maximum likelihood* estimation. R package version. 2023. https://doi.org/10.32614/CRAN.package.WeMix. 4.0.3.
2. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika. 1999;86(4):948–55. https://doi.org/10.1093/biomet/86.4.948.

3. Black P, Wiliam D. Assessment and classroom learning. Assess Education: Principles Policy Pract. 1998;5(1):7–74. https://doi.org/10.1080/0969595980050102.
4. Bronfenbrenner U. The ecology of human development: Experiments by nature and design. Harvard University Press; 1979.
5. Bronfenbrenner U, Morris PA. The bioecological model of human development. In: Lerner RM, Damon W, editors. Handbook of child psychology: Theoretical models of human development. Volume 1, 6th ed. Wiley; 2006. pp. 793–828.
6. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates; 1988.
7. Chinedu OF, Chinwendu OM. (2020). Relationship between effective classroom management and students' academic performance in biology in Anambra West LGA of Anambra State, Nigeria. Bulgarian J Sci Educ Policy, *14*(2).
8. Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychol Methods. 2007;12(2):121–38. https://doi.org/10.1037/1082-989X.12.2.121.
9. Floden RE. The measurement of opportunity to learn. In: Porter AC, Gamoran A, editors. Methodological advances in cross-national surveys of educational achievement. National Academy; 2002. pp. 231–66.
10. Grajcevci A, Shala A. A review of Kosovo's 2015 PISA results: Analysing the impact of teacher characteristics in student achievement. Int J Instruction. 2021;14(1):489–506. https://doi.org/10.29333/iji.2021.14129a.
11. Hattie J. Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge; 2009.
12. Huang FL. Using plausible values when fitting multilevel models with large-scale assessment data using R. Large Scale Assessments Educ. 2024;12:7. https://doi.org/10.1186/s40536-024-00192-0.
13. Hunter JE, Schmidt FL. Methods of meta-analysis: Correcting error and bias in research findings. 2nd ed. Sage; 2004.
14. Jirout J, Klahr D. Children's scientific curiosity: In search of an operational definition of an elusive concept. Dev Rev. 2012;32(2):125–60. https://doi.org/10.1016/j.dr.2012.04.002.
15. Kreft I, de Leeuw J. (1998). *Introducing multilevel modeling.* Sage. https://doi.org/10.4135/9781849209366
16. Kosovo Education Center (KEC). (2018). *The impact of teacher quality and in-school resources on Kosovar students' performance: Findings from PISA 2015 results.* Kosovo Education Center. https://kec-ks.org/wp-content/uploads/2018/10/PISA-2015-Results_ENG.pdf
17. Le Donné N, Fraser P, Bousquet G. Teaching strategies for instructional quality: Insights from the TALIS–PISA link data. OECD Publishing. 2016. https://doi.org/10.1787/9789264258075-en.
18. LeBreton JM, Senter JL. Answers to twenty questions about interrater reliability and interrater agreement. Organizational Res Methods. 2008;11(4):815–52.

https://doi.org/10.1177/1094428106296642.

19. Liu Z, Guo H, Zhou Z, Ma F, Zeng Y. How creative self-efficacy influences problem-solving skills in engineering education: The dual mediating role of critical thinking and metacognition. BMC Psychol. 2025;13(1):1278. https://doi.org/10.1186/s40359-025-02578-z.
20. Lumley T. (2020). *mitools: Tools for multiple imputation of survey data* (R package version 2.4). https://CRAN.R-project.org/package=mitools
21. Maas CJM, Hox JJ. Sufficient sample sizes for multilevel modeling. Methodology. 2005;1(3):86–92. https://doi.org/10.1027/1614-2241.1.3.86.
22. Mathieu JE, Aguinis H, Culpepper SA, Chen G. Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. J Appl Psychol. 2012;97(5):951–66. https://doi.org/10.1037/a0028380.
23. Murayama K, Pekrun R, Lichtenfeld S, vom Hofe R. Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. Child Dev. 2013;84(4):1475–90. https://doi.org/10.1111/cdev.12036.
24. Nilsen T, Gustafsson JE. Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time. Springer Nat. 2016. https://doi.org/10.1007/978-3-319-41252-8.
25. OECD. PISA 2022 results (I): The state of learning and equity in education. OECD Publishing. 2023. https://doi.org/10.1787/53f23881-en.
26. OECD. PISA 2022 results (III): Creative minds, creative schools. OECD Publishing. 2024. https://doi.org/10.1787/765ee8c2-en.
27. Onwunyili FC, Ifenyinwa OBI. Self-regulatory strategies and extrinsic motivation as predictors of effective learning of statistics among undergraduate students. Social Sci Educ Res Rev. 2020;7(2):107–25.
28. Onwunyili FC, Onwunyili MC. Assessment of classroom teachers' feedback in formative tests. Southeast J Res Sustainable Dev. 2019;2(1):166–81.
29. Onwunyili FC, Onwunyili MC. (2020). Effect of self-regulated learning on test anxiety: Academic achievement and metacognition among secondary school students in Anambra State. Available at SSRN 5489046.
30. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods. 2nd ed. Sage; 2002.
31. Rubin DB. Multiple imputation for nonresponse in surveys. Volume 81. Wiley; 2004.
32. Rutkowski L, Gonzalez E, Joncas M, Von Davier M. International large-scale assessment data: Issues in secondary analysis and reporting. Educational researcher. 2010;39(2):142–51.
33. Schoevers EM, Kroesbergen EH, Moerbeek M, Leseman PP. The relation between creativity and students' performance on different types of geometrical problems in elementary education. ZDM–Mathematics Educ. 2022;54(1):133–47. https://doi.org/10.1007/s11858-021-01315-5.

34. Sergeant-Perthuis G, Rudrauf D, Tisserand Y. Curiosity driven exploration through perspective transformation. Front Neurorobotics. 2022;16:1–15. https://doi.org/10.3389/fnbot.2022.737301.
35. Snijders TAB, Bosker RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. 2nd ed. Sage; 2012.
36. Spearman C. The proof and measurement of association between two things. Am J Psychol. 1904;15(1):72–101. https://doi.org/10.2307/1412159.
37. Uğraş H, Uğraş M, Papadakis S, Kalogiannakis M. ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices. Sustainability. 2024a;16(22). Article 9855. https://doi.org/10.3390/su162209855.
38. Uğraş H, Uğraş M, Papadakis S, Kalogiannakis M. (2024b). Innovative early childhood STEM education with ChatGPT: Teacher perspectives. *Technology, Knowledge and Learning.* Advance online publication. https://doi.org/10.1007/s10758-024-09804-8
39. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67. https://doi.org/10.18637/jss.v045.i03.
40. Von Stumm S, Hell B, Chamorro-Premuzic T. The hungry mind: Intellectual curiosity is the third pillar of academic performance. Perspect Psychol Sci. 2011;6(6):574–88. https://doi.org/10.1177/1745691611421204.
41. Zabelina DL, Robinson MD. Creativity as flexible cognitive control. Psychol Aesthet Creativity Arts. 2010;4(3):136–45. https://doi.org/10.1037/a0017379.