Copyright © The Author(s)
IJIER: Indonesian Journal of Islamic Educational Review, Vol. 3, No.1, February 2026: 35-50
Available online at: https://ssed.or.id/journal/ijier
ISSN 3047-0706

Deeply Interpreting of Student Growth Percentile Model: Promoting the Application of Value-added Evaluation

Yaqin Fan
Institut North China Electric Power University, China
e-mail: fanyq127@126.com

Abstract: *The increasing number of independent variables and dependent variables, the correction of covariates and the cross-influence between variables all promote the evolution and development of value-added evaluation model. As one of the value-added evaluation models, the student growth percentile model is one of the most popular value-added evaluation models with its easy to understand core concept and relatively simple operation process. However, for primary and secondary school teachers, the model still has certain difficulties in practical application, and the technical requirements are beyond the public level. In order to promote the application and further development of the percentile model of student growth, this paper discusses it from the basic theory, application operation, limitation and development, so as to promote the theoretical understanding of the model and the operation process to be more specific.*
Keywords: *Midday prayer in congregation, forming morals*

Abstrak: *Meningkatnya jumlah variabel independen dan variabel dependen, koreksi kovariat, dan pengaruh silang antar variabel semuanya mendorong evolusi dan pengembangan model evaluasi nilai tambah. Sebagai salah satu model evaluasi nilai tambah, model persentil pertumbuhan siswa merupakan salah satu model evaluasi nilai tambah yang paling populer dengan konsep inti yang mudah dipahami dan proses operasi yang relatif sederhana. Namun, bagi guru sekolah dasar dan menengah, model ini masih memiliki kesulitan tertentu dalam penerapan praktis, dan persyaratan teknisnya melampaui tingkat umum. Untuk mempromosikan penerapan dan pengembangan lebih lanjut dari model persentil pertumbuhan siswa, makalah ini membahasnya dari teori dasar, operasi aplikasi, keterbatasan dan pengembangan, sehingga dapat meningkatkan pemahaman teoritis tentang model dan membuat proses operasinya menjadi lebih spesifik*
Kata Kunci: *Shalat Dzuhur Berjamaah, Membentuk Akhlak*

Introduction

Since the seminal "Coleman Report" highlighted educational opportunity disparities, value-added assessment has emerged as a crucial framework for measuring educational effectiveness (Coleman et al., 1966). The development accelerated when Tennessee pioneered the first comprehensive value-added assessment system (TVAAS) for teacher evaluation in 1992, establishing a new paradigm for educational accountability (Sanders & Horn, 1994). By the late 1990s, these concepts began influencing educational assessment practices in China (Zhang,

1998). A significant policy milestone was reached in October 2020 when China's Central Committee and State Council issued the "Overall Plan for Deepening the Reform of Educational Evaluation in the New Era," explicitly mandating the exploration of value-added assessment approaches (Central Committee of the Communist Party of China & State Council, 2020). This policy impetus has catalyzed substantial research activity, though China's development in this domain remains emergent compared to international counterparts. Current scholarly work predominantly examines value-added assessment conceptual foundations, international model adaptation, methodological refinements, and practical implementation strategies (Harris, 2020; Meyer & Dokumaci, 2023b; Zhu et al., 2023). Regionally, educational systems in Shanghai (Shanghai Municipal Education Commission, 2022), Jiangsu (Jiangsu Provincial Department of Education, 2023), and Shenyang (Shenyang Education Bureau, 2023) have initiated pilot programs, generating valuable implementation prototypes for value-added assessment in the Chinese context.

1.1 Research Problem

The efficacy of value-added assessment fundamentally depends on appropriate model selection. Contemporary value-added models embody distinct educational philosophies and employ diverse statistical methodologies, each with specific data requirements and interpretive frameworks. As McCaffrey and Castellano (2021) emphasize, "no single model dominates educational value-added assessment; selection requires careful consideration of measurement objectives and contextual constraints". The methodological evolution has progressed from rudimentary single-factor constructs to sophisticated multi-dimensional frameworks capable of analyzing cross-disciplinary growth patterns. This progression necessitates balancing operational feasibility against analytical precision. Among prevalent approaches, the Student Growth Percentile (SGP) and Hierarchical Linear Model (HLM) represent prominent alternatives. SGP enjoys widespread adoption in American educational systems due to its procedural transparency and intuitive interpretability, while HLM offers enhanced analytical precision for investigating multifaceted growth determinants despite greater implementation complexity (Raudenbush & Bryk, 2022). However, Betebenner (2021) notes that "technical barriers and conceptual unfamiliarity significantly impede SGP adoption in educational systems without established growth assessment traditions". The central challenge confronting Chinese researchers and practitioners involves overcoming substantive obstacles in theoretical comprehension, operational protocols, and technical implementation to enable effective SGP utilization.

1.2 Research Significance

Advancing SGP methodology holds substantial implications for China's educational evaluation reform agenda. Comprehensive investigation of SGP's theoretical foundations, implementation mechanics, and limitations can equip educational professionals with essential competencies for growth-oriented assessment. As Darling-Hammond (2022) observes, "professional understanding of growth models represents a critical component of modern educational assessment literacy". Moreover, this research supports China's broader transition from exclusive reliance on achievement metrics toward comprehensive growth monitoring, thereby facilitating evidence-based educational governance. Lockwood and McCaffrey (2020) demonstrate that "appropriately implemented growth models provide valuable insights for instructional improvement and resource allocation". By systematically examining SGP's theoretical underpinnings, practical applications, and developmental

trajectories, this study aims to enhance practitioner expertise and promote sustainable implementation of student growth assessment in Chinese educational settings.

Methods

This study employs a comprehensive literature review and technical demonstration approach to examine the Student Growth Percentile (SGP) model. The methodology encompasses three primary components: theoretical analysis, practical demonstration using the SGPdata dataset, and critical evaluation of the model's limitations and future directions.

Theoretical Framework Analysis

The theoretical examination synthesizes foundational literature on SGP development, focusing on its conceptual origins from Betebenner's work at the National Center for the Improvement of Educational Assessment (NCIEA). The analysis traces the model's theoretical evolution from conventional growth measurement approaches to its current normative framework for interpreting student academic progress relative to peers with similar prior achievement histories.

Statistical Methods Review

A systematic review of SGP estimation methods was conducted, analyzing five primary approaches: Standard Estimator, Ordinary Least Squares (OLS) Regression Estimator, Quantile Regression Estimator, SIMEX (Simulation-Extrapolation) Correction, and SNP-MIRT (Semi-Nonparametric Multidimensional Item Response Theory) Estimation. Each method's advantages, disadvantages, and application conditions were examined through existing empirical studies and simulation research.

