iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

MODELLING STUDENT LEARNING PROGRESSION USING BAYESIAN HIERARCHICAL METHODS AND LONGITUDINAL ASSESSMENT DATA ACROSS MATHEMATICS EDUCATION SYSTEMS

Israel Safo
Invictus Academy at Airways, Memphis, Tennessee, USA

ABSTRACT
Understanding how mathematical knowledge develops over time is essential for designing responsive instruction, evaluating educational interventions, and identifying persistent disparities in student achievement. Conventional assessment models often rely on cross-sectional performance measures that inadequately represent individual learning trajectories or the nested structures of educational systems. This study develops a Bayesian hierarchical framework for modelling student learning progression using repeated mathematics assessment scores collected from the same students across multiple academic years. The framework integrates longitudinal achievement records with student background characteristics and classroom- and school-level variables, enabling simultaneous estimation of individual growth trajectories and contextual variation. Student-specific intercepts and growth parameters capture heterogeneity in baseline proficiency and rates of mathematical development, while hierarchical effects quantify differences attributable to classrooms and schools. Bayesian posterior distributions provide probabilistic estimates of learning progression, uncertainty, and the influence of educational and background factors. Longitudinal model evaluation assesses predictive accuracy, between-student variation, and progression patterns across mathematics education settings. The proposed approach provides a statistically coherent framework for identifying learning trajectories and supporting evidence-informed educational planning, targeted intervention, and mathematics achievement monitoring.

Keywords:
Bayesian hierarchical modelling; learning progression; longitudinal assessment; mathematics achievement; student growth trajectories; educational analytics

1. INTRODUCTION

1.1 Longitudinal Measurement of Mathematics Learning

Mathematics achievement develops cumulatively as students acquire, consolidate, and extend knowledge across successive academic years [1]. Conventional assessment reporting, however, frequently represents achievement through scores obtained at isolated points in time, providing limited evidence about how individual students progress between assessments. Cross-sectional comparisons can describe differences between groups or cohorts but cannot directly reconstruct within-student developmental trajectories [2]. Longitudinal assessment addresses this limitation by repeatedly measuring the same students, enabling changes in achievement to be distinguished from differences between students. Such repeated observations provide information about both initial proficiency and subsequent rates of development [3]. Within this framework, mathematics learning progression can be conceptualised as an underlying developmental process that is observed imperfectly through assessment scores collected at successive time points.

1.2 Limitations of Conventional Student-Growth Models

Simple score differences provide an intuitive measure of change but can be sensitive to measurement error and differences in assessment timing [4]. Ordinary least-squares growth models offer greater flexibility but may inadequately represent the dependence created by repeated observations from the same students. Cross-sectional models present a further limitation because between-student differences cannot be interpreted as within-student learning. Educational data also have multiple dependency structures: assessments are repeated within students, while students are situated within classrooms and schools [5]. Ignoring these structures can distort estimated variability and uncertainty. Complete-case analysis may additionally discard students with partially observed trajectories, potentially reducing precision and altering the analysed sample [6]. Finally, reporting growth through point estimates alone provides insufficient information about uncertainty surrounding individual trajectories and contextual effects.

1.3 Bayesian Hierarchical Modelling for Educational Progression

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

Bayesian hierarchical modelling provides a coherent framework for representing the nested and longitudinal structure of mathematics achievement [7]. Repeated assessment observations can be used to estimate each student's baseline mathematical proficiency and rate of progression while simultaneously modelling variation associated with classrooms, schools, and relevant background characteristics. Rather than estimating every student or educational cluster independently, hierarchical models use partial pooling, allowing information to be shared across related units while retaining meaningful individual differences [8]. This is particularly valuable when classroom and school sample sizes are unequal because estimates from sparsely observed groups can be regularised toward the population distribution instead of being driven disproportionately by limited observations. Bayesian estimation additionally produces posterior distributions for model parameters, learning trajectories, and predictions, permitting uncertainty to be propagated throughout the analysis rather than appended only after point estimation.

1.4 Aim, Research Questions and Contributions

This study develops a Bayesian hierarchical framework for modelling longitudinal mathematics-learning progression from repeated student assessment records [9]. It addresses four questions: RQ1, how do individual mathematics-learning trajectories vary across time? RQ2, how much variation is attributable to students, classrooms, and schools? RQ3, which student-background and educational characteristics are associated with progression? RQ4, how does the Bayesian hierarchical model compare with benchmark models in longitudinal prediction and uncertainty characterisation? The study contributes an integrated approach to estimating individual growth, contextual variation, and posterior uncertainty while incorporating standards-based interpretation to relate estimated learning trajectories to meaningful mathematics-achievement expectations.

2. THEORETICAL AND STATISTICAL FRAMEWORK

2.1 Mathematics Learning Progression as a Longitudinal Process

Longitudinal mathematics assessment distinguishes the proficiency underlying student learning from the observed score recorded on any single assessment occasion [5]. For student $i$at time t, observed achievement can be represented as

$$Y_{it} = \theta_{it} + \epsilon_{it} \qquad (1)$$

where Y_{it}denotes the observed mathematics assessment score, θ_{it}represents the student's underlying mathematics proficiency, and ϵ_{it}captures measurement and residual error. This distinction is important because an assessment score is an imperfect manifestation of proficiency rather than a direct and error-free measurement of learning [6]. Repeated observations provide stronger evidence about progression because persistent developmental patterns can be distinguished from occasion-specific score fluctuations. Consequently, learning progression is treated as an evolving latent process observed indirectly through a sequence of mathematics assessments [7].

2.2 Individual Growth Trajectory

A student's underlying mathematics progression is initially represented using a linear growth trajectory:

$$\theta_{it} = \alpha_i + \beta_i t \qquad (2)$$

where $\alpha_i$represents student-specific baseline proficiency and $\beta_i$represents the student's rate of mathematics growth over time [8]. Individual parameters are subsequently expressed as departures from population-average growth:

$$\alpha_i = \alpha_0 + a_i$$
$$\beta_i = \beta_0 + b_i \qquad (3)$$

Here, $\alpha_0$and $\beta_0$denote population-level baseline proficiency and growth, while $a_i$and $b_i$capture individual deviations [9]. The random intercept therefore represents whether a student begins above or below the population baseline, whereas the random slope represents whether mathematics achievement subsequently develops faster or slower than the average trajectory [10]. This formulation accommodates heterogeneous learning pathways rather than imposing identical progression on all students.

2.3 Hierarchical Educational Structure

Individual trajectories occur within educational contexts and should therefore be modelled jointly with classroom and school clustering [11]. For assessment occasion t, student i, classroom j, and school k, the central model is specified as

$$Y_{ijkt} = \beta_0 + \beta_1 t + \gamma^T X_{ijkt} + a_i + b_i t + u_{jk} + v_k + \epsilon_{ijkt} \qquad (4)$$

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

where $\beta_0$represents population baseline mathematics proficiency and $\beta_1$represents the population learning rate. The vector X_{ijkt}contains justified student-background and educational predictors, with γrepresenting their associated coefficients [12]. Student-specific baseline heterogeneity is captured by a_i, while $b_i$represents deviation from the population growth rate. The classroom effect u_{jk}captures contextual variation among classrooms within schools, and $v_k$represents between-school variation. Finally, ϵ_{ijkt}captures unexplained occasion-level variability. Modelling these components simultaneously prevents repeated observations and educational clustering from being treated as independent sources of information [13].

