

The Construction and Application of Regional Education Quality Monitoring Databases: A Case Study of Suzhou's Education Quality Monitoring

Jian Shen,¹ Qiang Luo²

1. Suzhou High School-SIP, Suzhou 215000, Jiangsu, China
2. Suzhou Education Quality Monitoring Center, Suzhou 215000, Jiangsu, China

Abstract: *The development of school education depends on the quality of the education provided, and it is a key metric for assessing the effectiveness of schools in developing talent. Building specialized, intelligent education quality monitoring (EQM) databases is crucial for speeding EQM progress in the big data era. This article examines the development of regional EQM databases in the areas of operational procedure and logical structure based on the idea of data lakes by using the development of databases for the EQM data analysis system (DAS) in Suzhou City as a case study. The goal of this study is to assist in addressing the current issues with regional EQM data processing and ensuring EQM's successful implementation.*

Best Evidence in Chinese Education 2022; 12(2):1613-1628.

Doi: 10.15354/bece.22.re031

How to Cite: *Shen, J., & Luo, Q. (2022). The construction and application of regional education quality monitoring databases: A case study of Suzhou's education quality monitoring. Best Evidence in Chinese Education, 12(2):1613-1628.*

Keywords: *Education Quality Monitoring, Data Analysis System, Data Lake, Database*

About the Authors: Jian Shen, Suzhou High School-SIP, Suzhou 215000, Jiangsu, China. E-mail: james@suzhou.edu.cn

Qiang Luo, Suzhou Education Quality Monitoring Center, Suzhou 215000, Jiangsu, China. E-mail: 452384761@qq.com

Correspondence to: Qiang Luo at Suzhou Education Quality Monitoring Center of China.

Funding: This study is part of the study "Regional Education Quality Improvement Paths Based on EQM Big Data" (E-b/2020/16), funded by Jiangsu's 13th Five-year Plan for Education Sciences.

Conflict of Interests: None.

© 2022 Insights Publisher. All rights reserved.



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License

(<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed by the Insights Publisher.

Introduction

Education quality is crucial to the advancement of school education because it is an indicator of school performance, educational development, and talent training. Building a system to monitor the quality of compulsory education is an essential step in deepening the reform of educational evaluation in the new era and changing the unscientific orientation of educational evaluation. The State Council's Overall Plan for Deepening Educational Evaluation Reform in the New Era, released in October 2020, proposes the use of innovative evaluation tools as well as the application of artificial intelligence, big data, and other modern information technologies to develop a longitudinal evaluation system for the entire process of students' learning in all grades as well as a transversal system to comprehensively evaluate the results of moral, intellectual, physical, and social development (State Council, 2020).

A database is a warehouse that organizes, stores, and manages data in accordance with a data structure; it can connect and interact with data through tools such as data collection, organization, analysis, and visualization to provide evidence for scientific research and decision-making processes.

Education quality encompasses, among other things, the outcomes of school operation, teacher instruction, student learning, and family education. The education quality monitoring (EQM) database collects information about participating schools, teachers, students, and parents and stores and analyzes it using information technology, computer technology, and analytical data mining.

The domestic study on Compulsory Education Quality Monitoring (CEQM) focuses mostly on two topics. One is to examine its implementation measures and support mechanisms, such as the research and development of monitoring tools (Wang, 2016), the establishment and operation of monitoring institutions (Zhang, 2010), the application of monitoring results (Yang, 2019; Guo & Wang, 2016), and IT support for monitoring (Wang, 2016; Zhang et al., 2016). The second field of research focuses on EQM's implementing actors and the development of a monitoring system that includes EQM at the national, provincial, municipal, and district levels (Li & Chen, 2020; Xin & Zhao, 2020; Li et al., 2017; Zhou, 2016). These two interrelated topics are the primary focus of CEQM research. However, there is a paucity of research on how to leverage big data technology and appropriate theories to generate an EQM database and build a data analysis system (DAS) for EQM, which has impeded the development of education quality monitoring.

The purpose of this study is to examine the building of the regional EQM database using the construction of the EQM database in Suzhou as an example in order to address the difficulties in data analysis.

The Status Quo of EQM Development in Suzhou

Massive amounts of data are involved in EQM. Suzhou's CEQM 2021, for example, generated more than 67.15 million data records during the data collection stage, each of which is made up of data of various types and structures. It is necessary to retrieve relevant data from previous years for follow-up analysis during the data analysis stage. Every year, hundreds of millions of raw monitoring data points must be processed. With such a large amount of data and complex structures, the planning and construction of an EQM database is an urgent task that must be completed.

The “*Compulsory Education Academic Quality Monitoring Project*” was launched in Suzhou City, Jiangsu Province, in 2015, and the Suzhou Education Quality Monitoring Center (hereinafter referred to as the Suzhou EQM Center) has been tracking and monitoring junior secondary school students ever since. More than 1.3 million students had been tested by 2021, more than 180 million pieces of monitoring data been collected, and more than 14,000 monitoring reports been issued (Song & Luo, 2021). The Suzhou EQM Center strengthens the application of intelligent technology in monitoring practices in response to educational evaluation reform in the artificial intelligence era and builds a regional EQM database with Suzhou characteristics. It has also established a data analysis system and made incremental technological advances in standardized data governance, intelligent data analysis, and visualized data presentation.

Suzhou EQM Center created the DAS framework based on its own needs (**Figure 1**), which divides the data analysis procedure into six processes involving 12 functional modules.