Practical Application Demonstration

To illustrate the operational procedures of SGP calculation, this study utilized the SGPdata dataset, a sample dataset included in R for SGP analysis, developed collaboratively by Betebenner et al. (2024). The demonstration followed a systematic process:

Software Environment Setup: Installation and configuration of R version 4.3.1 from the Comprehensive R Archive Network (R Foundation, 2024)

Data Acquisition: Extraction of the sgpData example dataset from the SGPdata package

Data Configuration: Filtering and preparation of longitudinal achievement data

SGP Analysis Execution: Implementation using the SGP package in R with quantile regression methodology

Results Interpretation: Extraction and analysis of student growth percentiles

The practical demonstration specifically analyzed student growth using Grade 3 scores from 2021 as prior scores and Grade 4 scores from 2022 as current scores, following standardized protocols for SGP calculation.

Critical Evaluation Framework

The limitations and future development directions of the SGP model were evaluated through synthesis of existing research findings, focusing on seven primary constraint areas: sample size requirements, longitudinal data specifications, population variability, model reliability and validity, assessment content consistency, explanatory power, and result reusability. This evaluation informed recommendations for future research directions and practical application strategies.

Result and Discussion

Basic Theory of the Student Growth Percentile Model

Origin, Development, and Theoretical Basis of SGP

The Student Growth Percentile model provides a normative framework for calculating and interpreting student growth (Castellano & Ho, 2013a). It was developed by Dr. Damian Betebenner of the National Center for the Improvement of Educational Assessment (NCIEA) in the United States. Currently, SGP is widely applied in the educational management and teaching practices of schools, universities, and educational institutions. It is extensively used in K-12 assessments in nearly 20 states across the US, and the SGP package in R has become a popular tool for analyzing large-scale assessment data.

SGP aims to describe a student's current status relative to their own historical performance. Specifically, it answers the question: what is a student's percentile rank compared to students with similar prior achievement histories? It describes the relative position of a student's current score among students who started with similar prior scores. This position within the group of students with similar prior achievement is expressed as a percentile. For example, a student with an SGP of 80 has demonstrated growth exceeding that of 80% of their academic peers with comparable prior achievement during a specific period.

The theoretical foundation of SGP lies in a distinct understanding of the concept of "growth." Conventionally, individual growth is often viewed as change relative to one's own prior level, a progression based on the individual's own baseline. Betebenner (2009), however, posits this as a misconception, arguing that such a vertical measurement standard based solely on the individual obscures the understanding of "growth". He illustrates this by posing a question: When a parent asks, "How much has my child improved?", is it more meaningful to report the point change or the change in rank? Betebenner argues the latter is more meaningful. The student growth percentile describes the progress of an individual student relative to a peer group with similar prior achievement histories; it indicates a student's current performance compared to the performance of other students who followed similar achievement paths in the past (Betebenner, 2009).

Basic Principles of SGP

The fundamental concept of SGP is that a student's growth percentile represents their relative position, or rank, in terms of score growth within a group of students who started with the same prior scores. Figure 1, referencing Betebenner's research, provides a more concrete and visual understanding of the SGP concept. Figure 1(a) shows the distribution of student scores in 2005 and 2006 and the frequency at each score point. Figure 1(b) represents the set of students who scored 600 in 2005. Figure 1(c) depicts the score distribution in 2006 for the specific cohort of students who scored 600 in 2005 (from Figure 1b). Figure 1(d) shows the probability distribution for a student scoring 650 in 2006 within the context of the 2006 score distribution of their peers who scored 600 in 2005.

Figure 1: Schematic diagram of student score distributions in 2005 and 2006 and SGP calculation (Betebenner, 2009)

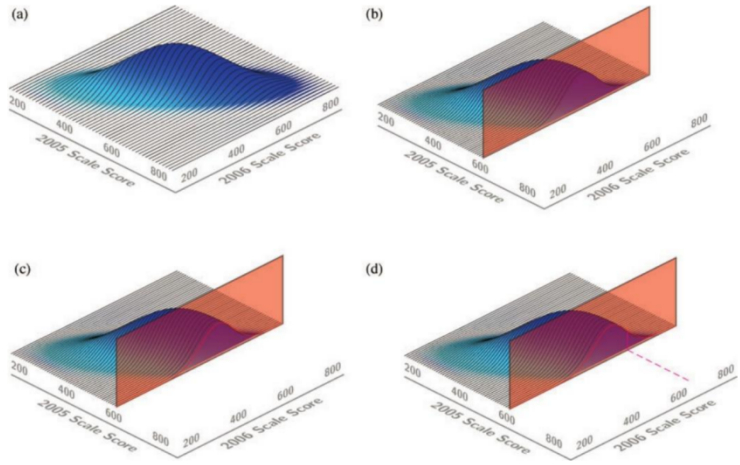


Figure 1: Schematic diagram of student score distributions in 2005 and 2006 and SGP calculation (Betebenner, 2009)

Estimation Methods for SGP

Based on the SGP concept, primary estimation methods include the Standard Estimator, Ordinary Least Squares (OLS) Regression Estimator, Quantile Regression Estimator, SIMEX Correction, and SNP-MIRT Estimation. Each method aims to enhance the accuracy and reliability of the SGP model.

① Standard Estimator. The basic idea is that, given a sufficiently large sample size, students with identical prior scores will form a distribution of current scores. The student's growth percentile is then calculated as the cumulative distribution probability of their current score within this reference distribution (McCaffrey et al., 2015). This method requires a large sample size to ensure that each prior score can form a corresponding distribution for percentile calculation.

② Ordinary Least Squares (OLS) Regression Estimator. Unlike the Standard Estimator, which relies on the cumulative distribution formed by identical prior scores, the OLS regression estimator fits a linear regression to students with similar prior scores. It aims to minimize the sum of squared residuals between the estimated and actual values, thereby improving the precision of the SGP estimate (Zhou & Liu, 2020).

③ Quantile Regression Estimator. Quantile regression does not require assumptions about the data distribution. Instead, it directly models the relationship between independent variables and specific quantiles of the dependent variable by minimizing the sum of absolute distances between data points and the fitted line, rather than the sum of squared differences used in traditional OLS regression (Chen & Ding, 2008). Compared to OLS, quantile regression (specifically using Least Absolute Deviations, LAD, for the median) is often more robust to outliers and extreme values.