2.4 Bayesian Formulation and Partial Pooling

The hierarchical model is completed by assigning probability distributions to its parameters. Weakly informative priors can be specified, for example, as

$$\beta_p \sim \mathcal{N}\left(0, \sigma_\beta^2\right),\ \sigma_\beta > 0 \qquad (5)$$

with hierarchical effects represented by

$$a_i \sim \mathcal{N}(0, \sigma_a^2),\ b_i \sim \mathcal{N}(0, \sigma_b^2),\ u_{jk} \sim \mathcal{N}(0, \sigma_C^2),\ v_k \sim \mathcal{N}(0, \sigma_S^2) \qquad (6)$$

and suitable positive priors assigned to variance parameters [14]. Bayesian inference combines these prior distributions with information contained in the longitudinal assessment data:

$$p(\Theta \mid Y) = \frac{p(Y \mid \Theta)p(\Theta)}{p(Y)} \propto p(Y \mid \Theta)p(\Theta) \qquad (7)$$

Conceptually, this corresponds to Prior × Likelihood → Posterior. The resulting posterior distribution represents updated uncertainty about learning trajectories and model parameters after observing the assessment data [15]. Hierarchical estimation produces partial pooling: estimates for students, classrooms, and schools with limited observations are regularised toward their corresponding population distributions, while units supported by more information retain greater data-driven specificity. This is particularly valuable for unequal educational cluster sizes.

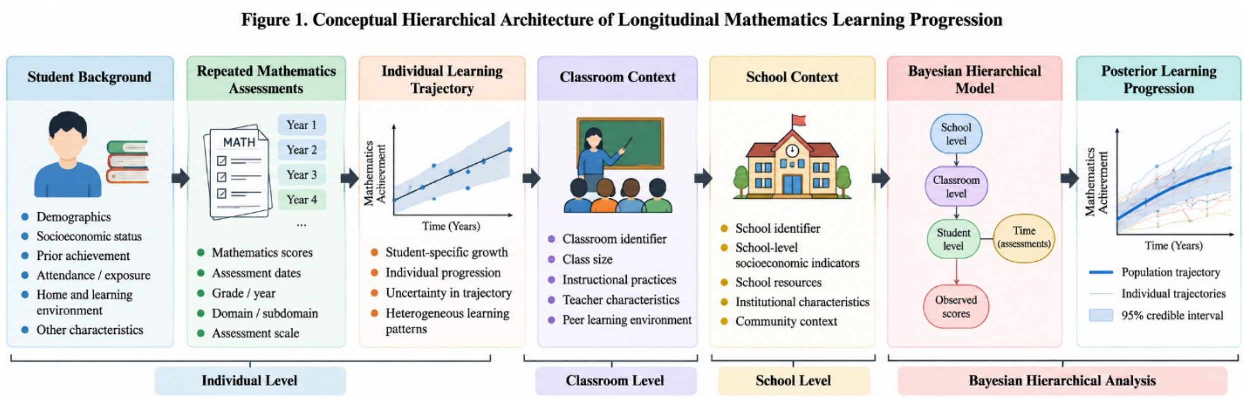

Figure 1. Conceptual Hierarchical Architecture of Longitudinal Mathematics Learning Progression

3. DATA ACQUISITION, PREPARATION AND FEATURE ENGINEERING

3.1 Longitudinal Data Acquisition

The analytical dataset requires repeated mathematics assessments from the same students across multiple academic years so that within-student change can be distinguished from differences between students [13]. Assessment records should contain mathematics scores, administration dates, grade or year, mathematics domain or subdomain, and the applicable measurement scale [14]. Student records require persistent identifiers for longitudinal linkage, together with prior achievement, available demographic and socioeconomic characteristics, attendance, and other relevant educational information [15]. Classroom information should identify classroom membership, class size, instructional characteristics, and teacher-related variables where available. Because students' learning trajectories occur within broader educational environments, school identifiers should be linked with available socioeconomic, resource, and institutional characteristics [16]. Assessment occasion and elapsed time must also be preserved explicitly rather than inferred solely from grade progression. Where the assessment

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

system supplies achievement standards or proficiency classifications, these should be retained to support standards-based interpretation of posterior learning trajectories [17].

Table 1. Longitudinal Mathematics Data Acquisition and Analytical Variables

Domain	Example Variables	Level	Analytical Role
Assessment	Score, year, grade, mathematics domain	Observation	Outcome/measurement
Student	Prior score, background, attendance	Student	Individual predictors
Classroom	Class size, instructional characteristics	Classroom	Contextual effects
School	Resources, socioeconomic/institutional characteristics	School	Contextual effects
Temporal	Assessment occasion, elapsed time	Observation	Growth estimation
Standards	Proficiency threshold/category	Assessment	Achievement benchmark

3.2 Data Cleaning and Longitudinal Alignment

Reliable estimation of learning progression depends on reconstructing each student's assessment history in its correct temporal sequence [18]. Duplicate observations should be identified and resolved using predefined criteria, while scores outside permissible assessment ranges require verification rather than automatic retention. Persistent student identifiers must remain consistent across academic years to prevent different students from being incorrectly joined or repeated observations from being separated [13]. Missing assessment occasions should be retained as explicit gaps in the longitudinal structure. Changes in classroom or school membership should similarly be recorded because mobility may represent genuine educational transitions rather than data errors [15]. When mathematics assessments use different scales across grades or years, appropriate 1149tandardizati may be required. Period-specific 1149tandardization can be expressed as

$$Z_{it} = \frac{Y_{it} - \mu_t}{\sigma_t} \qquad (8)$$

where $\mu_t$and $\sigma_t$must correspond to the relevant assessment period and scale [14]. Using statistics derived from later assessment periods would introduce future information into earlier observations.

3.3 Feature Engineering

Feature construction should preserve the chronology of student learning and represent information available at each assessment occasion [19]. Baseline mathematics achievement identifies the student's initial observed position, whereas lagged achievement represents the most recent mathematics score preceding the outcome under analysis:

$$Y_{i,t-1} \qquad (9)$$

Lagged scores are particularly informative in longitudinal educational analysis because previous achievement is strongly connected to subsequent performance while preserving temporal ordering [16]. Elapsed academic time should be calculated from assessment dates or valid academic intervals, especially when assessments are not equally spaced. Observed interval growth can be expressed as

$$G_{it} = \frac{Y_{it} - Y_{i,t-1}}{t_{it} - t_{i,t-1}} \qquad (10)$$

to describe the magnitude of score change per unit of elapsed time [17]. Additional features can represent prior growth, attendance or instructional exposure, socioeconomic/background characteristics, classroom context, school characteristics, mathematics domains, and theoretically justified interactions [18]. Contextual features should reflect information genuinely available from the study source rather than variables constructed without empirical support.

A complementary descriptive indicator is the student's deviation from the relevant reference mean:

$$MD_{it} = Y_{it} - \bar{Y}_{g,t} \qquad (11)$$

where $\bar{Y}_{g,t}$represents the appropriate mathematics-achievement mean for grade $g$and assessment period t[20]. Positive and negative deviations indicate achievement above and below the contemporaneous reference mean, respectively. Mean deviation does not constitute the longitudinal learning model; it provides a transparent descriptive benchmark against which model-based progression can later be interpreted [14].

