Moving Beyond Scores: Reforming Examination Feedback to Foster Student Development

Feifei Huang[1], Zhaofeng Huang[2*]

[1] School of Education, Guangdong Polytechnic Normal University, Guangzhou, Guangdong, China

[2] School of Life Sciences, Guangzhou University, Guangzhou, Guangdong, China

Abstract: Recent educational assessment reforms in China mandate a fundamental shift from a singular focus on test scores toward comprehensive, diagnostic evaluations that support student development. However, a significant gap persists between these policy ambitions and current practices, which remain predominantly reliant on variable-centered analyses of composite scores. This article proposes a paradigm shift from a variable-centered to a person-centered analytical framework. We advocate for the application of two advanced statistical techniques, Latent Profile Analysis (LPA) and Multilevel Latent Class Analysis (MLCA). LPA allows for the identification of distinct, meaningful subgroups of students based on their multidimensional profiles of abilities and knowledge structures, moving beyond mere ranking to reveal heterogeneous learning patterns. Building on this, MLCA extends the analysis to account for the nested structure of educational data, enabling the simultaneous identification of latent typologies at both the individual and group levels. This integrated approach facilitates a more nuanced understanding of academic proficiency, providing a robust methodological foundation for generating diagnostic "student portraits" and "school profiles". The article concludes that this person-centered, multilevel framework is not only a technical advancement but also a strategic imperative for aligning educational assessment with the national reform agenda, ultimately fostering equitable and high-quality education.

***Keywords*:** latent profile analysis, multilevel latent class analysis, educational assessment reform

1. The Policy-Practice Gap in Educational Assessment Reform

In China, the feedback on examination results remains largely confined to reporting a score or grade, falling significantly short of leveraging assessment outcomes to inform and improve instructional practices. This status quo stands in stark contrast to the high-level design of educational evaluation reform in the new era. The "Outline of Plan for Building China into a Leading Country in Education(2024–2035)", issued by the Central Committee of the Communist Party of China and the China State Council in January 2025, serves as a guiding programmatic document for educational development over the next decade. It strongly reiterates and deepens the commitment to "overcoming the chronic malpractices of solely focusing on test scores". Further, it issues an explicit directive to improve outcome evaluation, strengthen process assessment, explore value-added methods, and enhance comprehensive appraisal. Similarly, the "Evaluation Guide for the Quality of High School Operations", issued by the Ministry of Education in January 2022, focuses on the operational level, requiring to "refine evaluation content, improve evaluation methods... and strive to overcome the tendency of 'solely focusing on test scores'," and to "prevent and correct deviations such as score supremacy". These policy documents clearly indicate a fundamental shift at the national level in how examination and evaluation are perceived. This represents a move away from their function as a mere "screening tool" toward roles as a "catalyst" for holistic student development and a "stethoscope" for diagnosing the educational process. However, a significant gap persists between the high-level policy guidance and its practical implementation.

The examination reform policies have introduced new requirements for outcome assessment to be "scientific, equitable, and comprehensive". However, most current practices in outcome assessment still predominantly rely on technical approaches that treat test scores as simple continuous variables for mean comparisons, rankings, or regression analysis. This conventional pathway fails to effectively address the deeper concerns underpinning the policy agenda. First, how can we scientifically evaluate the potential qualitative differences among individual students within educational instruction? Single composite scores obscure the heterogeneity in students' internal cognitive structures and competency proficiency. For example, two students with identical total scores may possess entirely different patterns of knowledge mastery, cognitive frameworks, and strengths in abilities. As emphasized in the "Outline of Plan for Building China into a Leading Country in Education(2024–2035)", the goal of education is to foster diverse talents through quality-oriented education. This necessitates that our assessments must be capable of identifying distinct competency profiles of students, moving beyond mere linear

ranking based on score levels. Second, how can the scores from outcome assessments more comprehensively reflect students' knowledge, abilities, and competencies, and more authentically represent instructional quality and educational development levels? Future outcome feedback must evolve from a single number to a diagnostic report. This report should be based on multidimensional data integration, incorporating not only detailed analyses of disciplinary core competencies but also process data, value-added assessments, and comprehensive quality performance indicators. Through this approach, the feedback provided to students would constitute a clear "competency map" and "personalized learning pathway suggestions". For teachers and schools, it would offer precise diagnostics regarding instructional effectiveness and deficiencies, revealing the structural patterns of collective strengths and common weaknesses at the group level, thereby providing an empirical basis for optimizing teaching strategies and equitably allocating educational resources.