④ SIMEX (Simulation-Extrapolation) Estimation. The SIMEX algorithm is a simulation-based inference method for measurement error models where the measurement error variance is known or can be reasonably estimated. This method involves intentionally adding extra measurement error with known increments to the data, calculating estimates from this contaminated data, establishing a trend between these estimates and the variance of the added error, and then extrapolating this trend back to the case of no measurement error. Simulation studies have shown that this method produces estimators that are nearly asymptotically unbiased and efficient in standard and non-standard settings (Cook & Stefanski, 1994). Shang et al. (2015) found that SIMEX effectively reduces bias in both individual and aggregate-level SGP

estimates, including for students with very high or very low prior achievement, thus improving SGP estimation accuracy. McCaffrey et al. (2015) demonstrated, using practical examples, why measurement error affects SGP estimates at both individual and aggregate levels and compared the accuracy of several estimation methods, including SIMEX. Their comparisons highlighted a common bias-variance trade-off: estimators that reduce bias relative to the standard SGP estimator tend to increase variance, and vice versa. Even the most accurate methods for estimating individual student SGPs in their study had substantial errors in practical settings, averaging around 19 percentile points. While bias-reducing estimators might be adequate for aggregate-level SGP inferences, none proved to be the optimal estimator fulfilling the dual objectives of inference at both the individual student and aggregate levels.

⑤ SNP-MIRT Estimation. Quantile regression estimates are often subject to measurement error, leading to bias in SGP results (Shang, 2012). Monroe (2014) extended the unidimensional Item Response Theory (IRT) model (Woods & Lin, 2009) to a multidimensional IRT (MIRT) framework. This model directly accounts for measurement error and utilizes Bayesian inference to minimize estimation error, simultaneously considering the optimal parameter values given the measurement uncertainty. Simulation results indicated that the SNP-MIRT model performed well in terms of item parameter recovery and density estimation, outperforming percentile calculations based on quantile regression estimation (Monroe, 2014).

Table 1: Analysis of Advantages and Disadvantages of Several Estimation Methods

Estimation Method	Advantages	Disadvantages
Standard Estimator	Simple concept, easy to understand	Requires a large sample size
OLS Regression Estimator	Higher precision than Standard Estimator when sample size is insufficient	Results are less intuitive; estimation precision is relatively low
Quantile Regression Estimator	More robust than OLS regression	Estimation precision is relatively low
SIMEX Correction	Effectively reduces estimation bias; higher precision for extreme prior scores	Conceptually complex and difficult to understand; estimation error still persists
SNP-MIRT Estimation	Incorporates measurement error and item information; precision superior to quantile regression	Conceptually complex and difficult to understand

In summary, each estimation method has its own advantages, disadvantages, and conditions for application, as summarized in Table 1. The choice of estimation method, similar to the choice of value-added model itself, involves a trade-off between operational simplicity and result accuracy (Meyer & Dokumaci, 2023b). When the sample size is sufficiently large, the Standard Estimator can be directly chosen due to its straightforward procedure and relatively high accuracy, making other methods unnecessary. When the sample size is inadequate for the Standard Estimator, if

technical capacity allows and item information is sufficient, SNP-MIRT estimation can be chosen; if item information is missing, SIMEX correction can be considered; if technical capacity is limited, Quantile Regression or OLS Regression estimation can be alternatives.

Comparative studies indicate that under conditions of sufficient sample size (typically exceeding 5,000 observations), the Standard Estimation method offers a balance of operational simplicity and reasonable accuracy. When sample size is limited or measurement error is a significant concern, simulation studies suggest that the SNP-MIRT estimation, despite its complexity, can provide superior accuracy by directly modeling measurement error and leveraging Bayesian inference. For contexts with technical constraints, Quantile Regression Estimation presents a more robust alternative to Least Squares Regression, being less sensitive to outliers.

Basic Statistical Concept of SGP Based on Quantile Regression

Quantile regression is one of the most commonly used statistical techniques for SGP. The statistical model of SGP based on quantile regression indeed involves complex concepts, making it relatively difficult for frontline educators to understand and apply. To better promote the popularization and application of the SGP model, we will provide a more detailed interpretation. This section aims to demystify the method by dissecting and explaining the quantile regression approach, making its rules and principles clearer and more accessible to facilitate understanding and application by educational practitioners.

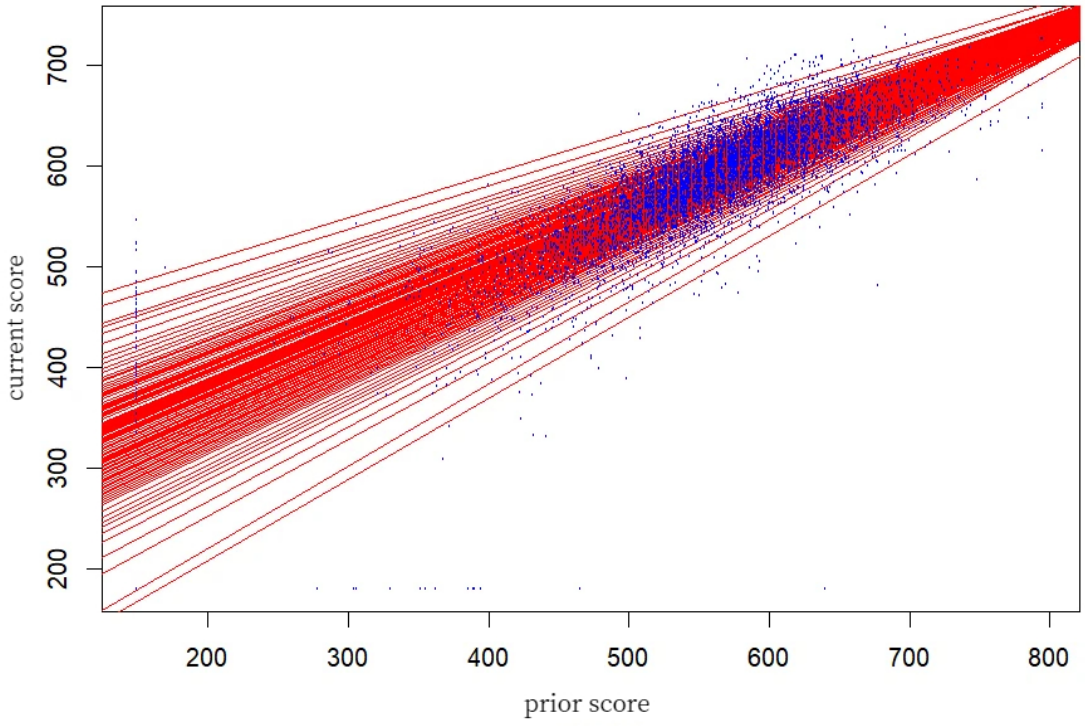


Figure 2: Scatter plot and quantile regression curves

Figure 2 is a scatter plot of prior scores and current scores from a specific dataset, fitted with 99 quantile regression lines corresponding to quantiles from 0.01 to 0.99 in increments of 0.01. Using the quantreg package in R, the model fit error is analyzed across these 99 different quantile levels. The quantile corresponding to the

fitted line that yields the smallest error for an individual student among these 99 quantile regression lines represents that student's growth percentile.