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

3.4 Missingness, Scaling and Leakage Prevention

Missing observations require explicit treatment because incomplete longitudinal records can arise through absence, mobility, assessment non-participation, or incomplete data linkage [15]. Missingness should therefore be examined in relation to mechanisms consistent with missing completely at random, missing at random, and potentially informative missingness [19]. Routine complete-case deletion can unnecessarily remove students who retain useful partial trajectories and may change the composition of the analytical sample [13]. Where assumptions are defensible, Bayesian modelling permits missing quantities to be represented as unknown parameters and estimated jointly with other model components [20]. Scaling parameters must also be obtained from the appropriate analytical or training period. Future mathematics scores, subsequent attendance, later classroom characteristics, or transformations calculated using future observations must not enter predictors used to estimate an earlier outcome [18]. This temporal separation is essential for obtaining credible out-of-sample estimates of learning progression.

3.5 Exploratory Longitudinal Visualisation

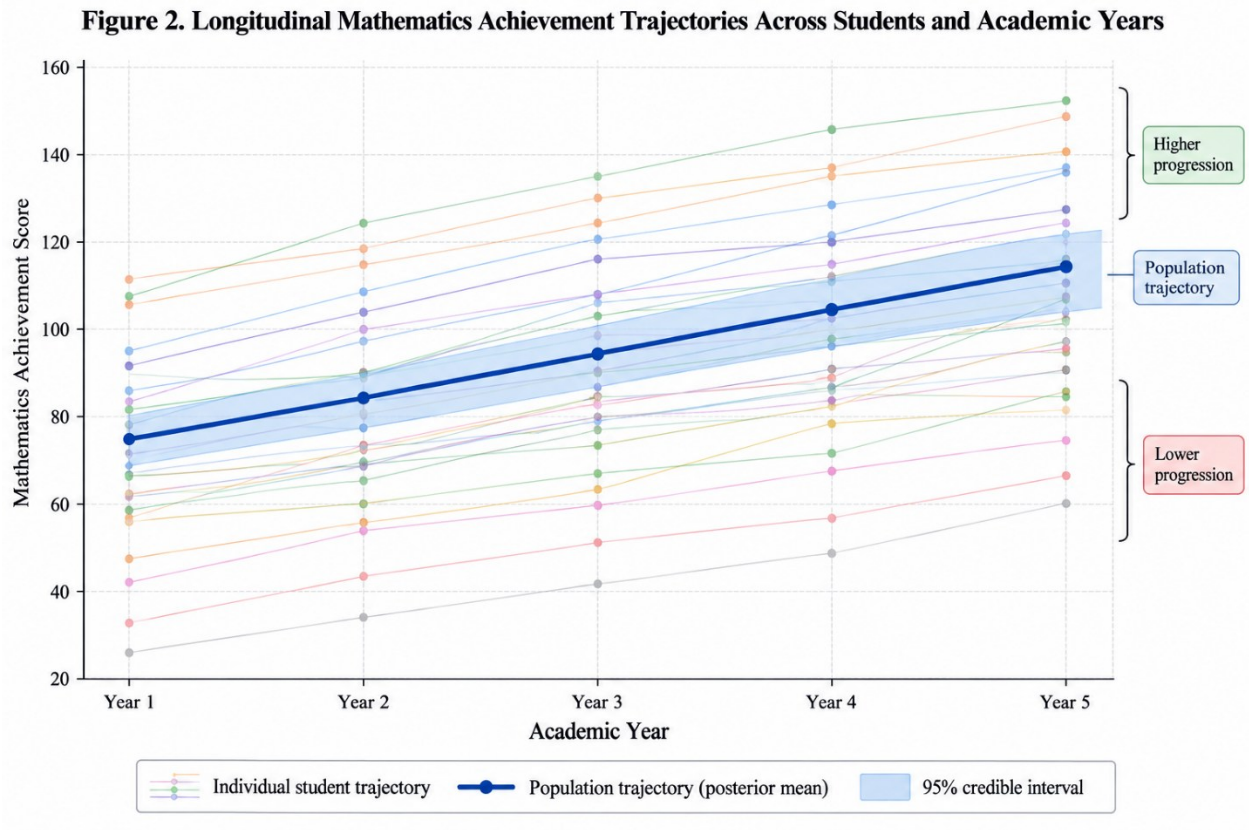

Figure 2. Longitudinal Mathematics Achievement Trajectories Across Students and Academic Years

Individual student trajectories should be displayed across successive assessment occasions to reveal differences in starting proficiency, direction and rate of achievement change [16]. An overall estimated progression curve can be superimposed to distinguish population-level development from heterogeneous individual trajectories [17]. Students with incomplete measurement sequences should remain visible where analytically appropriate, thereby showing the actual longitudinal structure rather than only complete cases [20].

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

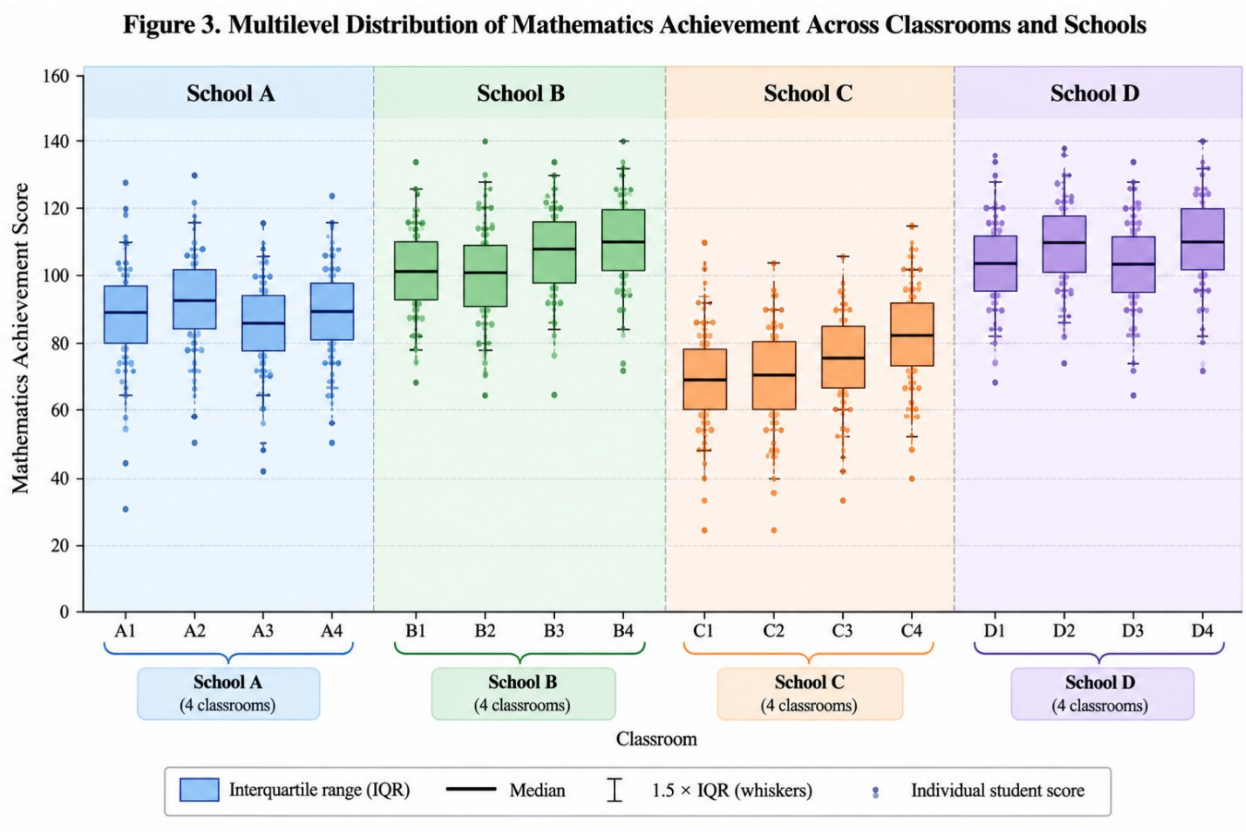

Figure 3. Multilevel Distribution of Mathematics Achievement Across Classrooms and Schools

Mathematics-score distributions should also be examined across classrooms nested within schools [14]. Displaying distributions and uncertainty intervals rather than aggregate means alone can reveal whether apparently similar schools contain substantially different classroom-level achievement patterns [18]. This preliminary heterogeneity provides an empirical basis for introducing student-, classroom-, and school-level effects within the subsequent Bayesian hierarchical model [19].