2. A New Paradigm for Multidimensional Academic Quality Assessment: Constructing a Student-Specific Profile

Examination reform policies have introduced a multitude of new requirements for the evaluation of assessment results. However, the majority of existing outcome evaluations still predominantly analyze test scores as primary variables. Therefore, how can we accurately assess the potential individual differences among students in the teaching-learning process? Furthermore, how can the scores from outcome evaluations more comprehensively reflect students' knowledge, abilities, and competencies? And how can the scores provide a more authentic representation of instructional quality and educational development levels?

As educational reform deepens, key priorities such as educational equity, equitable resource allocation, scientific assessment, and diversified pathways have emerged as major trends. Within this context of diversification, significant technical challenges arise in conducting multidimensional evaluations and monitoring of academic quality across different regions and disciplines, and in providing timely, effective feedback. These challenges represent urgent issues that basic education academic quality monitoring must address. The "Compulsory Education Academic Proficiency Monitoring" plays a pivotal role in this landscape. Its purpose is to implement quality monitoring in primary and secondary schools to better understand and grasp the effectiveness of teachers' instruction, students' learning, and school principals' management. This process generates first-hand educational data, providing reliable and accurate information for government decisions on educational investment and reform, thereby supporting the provision of higher-quality educational services.

For a long time, the analysis and evaluation of academic proficiency data by examination authorities have primarily focused on test scores, using them as the benchmark for judging performance. Consequently, the utilization of examination data to provide feedback for instructional improvement remains severely underdeveloped. Research indicates that most current outcome evaluations treat examination scores as variables for analysis, overlooking the potential individual differences among students (Bergman & Lundh, 2015)[1]. Other studies argue that this variable-centered approach tends to lead to an overemphasis on scores, which runs counter to the reform policy's objective of moving beyond the "score-only" (Guan, Han, & Zhang, 2021)[2]. However, the educational process involves students as the main participants, with parents and teachers engaged, and is managed by schools, higher administrative departments, and the government. Therefore, basic education academic proficiency monitoring should comprehensively investigate the current state, interrelationships, and developmental patterns of student learning, teacher instruction, and school management.

Currently, China's basic education academic proficiency monitoring system has matured significantly, with its collected data characterized by typical multidimensional, multilevel, and dynamic features. Yet, in stark contrast to this rich data structure, traditional outcome evaluation paradigms remain confined to simplified reporting models. These reports largely rely on single composite scores or grades to categorically define students, and simplistically use average student achievement as the core — often sole — indicator for evaluating teaching effectiveness and school quality. Methodologically, this practice severely obscures the wealth of information inherent in the monitoring data concerning student knowledge structures, competency gaps, group heterogeneity, and patterns of inter-school variation. Consequently, the value of this precious monitoring data is vastly underutilized. Advancing educational assessment reform presents a critical and pressing challenge.

A closer examination of current educational assessment practices reveals that mainstream analytical methods predominantly adhere to a variable-centered paradigm. This approach treats variables as the central constructs and fundamental units of analysis. This process involves inferring overall trends by examining correlations or average effects between these variables. This method can uncover general associations between certain influencing factors and academic outcomes at a macro-theoretical level. However, its limitations for guiding educational practice are inherent and significant. On one hand, it operates on the assumption of sample homogeneity, neglecting the prevalent and educationally critical reality of individual differences. On the other hand, this approach inevitably leads to an excessive emphasis on isolated score

metrics in practice. By focusing on variables, it subtly reinforces a score-only mentality, which fundamentally contradicts the core principle of current educational assessment reform.

Therefore, within the broader push to deepen the application of basic education academic proficiency monitoring, a methodological restructuring is imperative. We should vigorously advocate for a strategic shift from the variable-centered to a person-centered analytical paradigm. This paradigm treats the whole, contextually situated student as the fundamental unit of analysis. It aims to employ statistical techniques to identify latent subgroups within the student population that exhibit distinct, multidimensional profiles of characteristics.

3. Latent Profile Analysis and Multilevel Latent Class Analysis

While traditional variable-centered methods can effectively isolate specific individual characteristics for examination, they treat these characteristics in isolation. This approach fundamentally limits the ability to explore the complex interactions that occur between different traits across individuals. In contrast, the person-centered approach refers to an analytical technique that classifies individuals into distinct subgroups based on shared response patterns. In essence, it is a classification methodology aimed at identifying population heterogeneity (Schmiege, Masyn, & Bryan, 2018)[3]. However, the application of this method to educational assessment using academic data has been scarcely explored in prior research. As examination reform policies introduce numerous new requirements for evaluating assessment outcomes, traditional methods of processing exam data fall short of meeting the evolving needs of educational assessment. The person-centered approach offers a novel perspective to address this critical gap.