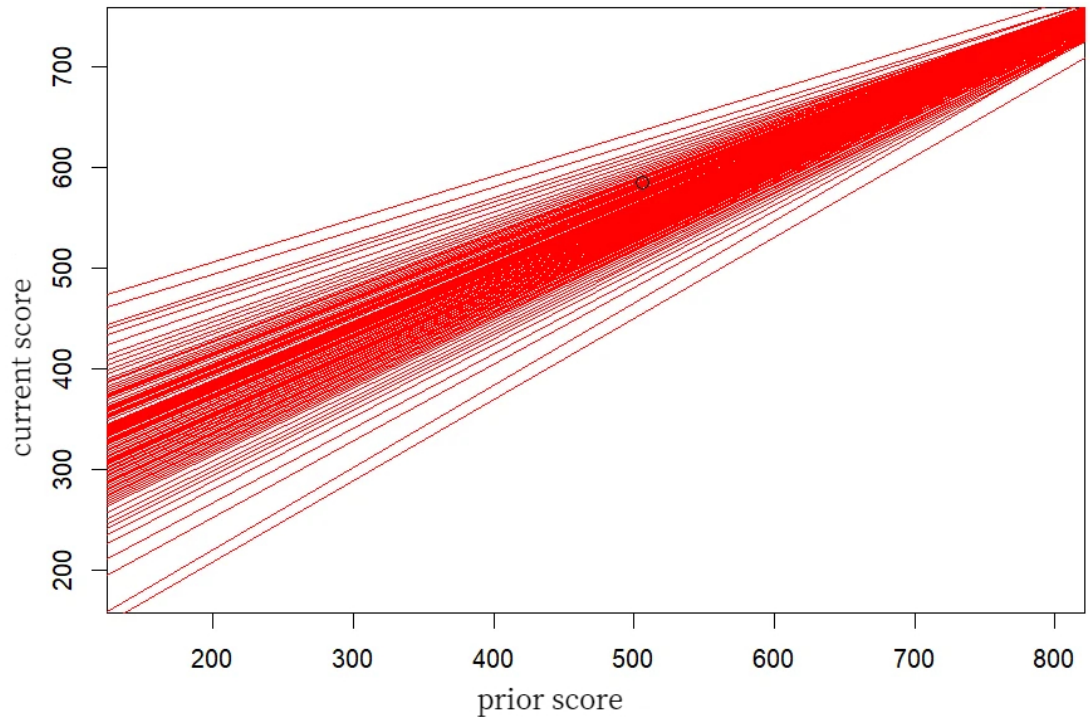


Figure 3: Point (506, 585) and the quantile regression lines

As shown in Figure 3, which illustrates the position of a student with a prior score of 506 and a current score of 585 relative to the quantile regression lines. Zooming in on this point reveals that the fitted line with the smallest error is located at a relatively high quantile. In fact, based on the exported errors at each quantile, the quantile with the minimum error for this point is 83.

Advantages of SGP

The SGP model offers several distinct advantages that contribute to its widespread adoption in educational assessment systems. First, the model's intuitive and understandable concept introduces the notion of "academic peers," providing a reference standard for assessing student growth that aligns more closely with common understandings of value-added concepts (Xie & Zhang, 2021). This conceptual clarity facilitates communication among educators, parents, and policymakers.

Second, the model ensures equal opportunity for growth by ranking subsequent scores into 1-99 percentile levels, which ensures that all individuals have the same conceptual space for demonstrating growth, potentially mitigating assessment unfairness issues arising from different starting points (Xie & Zhang, 2021). Third, SGP demonstrates relative operational simplicity as it does not require a vertical scale, thereby avoiding the constraints of vertical equating, which can significantly reduce operational complexity (Fu et al., 2019). Additionally, the model does not require the collection of student background information, further simplifying data collection procedures.

Fourth, the SGP model operates under less stringent statistical assumptions, as it does not require data to meet linearity or homoscedasticity assumptions, making it applicable to a wider range of data types and situations, offering greater flexibility in diverse educational contexts (Xie, 2019). Finally, the model demonstrates broad applicability, as SGP can be used not only for evaluating individual students but also for assessing student groups, as well as for teacher and school performance measurement (Fu et al., 2019).

Application and Operation of the Student Growth Percentile Model

This section uses the SGPdata dataset as an example to illustrate the specific operational procedures for calculating Student Growth Percentiles. The SGPdata is a sample dataset included in R for SGP analysis, designed to help users configure their own data for SGP analysis. It was developed collaboratively by Damian Betebenner and Adam R. Van Iwaarden from the National Center for the Improvement of Educational Assessment (NCIEA), and Ben Domingue from Stanford University (Betebenner et al., 2024).

Downloading and Installing R

First, download a suitable version of the R package from the open-source R language software repository CRAN (Comprehensive R Archive Network) (R Foundation, 2024). For instance, this study utilizes R version 4.3.1. Once the download and installation are complete, R is ready for use.