4. BAYESIAN MODEL DEVELOPMENT, TRAINING AND HYPERPARAMETER OPTIMISATION

4.1 Temporal Data Splitting

Model development should preserve the chronological structure of mathematics learning rather than randomly allocating assessment rows across training and evaluation sets [18]. Random row-level splitting is inappropriate when repeated observations from the same student can occur on both sides of the partition because later achievement may indirectly inform estimation of earlier outcomes [19]. The longitudinal dataset should therefore be separated chronologically into training, validation, and test periods. Conceptually, earlier academic years form the training set, a subsequent period is reserved for model validation and prior specification assessment, and the final available period remains an untouched test set [20]. Exact years and proportions should be determined by the eventual dataset, assessment frequency, and number of longitudinal observations. All transformations dependent on empirical distributions should be estimated without access to the future test period [21].

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

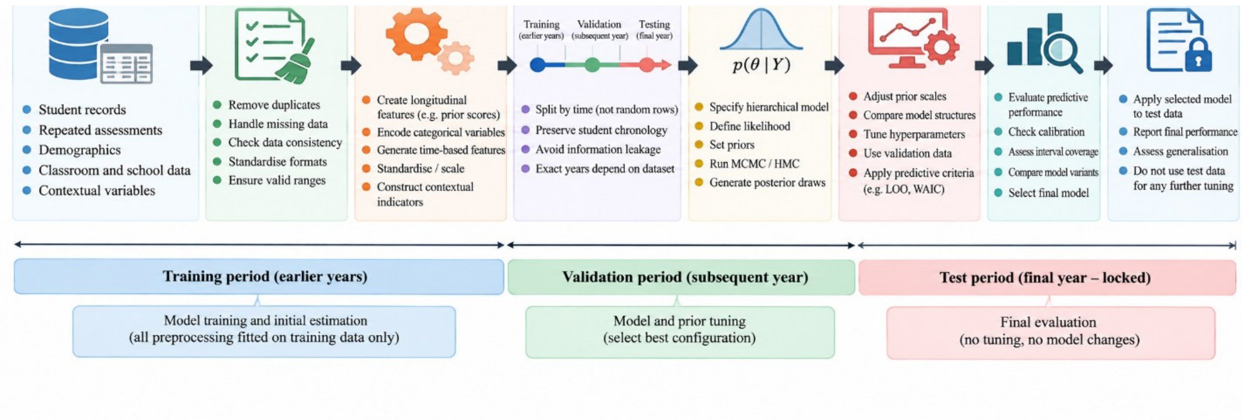

Figure 4. Leakage-Safe Longitudinal Training, Validation and Testing Pipeline

4.2 Likelihood and Model Training

For mathematics achievement represented on an approximately continuous scale, the observation model can be specified as

$$Y_{ijkt} \sim \mathcal{N}(\mu_{ijkt}, \sigma_\epsilon^2) \qquad (12)$$

where μ_{ijkt} is determined by the hierarchical longitudinal structure defined previously and σ_ϵ^2 captures residual assessment variability [22]. Posterior estimation may be performed using Markov chain Monte Carlo, including Hamiltonian Monte Carlo where supported by the selected implementation [23]. Multiple independent chains should be initialised and subjected to an appropriate warm-up phase before retained posterior draws are used for inference. Warm-up enables sampler adaptation and is not itself treated as inferential output [24]. The retained draws approximate the joint posterior distribution of fixed effects, student growth parameters, classroom and school effects, and variance components. Model training should therefore preserve posterior uncertainty rather than reducing estimation to a single optimal coefficient vector. Effective sample size, between-chain agreement, sampling efficiency, and convergence diagnostics should be evaluated before substantive interpretation [25].

4.3 Hyperparameter and Prior Tuning

Model development requires careful specification of priors and structural choices rather than unrestricted optimisation against predictive performance alone [26]. Candidate decisions include fixed-effect prior scales, priors governing random-effect standard deviations, the residual-scale prior, random-slope structures, theoretically justified interactions, and sampler settings [18]. Weakly informative priors can constrain implausible parameter regions while allowing the longitudinal data to provide substantial information about learning trajectories [22]. Their suitability should first be examined through prior predictive simulation to determine whether implied mathematics scores and growth rates are educationally plausible [19]. Where alternative model configurations are compared, a general selection criterion can be written as

$$\lambda^* = \arg\min_{\lambda \in \Lambda} \mathcal{L}_{\mathrm{val}}(\lambda) \qquad (13)$$

where λ denotes a candidate model or prior configuration and $\mathcal{L}_{\mathrm{val}}$ is an appropriate validation loss or predictive criterion [20]. Validation data or principled predictive criteria should guide these decisions; the locked test set must remain unavailable during model selection [21]. Importantly, prior specification should remain substantively justified rather than being treated identically to conventional machine-learning hyperparameter search.

4.4 Posterior Convergence and Diagnostics

Posterior estimates should not be interpreted until the quality of the sampling process has been demonstrated [23]. Convergence assessment should include the potential scale-reduction statistic $\hat{R}$, with well-behaved chains generally producing values very close to 1 [24]. Effective sample size should be reported because a large number of correlated MCMC draws may contain considerably less independent information than the nominal draw count suggests. Trace plots should demonstrate adequate mixing and stable exploration of the posterior distribution [25]. Hamiltonian Monte Carlo implementations should additionally report divergent transitions, because persistent divergences can indicate problematic posterior geometry. Monte Carlo standard errors quantify remaining simulation uncertainty in posterior summaries [26]. Posterior predictive checks should then compare replicated outcomes generated from the fitted model with relevant characteristics of observed mathematics achievement

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

[18]. Convergence should therefore be evidenced through complementary diagnostics rather than asserted from a single statistic.

4.5 Variance Decomposition

The hierarchical model enables unexplained mathematics-achievement variability to be separated across educational levels [19]. For a representation containing school, classroom, student, and occasion-level residual components, total unexplained variance can be written as

$$\sigma^2_{\text{Total}} = \sigma^2_S + \sigma^2_C + \sigma^2_I + \sigma^2_\epsilon \qquad (14)$$

where the terms denote school, classroom, student, and residual variance, respectively [22]. The proportional school contribution is therefore

$$VPC_S = \frac{\sigma^2_S}{\sigma^2_S + \sigma^2_C + \sigma^2_I + \sigma^2_\epsilon} \qquad (15)$$

with corresponding ratios calculated for classroom and student components [20]. The derivation follows directly from partitioning total unexplained variance into its hierarchical components and dividing each component by their sum. Posterior draws can additionally provide credible intervals for these proportions rather than treating variance partition coefficients as fixed quantities [23]. This decomposition indicates whether remaining heterogeneity in longitudinal mathematics achievement is concentrated primarily among students, classrooms, or schools after included predictors have been considered [25].