Conventional educational assessment methods have largely been confined to using examination scores to evaluate individuals. In reality, there are often qualitative differences between an individual's observable performance and their underlying competencies and literacy. For instance, different students who both score 85 on a mathematics exam may exhibit distinct response patterns across individual items, indicating potential qualitative differences in the essential knowledge and key competencies they have mastered. Existing research has found that LPA outperforms hierarchical cluster analysis in classifying heterogeneous populations. LPA is a statistical method that explains the density distribution of responses to observed variables through a few mutually exclusive latent categorical variables (Collins & Lanza, 2010)[4]. It primarily handles the latent classification of continuous observed variables. Similar to traditional cluster analysis, LPA offers several methodological advantages: it imposes less stringent requirements on the scale and distribution of observed variable data. Further, it provides estimates of classification error, enhancing precision. And it accounts for mixed data distributions in its statistical assumptions, often aligning more closely with the true nature of educational data.

The application of a latent profile model rests on two fundamental assumptions. Primarily, all latent profiles must be mutually exclusive. Additionally, the observed variables within each profile must be locally independent. LPA generally involves four key steps. First, model estimation begins with a null model (one profile) as a baseline. Sequentially increase the number of latent profiles, estimating parameters and calculating model fit indices for each. Then, compute the means and conditional probabilities for the latent profiles. Second, Compare the fit of the competing models to determine the optimal solution. Commonly used fit indices include information criteria (e.g., AIC, BIC, cAIC, aBIC), likelihood ratio tests (e.g., LMR, aLMR, BLRT), and classification accuracy measures (e.g., Entropy). Generally, lower values for information criteria and higher entropy values indicate better model-data fit. Third, based on the conditional probabilities of the observed variables within each profile, characterize the distinctive features of each latent profile and assign descriptive labels. Last but not the least, after assigning individuals to profiles based on posterior probabilities, subsequent analyses can be conducted to examine how profile membership relates to other external variables.

In summary, LPA focuses on heterogeneity among individuals, capable of not only accurately depicting quantitative differences but also summarizing multidimensional qualitative differences between individuals. Therefore, applying LPA to outcome assessment aligns more closely with real educational contexts, resonates with the pedagogical principle of "teaching students according to their aptitude", and can provide educators with more targeted instructional strategies.

MLCA is a non-parametric model that posits a categorical group-level latent variable. This proposition means that latent classes exist at the group level, and within them, the model parameters are homogeneous (Vermunt, 2003)[5]. Compared to traditional latent class models, MLCA often demonstrates superior model fit. This indicates that MLCA offers two key advantages. First, it optimizes both parameter estimation and classification accuracy. Second, it enables the investigation of potential types among regional or school groups (Henry & Muthén, 2010)[6]. This capability allows educational policies, guidelines, and management measures to be tailored to different educational levels (Zhang, Zhang , Jiao, & Wang, 2013) [7]. Building on the differentiation of individual knowledge and ability structures, we can further classify groups by their educational level types. It will significantly contribute to a more comprehensive consideration of educational resource

disparities, thereby facilitating more equitable and effective resource allocation.

Acknowledgments

This paper was supported by National Education Examinations Authority Ministry of Education Special Project in Support of the 14th Five-Year Plan, “Research on a Multi-dimensional Assessment Model for Data based on the China College Entrance Examination assessment framework” (NEEA2021037).

References

[1] Bergman, L. R., & Lundh, L. G. Introduction: The person-oriented approach: Roots and roads to the future[J]. Journal for Person-Oriented Research, 2015, 1(1): 1–6.
[2] Guan, D.D., Han, N., & Zhang, J.S. Deepening the Reform of Gaokao Evaluation by Focusing on “Four Evaluations”and Serving“Five Main Bodies”[J]. Journal of China Examination, 2021, (3): 1-8.
[3] Schmiege, S. J., Masyn, K. E., & Bryan, A. D. Confirmatory latent class analysis: Illustrations of empirically driven and theoretically driven model constraints[J]. Organizational Research Methods, 2018, 21(4): 983–1001.
[4] Collins, L. M., & Lanza, S. T. Latent class and latent transition analysis: wth applications in the social, behavioral, and health sciences. Hoboken, NJ: Wiley; 2010.
[5] Vermunt, J.K. Multilevel latent class models[J]. Sociological Methodology, 2003, 33: 213-239.
[6] Henry, K.L., & Muthén, B. Multilevel Latent Class Analysis: An Application of Adolescent Smoking Typologies With Individual and Contextual Predictors[J]. Structural Equation Modeling: A Multidisciplinary Journal, 2010, 17: 193-215.
[7] Zhang, J.T., Zhang, M.Q., Jiao, C., & Wang, L.T. Application of the Multilevel Latent Class Model to Educational Assessment: An Example from an English Academic Proficiency Test[J]. Educational Research and Experiment, 2013, (3): 78-84.