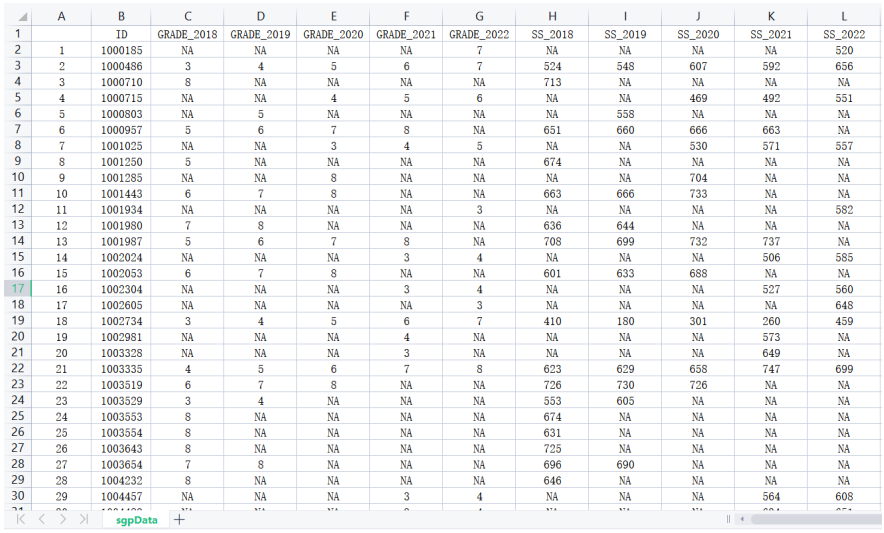
Obtaining the Example Dataset

Execute the following code to obtain the sgpData example dataset.

```
rsetwd("F:/sgpdata") # Set the working directory
getwd() # Verify the current working directory
require(SGPdata) # Load the SGPdata package
help(package="SGPdata") # View package documentation
write.csv(sgpData, file="F:/sgpdata/sgpData.csv") # Save the dataset
```

This code specifies the full path and filename for the output file. In this example, the data will be saved to the F:/sgpdata/ directory under the filename sgpData.csv. Consequently, the contents of sgpData are saved in CSV format at the specified location. The specific contents of the sgpData.csv file are shown in Figure 4, including 10 variables: GRADE_2018, GRADE_2019, GRADE_2020, GRADE_2021, GRADE_2022, SS_2018, SS_2019, SS_2020, SS_2021, SS_2022. The first five are grade-level variables, and the last five are scale score variables, corresponding sequentially.

Note that using the write.csv() function in R typically does not require installing an additional package, as it is a function provided in base R. write.csv() is part of R's basic functionality for saving data frames as CSV files.



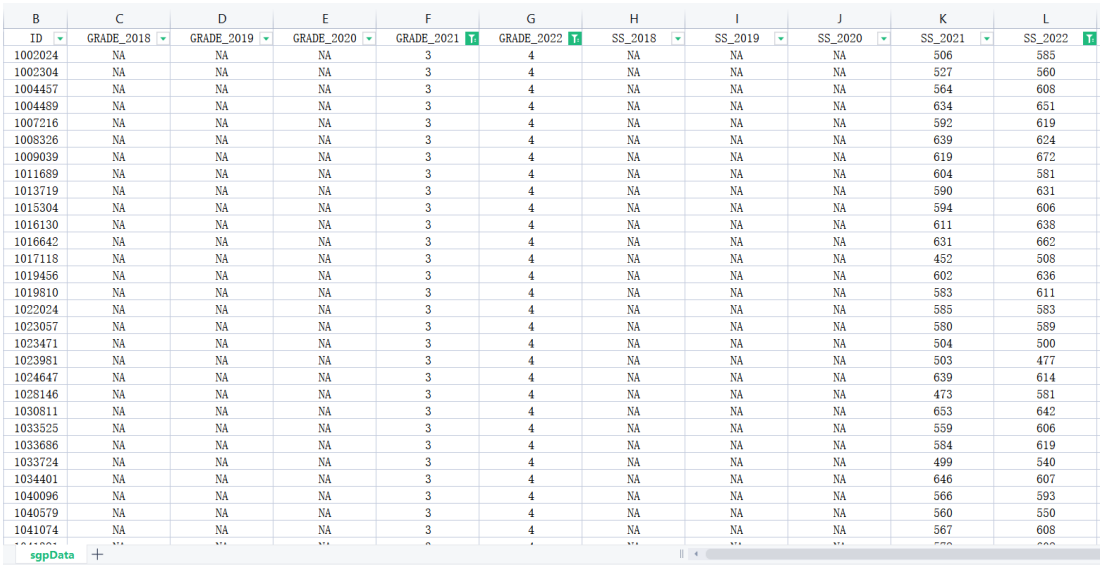
	A	B	C	D	E	F	G	H	I	J	K	L
1		ID	GRADE_2018	GRADE_2019	GRADE_2020	GRADE_2021	GRADE_2022	SS_2018	SS_2019	SS_2020	SS_2021	SS_2022
2	1	1000185	NA	NA	NA	NA	7	NA	NA	NA	NA	520
3	2	1000486	3	4	5	6	7	524	548	607	592	656
4	3	1000710	8	NA	NA	NA	NA	713	NA	NA	NA	NA
5	4	1000715	NA	NA	4	5	6	NA	NA	469	492	551
6	5	1000803	NA	5	NA	NA	NA	NA	558	NA	NA	NA
7	6	1000957	5	6	7	8	NA	651	660	666	663	NA
8	7	1001025	NA	NA	3	4	5	NA	NA	530	571	557
9	8	1001250	5	NA	NA	NA	NA	674	NA	NA	NA	NA
10	9	1001285	NA	NA	8	NA	NA	NA	NA	704	NA	NA
11	10	1001443	6	7	8	NA	NA	663	666	733	NA	NA
12	11	1001934	NA	NA	NA	NA	3	NA	NA	NA	NA	582
13	12	1001980	7	8	NA	NA	NA	636	644	NA	NA	NA
14	13	1001987	5	6	7	8	NA	708	699	732	737	NA
15	14	1002024	NA	NA	NA	3	4	NA	NA	NA	506	585
16	15	1002053	6	7	8	NA	NA	601	633	688	NA	NA
17	16	1002304	NA	NA	NA	3	4	NA	NA	NA	527	560
18	17	1002605	NA	NA	NA	NA	3	NA	NA	NA	NA	648
19	18	1002734	3	4	5	6	7	410	180	301	260	459
20	19	1002981	NA	NA	NA	4	NA	NA	NA	NA	573	NA
21	20	1003328	NA	NA	NA	3	NA	NA	NA	NA	649	NA
22	21	1003335	4	5	6	7	8	623	629	658	747	699
23	22	1003519	6	7	8	NA	NA	726	730	726	NA	NA
24	23	1003529	3	4	NA	NA	NA	553	605	NA	NA	NA
25	24	1003553	8	NA	NA	NA	NA	674	NA	NA	NA	NA
26	25	1003554	8	NA	NA	NA	NA	631	NA	NA	NA	NA
27	26	1003643	8	NA	NA	NA	NA	725	NA	NA	NA	NA
28	27	1003654	7	8	NA	NA	NA	696	690	NA	NA	NA
29	28	1004232	8	NA	NA	NA	NA	646	NA	NA	NA	NA
30	29	1004457	NA	NA	NA	3	4	NA	NA	NA	564	608

sgpData

Figure 4: Specific contents of the sgpData dataset

Data Configuration

After obtaining the sgpData file (shown in Figure 4), simple filtering and adjustments are necessary before proceeding. For example, to analyze SGPs using "Grade 3 scores from 2021 as the prior scores and Grade 4 scores from 2022 as the current scores," filter the data to include only records where GRADE_2021 is 3 and GRADE_2022 is 4, remove records with NA in SS_2021 or SS_2022, and potentially filter out other variables as needed. The result after filtering and adjustment is shown in Figure 5. Unlike Figure 4, Column A from the original data in Figure 4 needs to be deleted to align with the SGP package's code expectations.