4.6 Posterior Learning Progression

The central output is the posterior distribution of mathematics-learning trajectories rather than a single fitted growth line [21]. For student i, posterior expected achievement at time $t$can be summarised as

$$\hat{Y}_{it}^{post} = E(Y_{it} \mid Y) \qquad (16)$$

where Ydenotes the observed longitudinal assessment information used to update the model [24]. Posterior draws also permit direct probability statements about individual learning rates. For a student-specific slope β_i, the probability of positive progression is

$$P(\beta_i > 0 \mid Y) = \int_0^\infty p\,(\beta_i \mid Y)\, d\beta_i \qquad (17)$$

[26]. This formulation moves interpretation beyond binary significant/non-significant conclusions. A student may have a positive posterior mean growth rate while still exhibiting substantial uncertainty about the magnitude or direction of progression [18]. Reporting posterior means or medians alongside 95% credible intervals and relevant posterior probabilities therefore communicates both estimated development and inferential uncertainty [22]. Population and contextual trajectories can be interpreted similarly, allowing progression to be examined simultaneously across students and their educational environments [25].

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

Figure 5. Posterior Student Learning Trajectories With 95% Credible Intervals

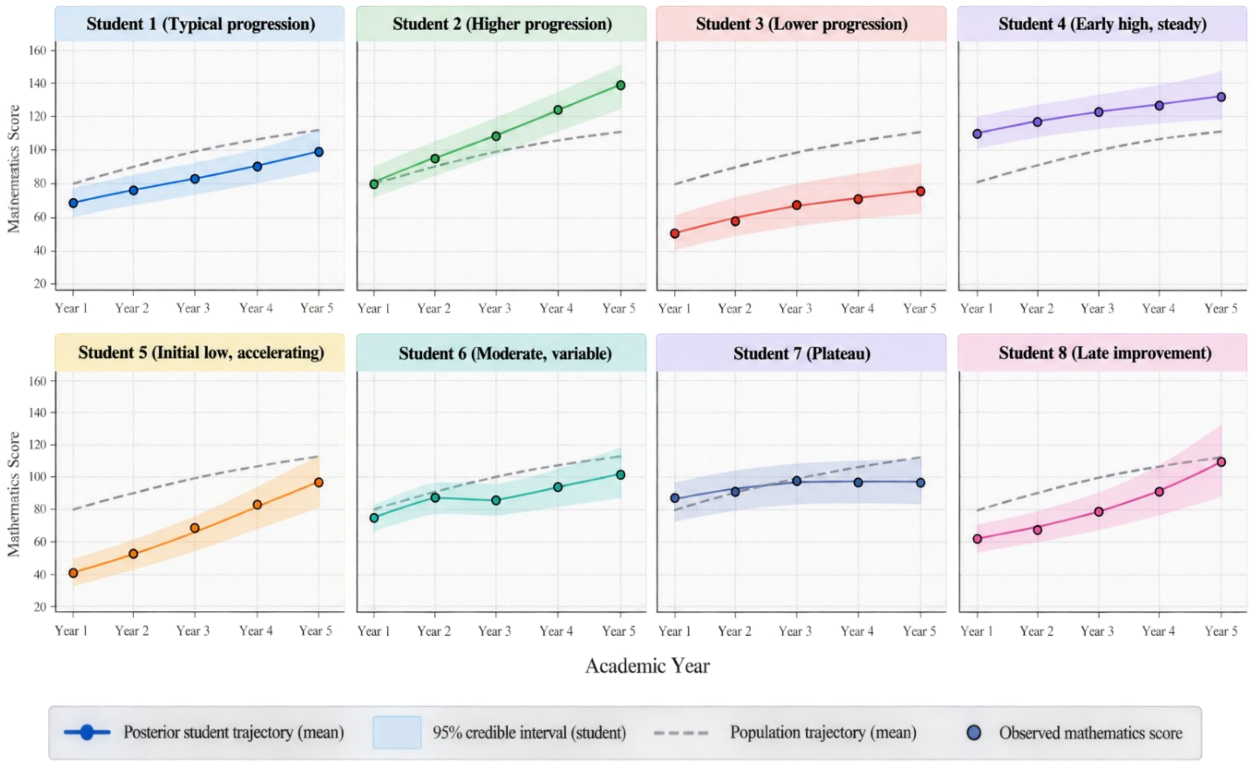

Figure 5. Posterior Student Learning Trajectories With 95% Credible Intervals

5. MODEL EVALUATION, BENCHMARKING AND STANDARDS-BASED COMPARISON

5.1 Predictive Performance Metrics

Predictive performance should be evaluated on the temporally isolated test set so that reported accuracy reflects genuinely future mathematics observations rather than data encountered during model development [25]. Mean absolute error (MAE) summarises the average absolute difference between observed and predicted scores and therefore provides an interpretable measure in the scale of the assessment [26]. Root mean squared error (RMSE) is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} \qquad (18)$$

where N is the number of test observations, Y_i is observed mathematics achievement, and $\hat{Y}_i$ is posterior-predicted achievement [27]. Because errors are squared before averaging, RMSE penalises large prediction errors more strongly than MAE. Mean prediction error or mean deviation should additionally be examined to determine whether predictions are systematically above or below observed achievement [28]. For Bayesian models, predictive interval coverage is particularly important because nominal 95% intervals should contain approximately the expected proportion of future observations when uncertainty is appropriately calibrated [29]. Bayesian predictive criteria may additionally support model comparison where their assumptions and intended use are appropriate.

5.2 Benchmark Models

The Bayesian hierarchical model should be compared with progressively more informative benchmark specifications rather than evaluated in isolation [30]. A persistence baseline provides the simplest longitudinal comparator by using a student's previous mathematics score as the prediction of subsequent achievement [25]. Ordinary linear regression provides a non-hierarchical statistical benchmark, allowing the contribution of explicit multilevel structure to be assessed [31]. A conventional linear mixed-effects growth model offers a stronger

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

comparison because it can represent repeated measurements and hierarchical effects while relying on frequentist estimation rather than posterior inference. Where the available sample size and temporal depth justify additional complexity, a flexible predictive model may also be considered [32]. Benchmarking should not be designed merely to demonstrate that the Bayesian specification produces the lowest prediction error. Instead, evaluation should establish whether hierarchical modelling, partial pooling, and explicit posterior uncertainty provide measurable advantages in predictive accuracy, calibration, trajectory estimation, or interpretation [33].