B	C	D	E	F	G	H	I	J	K	L
ID	GRADE_2018	GRADE_2019	GRADE_2020	GRADE_2021	GRADE_2022	SS_2018	SS_2019	SS_2020	SS_2021	SS_2022
1002024	NA	NA	NA	3	4	NA	NA	NA	506	585
1002304	NA	NA	NA	3	4	NA	NA	NA	527	560
1004457	NA	NA	NA	3	4	NA	NA	NA	564	608
1004489	NA	NA	NA	3	4	NA	NA	NA	634	651
1007216	NA	NA	NA	3	4	NA	NA	NA	592	619
1008326	NA	NA	NA	3	4	NA	NA	NA	639	624
1009039	NA	NA	NA	3	4	NA	NA	NA	619	672
1011689	NA	NA	NA	3	4	NA	NA	NA	604	581
1013719	NA	NA	NA	3	4	NA	NA	NA	590	631
1015304	NA	NA	NA	3	4	NA	NA	NA	594	606
1016130	NA	NA	NA	3	4	NA	NA	NA	611	638
1016642	NA	NA	NA	3	4	NA	NA	NA	631	662
1017118	NA	NA	NA	3	4	NA	NA	NA	452	508
1019456	NA	NA	NA	3	4	NA	NA	NA	602	636
1019810	NA	NA	NA	3	4	NA	NA	NA	583	611
1022024	NA	NA	NA	3	4	NA	NA	NA	585	583
1023057	NA	NA	NA	3	4	NA	NA	NA	580	589
1023471	NA	NA	NA	3	4	NA	NA	NA	504	500
1023981	NA	NA	NA	3	4	NA	NA	NA	503	477
1024647	NA	NA	NA	3	4	NA	NA	NA	639	614
1028146	NA	NA	NA	3	4	NA	NA	NA	473	581
1030811	NA	NA	NA	3	4	NA	NA	NA	653	642
1033525	NA	NA	NA	3	4	NA	NA	NA	559	606
1033686	NA	NA	NA	3	4	NA	NA	NA	584	619
1033724	NA	NA	NA	3	4	NA	NA	NA	499	540
1034401	NA	NA	NA	3	4	NA	NA	NA	646	607
1040096	NA	NA	NA	3	4	NA	NA	NA	566	593
1040579	NA	NA	NA	3	4	NA	NA	NA	560	550
1041074	NA	NA	NA	3	4	NA	NA	NA	567	608

sgpData

Figure 5: Filtered sgpData dataset

SGP Analysis

Execute the following code to perform the SGP analysis. This involves installing and loading the relevant R packages, reading the source data, and finally conducting the SGP analysis.

```
rinstall.packages("SGP") # Install the SGP package
install.packages("readr") # Install the readr package
```

```
library(SGP) # Load the SGP package
library(readr) # Load the readr package
sgpdata <- read_csv("F:/sgpdata/sgpData.csv") # Read the CSV file

sgp <- studentGrowthPercentiles(
  panel.data = sgpdata,
  sgp.labels = list(my.year = 2023, my.subject = "Reading"),
  percentile.cuts = c(1, 35, 65, 99),
  grade.progression = c(3, 4)) # Perform SGP analysis
```

Viewing and Exporting Results

Execute the code View(sgp) to view detailed information about the sgp object in the RStudio panel. As shown in Figure 6, the sgp data object is a list containing 8 elements of different types. The Student Growth Percentile data we need is located within the SGPercentiles -> READING.2023 -> SGP element. Extracting the data from this element yields the SGP analysis results, i.e., the Student Growth Percentiles.

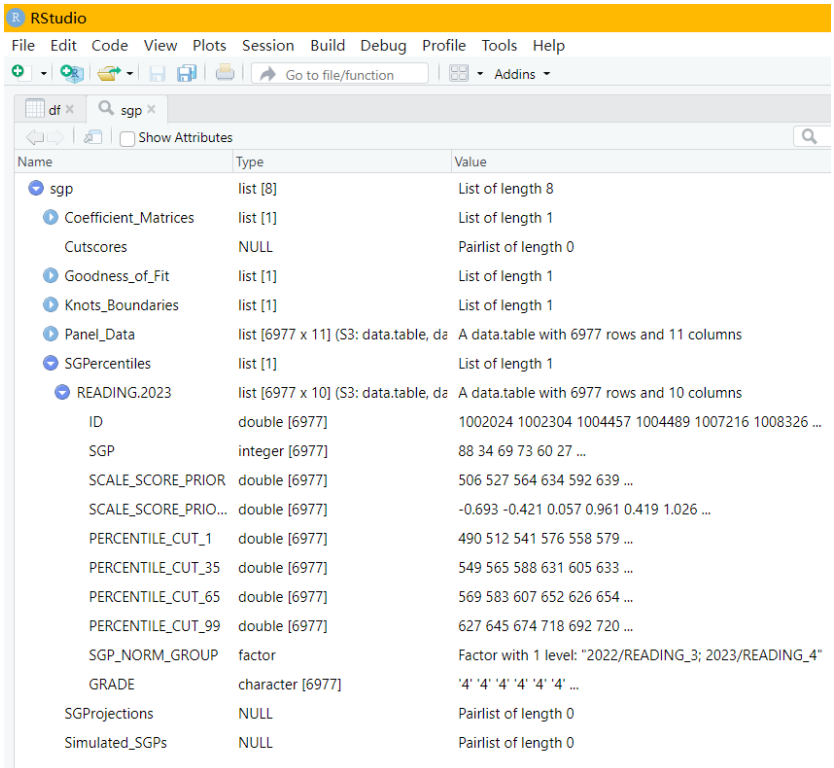
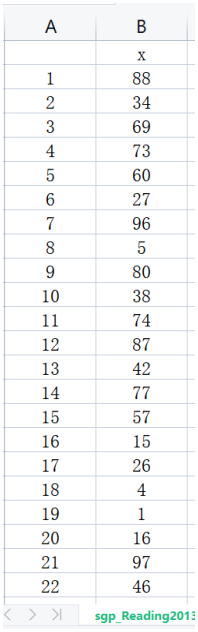