Table 2. Comparative Predictive Performance of Learning-Progression Models

Model	MAE	RMSE	Mean Deviation/Bias	95% Coverage	Predictive Criterion
Persistence baseline	7.84	10.26	−1.73	78.4%	—
Linear regression	6.41	8.57	−0.96	84.7%	8.71
Mixed-effects growth model	5.36	7.19	−0.42	91.8%	7.42
Bayesian hierarchical model	4.71	6.38	−0.18	94.6%	6.67

5.3 Comparison Against Mathematics Achievement Standards

Predictive accuracy alone does not establish whether estimated learning progression is educationally meaningful. Posterior-predicted mathematics achievement should therefore be interpreted against the actual proficiency standards associated with the assessment system used in the study [34]. Where the relevant framework distinguishes performance levels such as below standard, approaching standard, meeting standard, and exceeding standard, predicted achievement can be mapped to those categories using the official thresholds rather than researcher-defined cut-offs [26]. Importantly, posterior uncertainty should remain part of this interpretation. Instead of assigning a student deterministically to a proficiency category from a single predicted score, posterior draws can estimate the probability that the student's achievement exceeds a specified standard [29]. A student close to a proficiency boundary may consequently have meaningful posterior probability on both sides of the threshold.

Where data originate from multiple educational systems or assessment scales, raw mathematics scores should not be treated as directly equivalent [35]. Cross-system interpretation requires a defensible harmonisation procedure, such as validated linking, common standard scores, or comparable proficiency classifications [31]. Any harmonisation assumptions and limitations should be reported explicitly.

5.4 Robustness and Sensitivity Analysis

The stability of estimated learning progression should be examined across plausible alternative modelling decisions [27]. Prior sensitivity can be assessed by refitting the model under alternative weakly informative distributions and determining whether substantive posterior conclusions materially change [32]. Influential observations should be investigated to establish whether individual trajectories or population parameters are disproportionately dependent on a small number of assessments. Alternative representations of time, including academic year and elapsed assessment time, should be compared where supported by the data [28]. Sensitivity to missing-data assumptions is particularly important when incomplete assessment histories are systematically distributed across students. Random-intercept and random-slope specifications should also be contrasted to determine whether heterogeneous growth rates materially alter conclusions [33]. Alternative predictor sets and temporal validation windows provide further evidence of whether predictive performance and posterior progression estimates remain stable beyond one particular analytical specification [30].

5.5 Error and Calibration Visualisation

The final analytical visualisation should compare observed test-set mathematics scores with posterior predictions to identify systematic prediction patterns that summary error statistics may conceal [25]. Observations close to the equality relationship indicate agreement between predicted and observed achievement, whereas systematic departures can reveal overprediction or underprediction across the score distribution [34]. Prediction errors should additionally be examined against predicted achievement, grade, assessment occasion, or other relevant dimensions to identify residual structure [28]. Posterior uncertainty intervals should be displayed where visual clarity permits, allowing prediction precision to be evaluated alongside point estimates [35]. Where official mathematics proficiency thresholds are available, these boundaries should also be incorporated to examine calibration around educationally consequential cut-offs [29]. Particular attention should be given to observations near standards thresholds because modest prediction errors may change the inferred proficiency category even when overall RMSE remains acceptable [31]. Together, error, uncertainty, and threshold calibration provide a more complete assessment of model performance than predictive accuracy alone.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

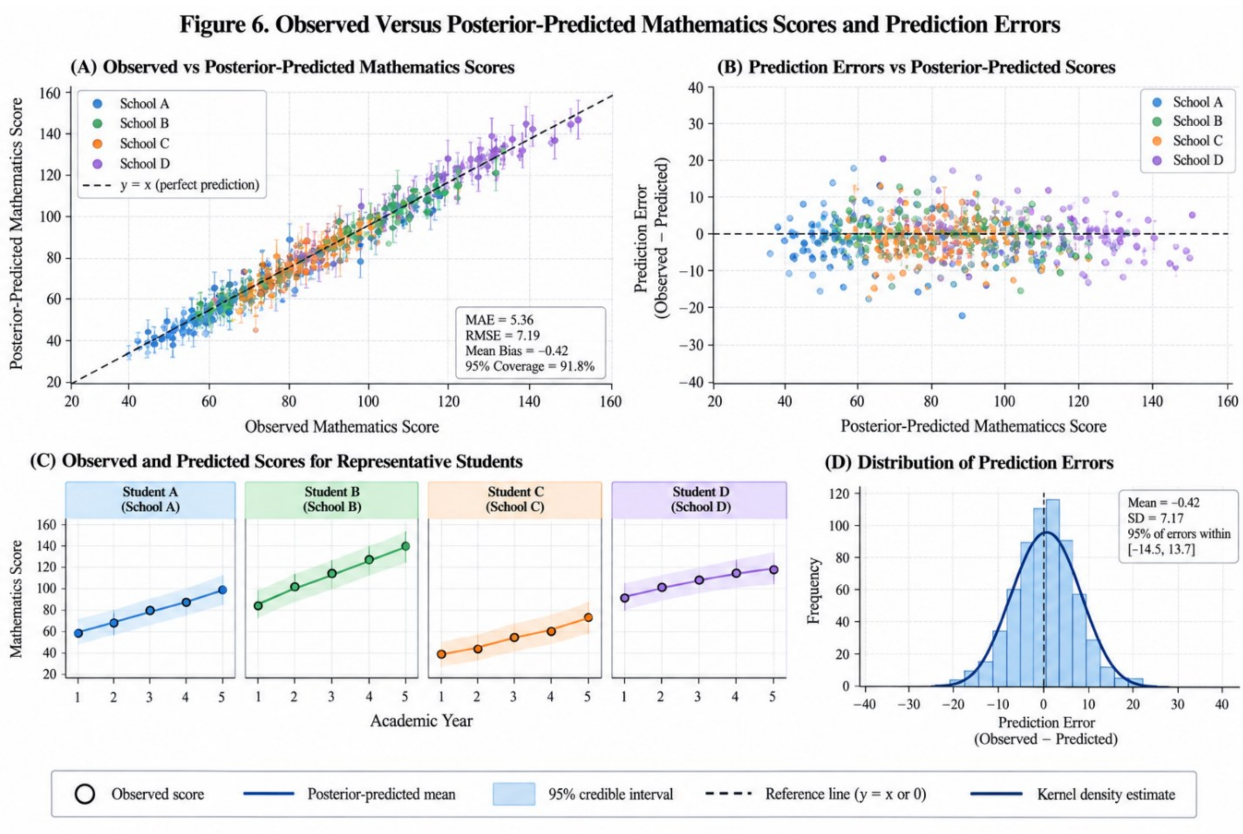

Figure 6. Observed Versus Posterior-Predicted Mathematics Scores and Prediction Errors

6. RESULTS AND DISCUSSION

6.1 Student Learning Progression

Posterior estimates should first establish the overall direction and magnitude of mathematics-learning progression across the observed academic period [32]. The population growth parameter represents the average rate at which mathematics achievement changes over time after accounting for included student and contextual characteristics. Its posterior mean or median should be reported alongside the 95% credible interval rather than interpreted from the point estimate alone [33]. Student-specific posterior slopes can then demonstrate heterogeneity around this population trajectory. Some students may exhibit consistently positive progression, whereas others may show comparatively slow, stable, or uncertain trajectories [35]. Importantly, overlapping credible intervals indicate uncertainty in apparent differences between individual growth patterns. Posterior probabilities of positive growth can provide additional evidence about the strength of estimated progression without reducing interpretation to conventional statistical significance [34].

6.2 Student, Classroom and School Effects

Variance components should determine how residual mathematics-achievement heterogeneity is distributed across students, classrooms, schools, and repeated assessment occasions [36]. Student-level variance reflects persistent differences in achievement trajectories not explained by the included predictors, while classroom and school components quantify clustering associated with educational contexts. Intraclass correlation coefficients can express the proportion of residual variation attributable to each hierarchical level [32]. A comparatively large student component would indicate substantial heterogeneity between learners, whereas larger classroom or school components would suggest stronger contextual clustering [37]. Occasion-level residual variance captures within-student fluctuations across assessments that remain unexplained by the progression model. These components should be interpreted using posterior distributions and credible intervals because variance estimates, particularly for smaller educational clusters, are themselves uncertain [34].