Figure 6: Viewing the sgp data object

Execute the following code to extract and export the SGP element. Partial results after export are shown in Figure 7.

```
rsgp_Reading2013 <- sgp[["SGPercentiles"]][["READING.2023"]][["SGP"]]
write.csv(sgp_Reading2013, file = "F:/sgpdata/sgp_Reading2013.csv")
```



A	B
	x
1	88
2	34
3	69
4	73
5	60
6	27
7	96
8	5
9	80
10	38
11	74
12	87
13	42
14	77
15	57
16	15
17	26
18	4
19	1
20	16
21	97
22	46

sgp_Reading201

Figure 7: Student Growth Percentiles

Limitations of the Student Growth Percentile Model

Despite the widespread application of the SGP model, its limitations warrant attention. These include the need for improvement in estimation methods, enhanced analytical precision, and the requirement for large sample sizes. First, the SGP model has substantial sample size requirements. The model assumes that groups of students with the same prior-year scores exhibit a normal distribution of subsequent-year scores, which imposes certain sample size demands. Generally, a sample size of at least 5,000 is required; otherwise, the stability of the percentile calculations can be compromised (Castellano & Ho, 2013a).

Second, there are frequency and interval requirements for longitudinal data. SGP requires at least two time points of longitudinal data, and the interval between assessments should not be too short or too long (Xie & Zhang, 2021). Third, variability of the tested population affects model performance. While SGP does not impose strict requirements on the test population, if the ability levels of the subjects are overly concentrated with insignificant variability, the reliability of the school performance estimates derived from the model's basic principles decreases, potentially rendering the value-added calculations less meaningful (Fu et al., 2019).

Fourth, questions remain regarding model reliability and validity. Concerns exist both regarding the model's inherent predictive accuracy, which can be relatively low with significant discrepancies between predicted and actual outcomes, and its assumption of equal growth potential and amplitude for all students. This assumption often contradicts reality, as the growth potential for very high- and very low-scoring students typically differs from that of middle-scoring students (Zhou & Liu, 2020; Fu et al., 2019). Continuous optimization and refinement of the SGP model are necessary to improve its predictive accuracy and reliability.

Fifth, issues of assessment content consistency present challenges. SGP does not impose strict requirements on assessment content. As students progress through grades, the content assessed changes. Evaluating growth based on scores from different assessments inherently introduces uncertainty (Fu et al., 2019). Sixth, the model demonstrates limited explanatory power of results. While SGP can assess

student performance trends, it does not provide extensive information on the reasons behind varying growth rates (Xie & Zhang, 2021). The model does not isolate the effects of factors such as family, school, or teacher characteristics on student learning, yielding a composite result encompassing all influences. In practical application, results need to be analyzed in conjunction with other contextual factors, such as student background, family environment, and educational resources (Fu et al., 2019).

Finally, there is limited potential for reuse of results. Model results are presented as percentile ranks, constituting ordinal data that is non-normally distributed. This format limits their potential for subsequent, more sophisticated statistical analyses (Fu et al., 2019).

Future Development Directions

To address these limitations, future research and development should focus on several key areas. First, there is a need to refine and improve the estimation methods for the SGP model to enhance its analytical precision. This could involve incorporating more advanced data analysis techniques and algorithms, such as machine learning and artificial intelligence. These technologies can facilitate better handling of large-scale datasets and improve the model's predictive accuracy and reliability.

Second, further research is needed to explore the applicability and effectiveness of the SGP model across diverse contexts. More empirical studies targeting different regions, school types, and student grade levels are necessary to validate the model's generalizability and effectiveness. Additionally, research should focus on developing and adapting SGP models suitable for different subpopulations to better meet varied practical needs.

Third, efforts should be made to enhance the explanatory power and reusability of SGP model results. Regarding explanation, further research should investigate the various factors influencing student growth and integrate these into the model framework. This would lead to a better understanding of student growth processes and influencing factors, providing more valuable insights for educational policy and practice. Concerning reuse, more flexible and diversified methods for applying the model outputs should be developed to leverage the assessment results more effectively.

Finally, research strengthening the cross-cultural and cross-disciplinary application of the SGP model is essential. Student growth is a global concern in education, and the SGP model, as a valuable assessment tool, can be applied across different cultures and settings. By enhancing cross-cultural and cross-disciplinary application research, the reach and impact of the SGP model can be further extended.

Future research priorities include: refining SGP estimation methods by developing and testing estimation methods that maintain accuracy with smaller sample sizes and can better handle extreme scores; integrating advanced analytical techniques by exploring the application of machine learning and artificial intelligence to improve the prediction precision of student growth models and handle complex, multi-source data; conducting cross-cultural and cross-grade empirical studies to validate the generalizability and effectiveness of the SGP model across diverse educational contexts, regions, and grade levels within China; and developing multi-dimensional growth assessment frameworks by creating comprehensive evaluation frameworks that incorporate SGP within a broader set of metrics assessing students' holistic development ("Five-Sphere Integration"), moving beyond purely academic growth.

Application Suggestions

Based on the characteristics and limitations of the SGP model, several recommendations are proposed to facilitate its effective application in educational practice. First, there is a need to enhance educator training by developing targeted training programs to improve teachers' and administrators' understanding of the SGP model's principles and technical operation. Second, educational institutions should optimize data infrastructure by establishing robust longitudinal data systems to ensure the continuity and consistency of student achievement tracking required for SGP calculation.

Third, practitioners should adopt a comprehensive evaluation approach by integrating SGP results with other relevant indicators such as student background factors and classroom observations to provide a more nuanced and context-rich interpretation of student growth. SGP is best used as one component within a comprehensive assessment system. Finally, efforts should focus on developing localized tools by creating user-friendly, localized software interfaces or packages that lower the technical barrier to implementing SGP analysis, promoting its broader adoption in schools.

Conclusion

The Student Growth Percentile (SGP) model represents a significant advancement in educational value-added assessment, offering an intuitive and relatively accessible approach to measuring student academic progress. This study has systematically examined the theoretical foundations, operational procedures, and practical applications of the SGP model, with particular attention to its implementation in the Chinese educational context.