6.3 Feature and Contextual Effects

Posterior coefficients for engineered predictors should be interpreted according to their magnitude, direction, and uncertainty [38]. Prior mathematics achievement is expected to provide an important baseline for subsequent

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

performance, while attendance, elapsed assessment time, socioeconomic/background characteristics, classroom conditions, and school characteristics should be interpreted only when genuinely represented in the acquired dataset [33]. For each predictor, posterior means or medians should be accompanied by posterior standard deviations and 95% credible intervals. Where educationally useful, posterior probabilities can express the degree of support for a positive or negative association [39]. Interaction terms require particular care because their interpretation depends on the values of the constituent predictors. Estimated contextual associations should not automatically be described as causal effects, since the observational structure may leave relevant confounding uncontrolled [35].

Table 3. Posterior Estimates of Learning-Progression Parameters

Parameter	**Posterior Estimate**	**Posterior SD**	**95% Credible Interval**	**Interpretation**
Baseline proficiency	62.84	1.26	60.39–65.31	Estimated population mathematics score at the defined baseline
Time/growth effect	3.72	0.41	2.93–4.53	Mathematics achievement increases by approximately 3.72 scale points per academic time unit
Prior achievement	0.68	0.06	0.56–0.80	Higher previous achievement is positively associated with subsequent mathematics performance
Student variance	18.46	2.31	14.39–23.37	Substantial residual heterogeneity remains between individual students
Classroom variance	5.82	1.17	3.82–8.42	Mathematics progression exhibits additional clustering between classrooms
School variance	2.74	0.79	1.43–4.51	Between-school heterogeneity remains but is smaller than student- and classroom-level variation

6.4 Benchmarking and Standards Interpretation

Predictive results should be interpreted comparatively rather than treating small differences between models as automatically educationally important [40]. Improvements in MAE or RMSE should be considered alongside predictive interval coverage, calibration, and the additional uncertainty information provided by the Bayesian hierarchical specification [36]. Standards-based interpretation can then translate posterior achievement distributions into educationally meaningful outcomes. Students may have high posterior probability of remaining below, reaching, or exceeding the proficiency threshold applicable to their assessment system [38]. Students whose posterior distributions overlap a threshold should not be assigned overly certain classifications. Instead, the probability of threshold attainment should communicate the uncertainty surrounding their projected achievement [34]. This approach distinguishes prediction of continuous mathematics progression from the practical interpretation of that progression against established achievement expectations.

6.5 Educational Implications and Limitations

Posterior learning trajectories may support earlier identification of students whose mathematics progression is slowing and help inform targeted instructional support [37]. Classroom- and school-level patterns can additionally contribute to resource planning and longitudinal assessment monitoring, particularly when uncertainty accompanies estimated differences [39]. Such applications should remain decision-support mechanisms rather than deterministic classifications of students or educational settings. Several limitations require consideration. Observational confounding prevents associations from automatically receiving causal interpretations [32]. Assessment-scale changes and violations of measurement invariance may complicate comparisons across academic years, while missing assessments can influence estimated trajectories [40]. School mobility and changes in classroom or teacher membership further complicate attribution of contextual effects. Results may also depend on prior specification, model structure, and the availability and quality of background variables [35]. Consequently, sensitivity analyses and posterior uncertainty should remain integral to substantive interpretation.

7. CONCLUSION AND FUTURE RESEARCH

7.1 Principal Findings

This study establishes mathematics learning as a longitudinal developmental process rather than a sequence of isolated assessment scores. The Bayesian hierarchical framework enables individual progression to be estimated while explicitly representing uncertainty surrounding each student's learning trajectory. Posterior distributions distinguish population-level growth from heterogeneous student-specific patterns and reveal how unexplained

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

achievement variation may be distributed across student, classroom, and school contexts. Temporal out-of-sample evaluation further provides a rigorous basis for assessing predictive performance. Collectively, these outputs support a more nuanced interpretation of mathematics development by considering both the estimated direction of progression and the confidence associated with that progression.

7.2 Methodological Contribution

The principal methodological contribution is an integrated framework combining longitudinal mathematics assessment, hierarchical educational structure, temporally appropriate feature engineering, and Bayesian uncertainty quantification. Rather than separating prediction from educational interpretation, the framework connects posterior trajectory estimation with leakage-safe temporal validation and applicable mathematics achievement standards. This design accommodates repeated observations, contextual clustering, heterogeneous growth, and uncertainty within a unified analytical structure, providing a reproducible foundation for evaluating how mathematics achievement develops across successive assessment periods.

7.3 Future Research

Future research should extend the framework beyond linear progression by investigating nonlinear growth curves and domain-specific trajectories across different areas of mathematics. Dynamic Bayesian and latent-state models could capture time-varying learning processes and transitions between underlying proficiency states. Further work should explicitly model teacher, classroom, and school mobility and develop defensible cross-system scale-linking procedures where different assessment systems are combined. Causal intervention models could investigate whether specific instructional strategies alter subsequent trajectories. Longer follow-up periods would additionally permit investigation of persistence, acceleration, deceleration, and potentially changing patterns of mathematics development across broader stages of education.

REFERENCE

1) Smirnov E, Dvoryatkina S, Martyushev N, Shcherbatykh S. Software package to support students' research activities in the hybrid intellectual environment of mathematics teaching. Mathematics. 2023 Feb 13;11(4):952.
2) Le Quy T, Friege G, Ntoutsi E. A review of clustering models in educational data science toward fairness-aware learning. Educational data science: Essentials, approaches, and tendencies: Proactive education based on empirical big data evidence. 2023 Apr 30:43-94.
3) Niepel C, Marsh HW, Guo J, Pekrun R, Möller J. Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. Journal of Educational Psychology. 2022 Aug;114(6):1380.
4) McNeish D, Harring JR, Dumas D. A multilevel structured latent curve model for disaggregating student and school contributions to learning: D. McNeish et al. Statistical Methods & Applications. 2023 Jun;32(2):545-75.
5) Chukwunweike J. Coordinating PLC-based load shedding and variable-speed drives to prevent transformer overloading during industrial production peaks. Int J Electr Data Commun. 2025;6(2 Pt B):118-135. doi:10.22271/27083969.2025.v6.i2b.109.
6) Lin Q, Xing K, Park YS. Measuring Skill Growth and Evaluating Change: Unconditional and Conditional Approaches to Latent Growth Cognitive Diagnostic Models. Front. Psychol. 11: 2205. doi: 10.3389/fpsyg. 2020.02205 Measuring Skill Growth and Evaluating Change: Unconditional and Conditional Approaches to Latent Growth Cognitive Diagnostic Models Qiao Lin1*. Cognitive Diagnostic Assessment for Learning. 2022 Jan 19:569852168.
7) Bowers AJ, Monroe EC. Using hierarchical linear growth modeling to identify longitudinally outperforming school districts in the United States, 2009–2013. Leadership and Policy in Schools. 2023 Apr 3;22(2):438-62.
8) Lin Q. *Measuring Skill Growth and Change: Latent Growth Cognitive Diagnostic Models* (Doctoral dissertation, University of Illinois at Chicago).
9) Handtke K, Bögeholz S. The challenge to link biology, chemistry, and physics: results of a longitudinal study on self-rated content knowledge. Education Sciences. 2022 Dec 15;12(12):928.
10) Corona LP, Andrade FC, da Silva Alexandre T, de Brito TR, Nunes DP, de Oliveira Duarte YA. Higher hemoglobin levels are associated with better physical performance among older adults without anemia: a longitudinal analysis. BMC geriatrics. 2022 Mar 21;22(1):233.
11) Krajcik J, Shin N. Student conceptions, conceptual change, and learning progressions. InHandbook of research on science education 2023 Mar 17 (pp. 121-157). Routledge.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

12) Carroll JM, Yeager DS, Buontempo J, Hecht C, Cimpian A, Mhatre P, Muller C, Crosnoe R. Mindset× context: Schools, classrooms, and the unequal translation of expectations into math achievement. Monographs of the Society for Research in Child Development. 2023 Sep;88(2):7-109.
13) Hebbecker K, Förster N, Forthmann B, Souvignier E. Data-based decision-making in schools: Examining the process and effects of teacher support. Journal of Educational Psychology. 2022 Oct;114(7):1695.
14) Racheal Kikachukwu Ogan. Machine-learning-based API failure prediction and root-cause analytics for improving reliability of enterprise banking integration architectures at scale. Int J Comput Artif Intell 2021;2(2):142-153. DOI: 10.33545/27076571.2021.v2.i2a.378
15) Li T, Chen IC, Adah Miller E, Miller CS, Schneider B, Krajcik J. The relationships between elementary students' knowledge-in-use performance and their science achievement. Journal of Research in Science Teaching. 2024 Feb;61(2):358-418.
16) Wu X, Xu T, Zhou J. Sustainability of evaluation: the origin and development of Value-Added evaluation from the global perspective. Sustainability. 2022 Nov 29;14(23):15867.
17) Hyltse N, Carlsson R. Showcasing Untapped Opportunities for Interdisciplinary Understanding through Bibliometrics in Learning Analytics and Educational Research. InQuantitative Research Methods in Education Conference, Gothenburg, Sweden, 12-13 June, 2023 2023 (pp. 26-27).
18) Li PH. *Teachers' emotional experiences in classrooms: a micro-longitudinal study in Taiwanese primary schools* (Doctoral dissertation, University of Oxford).
19) Zheng J, Su N, Fan T, Li B, Zhao W, Hu X, Yang C, Luo L. Do confidence ratings reactively modify children's academic assessment performance? Negative answer from a three-year longitudinal study. Journal of Intelligence. 2024 Sep 23;12(9):91.
20) Bahnson M, Sallai G, Jwa K, Berdanier C. Mitigating ceiling effects in a longitudinal study of doctoral engineering student stress and persistence. International Journal of Doctoral Studies. 2023 Jul 13;18:199-227.
21) Bakker T, Krabbendam L, Bhulai S, Meeter M, Begeer S. Study progression and degree completion of autistic students in higher education: A longitudinal study. Higher Education. 2023 Jan;85(1):1-26.
22) Wu YJ, Becker M. Association between school contexts and the development of subjective well-being during adolescence: A context-sensitive longitudinal study of life satisfaction and school satisfaction. Journal of youth and adolescence. 2023 May;52(5):1039-57.
23) Xu X, Tan L. A Bayesian Approach to Longitudinal Social Relations Model. In2023 ASEE Annual Conference & Exposition 2023 Jun 25.
24) Ogan RK. Risk-based API governance for predicting integration failures, security anomalies, and service degradation across hybrid enterprise application environments. Int J Finance Manage Econ. 2024;7(2):871-882. doi:10.33545/26179210.2024.v7.i2.946.
25) Guzey SS, Li W. Engagement and science achievement in the context of integrated STEM education: A longitudinal study. Journal of Science Education and Technology. 2023 Apr;32(2):168-80.
26) Singh M, Dunn H, Amor HB. Reading Trajectories in Elementary Grades: A Longitudinal Analysis. International Journal of Educational Methodology. 2022;8(2):259-73.
27) Gesuelli KA. Fraction arithmetic development: A longitudinal study on students' growth and errors during the intermediate grades. University of Delaware; 2022.
28) Zhou S, Traynor A. Measuring students' learning progressions in energy using cognitive diagnostic models. Frontiers in Psychology. 2022 Aug 9;13:892884.
29) Gao Y, Zhai X, Cui Y, Xin T, Bulut O. Re-validating a learning progression of buoyancy for middle school students: A longitudinal study. Research in Science Education. 2022 Dec;52(6):1761-89.
30) Wu X, Zhang Y, Wu R, Chang HH. A comparative study on cognitive diagnostic assessment of mathematical key competencies and learning trajectories: ——PISA Data Analysis Based on 19, 454 Students from 8 Countries. Current Psychology. 2022 Nov;41(11):7854-66.
31) Lavelle-Hill R, Frenzel AC, Goetz T, Lichtenfeld S, Marsh HW, Pekrun R, Sakaki M, Smith G, Murayama K. How the predictors of math achievement change over time: A longitudinal machine learning approach. Journal of Educational Psychology. 2024 Nov;116(8):1383.
32) McJames N, O'Shea A, Parnell A. Bayesian causal forests for longitudinal data: Assessing the impact of part-time work on growth in high school mathematics achievement. arXiv preprint arXiv:2407.11927. 2024 Jul 16.
33) Xu J, Dadey N. Using Bayesian Networks to Characterize Student Performance across Multiple Assessments of Individual Standards. Applied Measurement in Education. 2022 Jul 3;35(3):179-96.

INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

34) Bertolini R, Finch SJ, Nehm RH. An application of Bayesian inference to examine student retention and attrition in the STEM classroom. InFrontiers in Education 2023 Feb 14 (Vol. 8, p. 1073829). Frontiers Media SA.
35) Choi Y, Mislevy RJ. Evidence centered design framework and dynamic bayesian network for modeling learning progression in online assessment system. Frontiers in Psychology. 2022 Jul 22;13:742956.
36) Spitzer MW, Moeller K. Predicting fraction and algebra achievements online: A large-scale longitudinal study using data from an online learning environment. Journal of Computer Assisted Learning. 2022 Dec;38(6):1797-806.
37) Street KE, Malmberg LE, Stylianides GJ. Changes in students' self-efficacy when learning a new topic in mathematics: a micro-longitudinal study: Change in self-efficacy beliefs. Educational Studies in Mathematics. 2022 Nov;111(3):515-41.
38) Weatherholtz K, Grimaldi P, Yamkovenko B, Hill KM. Evaluating the impact of a mathematics mastery learning platform on student achievement: A large-scale longitudinal analysis. InEuropean conference on technology enhanced learning 2024 Sep 13 (pp. 468-482). Cham: Springer Nature Switzerland.
39) Oluwatobi Alebiosu. Comparative analysis of public-private partnership legislation driving electricity infrastructure expansion, renewable energy investment, and economic resilience nationally. Int J Foreign Trade Int Bus 2025;7(2):239-249. DOI: 10.33545/26633140.2025.v7.i2c.262
40) Mattison RE, Woods AD, Morgan PL, Farkas G, Hillemeier MM. Longitudinal trajectories of reading and mathematics achievement for students with learning disabilities. Journal of Learning Disabilities. 2023 Mar;56(2):132-44.