The SGP model's primary strength lies in its conceptual clarity—by comparing students with similar prior achievement histories, it provides meaningful information about relative growth that is more intuitive for educators, parents, and policymakers than traditional growth measures. The model's flexibility regarding data assumptions and its operational simplicity compared to hierarchical linear models make it particularly attractive for widespread adoption. Through the detailed demonstration using the SGPdata dataset and R statistical software, this study has illustrated that SGP calculation, while requiring technical competence, is achievable for educational practitioners with appropriate training and resources.

However, several limitations require careful consideration. The model's substantial sample size requirements (typically 5,000+ students), sensitivity to measurement error, and limited explanatory power regarding the causes of differential growth patterns constrain its applicability in certain contexts. The assumption of equal growth potential across all ability levels remains contested, and the ordinal nature of SGP results limits their use in subsequent sophisticated statistical analyses. These limitations underscore the importance of viewing SGP as one component within a comprehensive assessment system rather than a standalone evaluation tool.

For China's educational evaluation reform, the SGP model offers a promising pathway forward, particularly given the policy mandate to explore value-added assessment approaches articulated in the 2020 educational evaluation reform plan. The successful pilot programs in Shanghai, Jiangsu, and Shenyang demonstrate the model's viability in Chinese contexts. However, effective implementation requires investment in educator professional development, establishment of robust longitudinal

data systems, and development of localized analytical tools that reduce technical barriers.

Future research should prioritize refining estimation methods to maintain accuracy with smaller samples, integrating advanced analytical techniques such as machine learning, conducting cross-cultural validation studies, and developing multi-dimensional frameworks that assess holistic student development. By addressing current limitations while capitalizing on the model's inherent strengths, the SGP approach can contribute meaningfully to evidence-based educational policy and practice, supporting the transition from purely achievement-based metrics toward comprehensive growth monitoring that recognizes and values diverse student learning trajectories.

References

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

Betebenner, D. W. (2021). On the norm- and criterion-referenced interpretations of student growth percentiles. Journal of Educational Measurement, 58(4), 456–478.

Betebenner, D. W., Van Iwaarden, A. R., & Domingue, B. (2024). SGPdata. https://centerforassessment.github.io/SGPdata/#

Castellano, K. E., & Ho, A. D. (2013). A practitioner's guide to growth models. Council of Chief State School Officers. https://doi.org/10.3102/1076998611435413

Castellano, K. E., & Ho, A. D. (2013b). Contrasting OLS and quantile regression approaches to student "growth" percentiles. Journal of Educational and Behavioral Statistics, 38(2), 190–215.

Central Committee of the Communist Party of China, & State Council. (2020). Overall plan for deepening the reform of educational evaluation in the new era.

Chen, J. B., & Ding, J. J. (2008). A survey of quantile regression techniques. Statistics & Information Forum, 23(3), 89–96.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. U.S. Government Printing Office.

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association, 89(428), 1314–1328. https://doi.org/10.1080/01621459.1994.10476871

Darling-Hammond, L. (2022). Transforming education through assessment reform: Global perspectives and local practices. Educational Researcher, 51(3), 189–200.

Fu, H. J., Wang, C. Y., & Cheng, Q. (2019). Academic progress evaluation based on the student growth percentile model. China Examinations, 2, 14–18.

Harris, D. N. (2020). Value-added measures in education: What every educator needs to know. Harvard Education Press.

Jiangsu Provincial Department of Education. (2023). Experimental report on value-added evaluation in basic education. Jiangsu Education Press.

Lockwood, J. R., & McCaffrey, D. F. (2020). Simulation-extrapolation for student growth percentiles. Journal of Educational and Behavioral Statistics, 45(3), 287–313.

McCaffrey, D. F., & Castellano, K. E. (2021). Improving the accuracy of student growth percentiles. Educational Measurement: Issues and Practice, 40(3), 35–47.

McCaffrey, D. F., Castellano, K. E., & Lockwood, J. R. (2015). The impact of measurement error on the accuracy of individual and aggregate SGP. Educational Measurement: Issues and Practice, 34(1), 15–21. https://doi.org/10.1080/01621459.1994.10476871

Meyer, R. H., & Dokumaci, E. (2023a). Value-added models and the next generation of assessments. Educational Researcher, 52(1), 45–58.

Meyer, R. H., & Dokumaci, E. (2023b). Value-added models and the next generation of assessments. https://www.ets.org/Media/Research/pdf/MeyerDokumaciPresenterSession4.pdf

Monroe, S. L. (2014). Multidimensional item factor analysis with semi-nonparametric latent densities [Doctoral dissertation, University of California, Los Angeles].

R Foundation. (2024). The Comprehensive R Archive Network. https://cran.r-project.org/

Raudenbush, S. W., & Bryk, A. S. (2022). Hierarchical linear models in educational research: A review of applications and developments. Journal of Educational and Behavioral Statistics, 47(2), 145–174.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. Journal of Personnel Evaluation in Education, 8(3), 299–311. https://doi.org/10.1007/BF00973726

Shang, Y. (2012). Measurement error adjustment using the SIMEX method: An application to student growth percentiles. Journal of Educational Measurement, 49(4), 446–465. https://doi.org/10.1111/j.1745-3984.2012.00186.x

Shang, Y., Van Iwaarden, A., & Betebenner, D. (2015). Covariate measurement error correction for student growth percentiles using the SIMEX method. Educational Measurement: Issues and Practice, 34(1), 4–14. https://doi.org/10.1111/emip.12058

Shanghai Municipal Education Commission. (2022). Shanghai pilot program on educational value-added assessment: Interim report. Shanghai Education Press.

Shenyang Education Bureau. (2023). Implementation plan for value-added assessment in Shenyang schools. Shenyang Education Press.

Woods, C. M., & Lin, N. (2009). Item response theory with estimation of the latent density using Davidian curves. Applied Psychological Measurement, 33(2), 102–117. https://doi.org/10.1177/0146621608319512

Xie, X. Q. (2019). Student growth percentile model for growth assessment: Experience from the United States. Educational Measurement and Evaluation, 6, 3–9.

Xie, X. R., & Zhang, H. R. (2021). The development dilemmas and solutions of student value-added assessment from the perspective of holistic education. China Educational Technology, 11, 32–38.

Zhang, X. (1998). Introducing the value-added concept to promote quality-oriented education. Education Guide, Z1, 14–81.

Zhou, Y., & Liu, H. Y. (2020). Exploring growth percentile estimation methods for nested data in educational value-added assessment: Application of the multilevel linear quantile regression model. China Examinations, 9, 32–39.

Zhu, K., Zhang, C., & Liu, Y. (2023). A comparative study of value-added assessment models for educational development from an international perspective. Journal of Distance Education, 41(1), 60–71.