The DAS provides data assistance for the data presentation step after the data gathering stage. Data import, data cleaning, quality analysis, project establishment, calculation analysis, and data push are the six steps that make up the data analysis process.

Based on the aforementioned structure, two additional structural layers - algorithm support and Data Lake - are added to the data analysis system to create a full analysis system design. To be exact, the data lake contains six databases, including raw data, virtual hierarchical data, user rights control, desensitized project data, algorithm rule resources, and result presentation data. The algorithm support layer also contains four algorithm libraries (**Figure 2**).

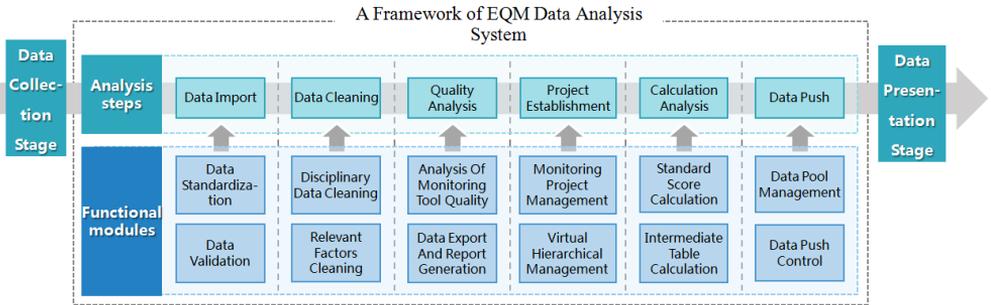


Figure 1. Procedure and Architecture of Data Analysis System for Suzhou EQM.

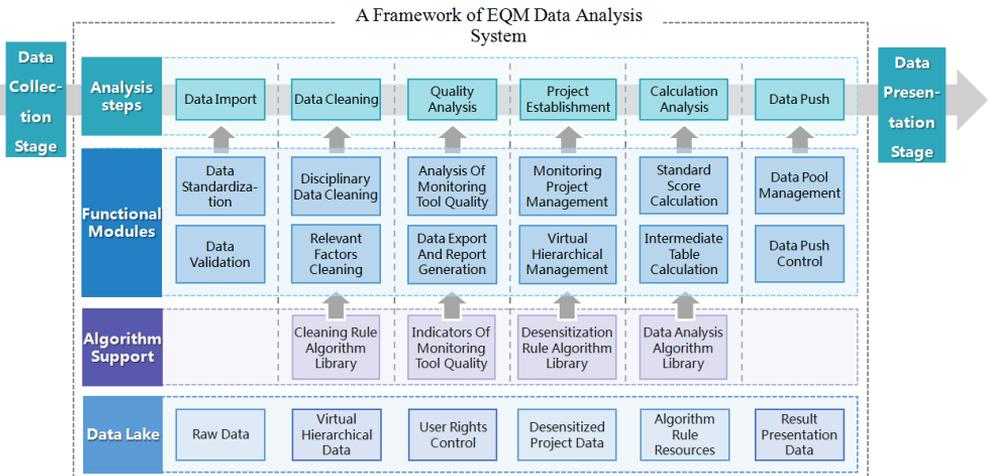


Figure 2. A Plan of Suzhou EQM Data Analysis Databases.

The Construction Paths of Regional EQM Databases

In the new era of big data and data science, having a centralized data architecture that is consistent with operational processes is critical. This is also true in EQM. Good database architecture should be able to grow with the monitoring scale and evolve with technological advancements.

Suzhou EQM Center creates databases for EQM data analysis based on the “data lake” concept. The data lake unifies the storage of all organizational data, including both the original data in the source system and the converted data (Campbell, 2017). It has become an important tool for organ-

izations wishing to make use of big data. Structured data (relational database data), semi-structured data (CSV, XML, JSON, etc.), unstructured data (emails, documents, PDFs), and binary data (images, audio, and video) are all stored in the data lake, forming a centralized data storage that holds all types of data. It aggregates and stores streams from various data sources, much like a large lake in nature, and outputs valuable data based on specific needs. In terms of monitoring data, the data lake contains not only data from various platforms, such as a question bank system, an examination service system, a scanning and marking system, and so on, but also a wide range of files, such as spreadsheets, scanned images, databases, and so on. It also saves process and result data from various data analysis processes. Its inclusiveness enables the cross-analysis of diverse data information and the use of large capacity and high-speed data pipelines, as well as the management of the entire data lifecycle to make data flow processes such as access, storage, processing, and application traceable (**Figure 3**).

The logical structure design that stresses the security, integrity, and efficiency of all databases is also a crucial issue since the database must be constructed to support the monitoring procedure, making the design based on data analysis procedures the most significant component. The location of various storage types can only be determined by a database architecture that takes into account the two factors mentioned above. This architecture can also guarantee that data is safeguarded, effectively stored, and accurately processed.

Database Design Based on Analytical Processes

Separation of Raw Data and Project Data

Importing different types of data from various platforms is frequently necessary for EQM data analysis. Examples include participant, school, and regional information from the examination system; monitoring tool information, dimensional information, scoring instructions, and other data from the question bank system; and original response records, scoring records, scanned images, and other data from the scanning and marking system. The raw databases for EQM are made up of all this data.

Different logical principles must be called upon and applied to the raw data in accordance with the demands of monitoring programs. In the Regional School Quality Analysis Project, data must be retrieved on a regional basis, and in the Private School Project, data must be based on the classification of private schools for children of migrant workers or non-migrant workers. As an example, it is necessary to include the monitoring

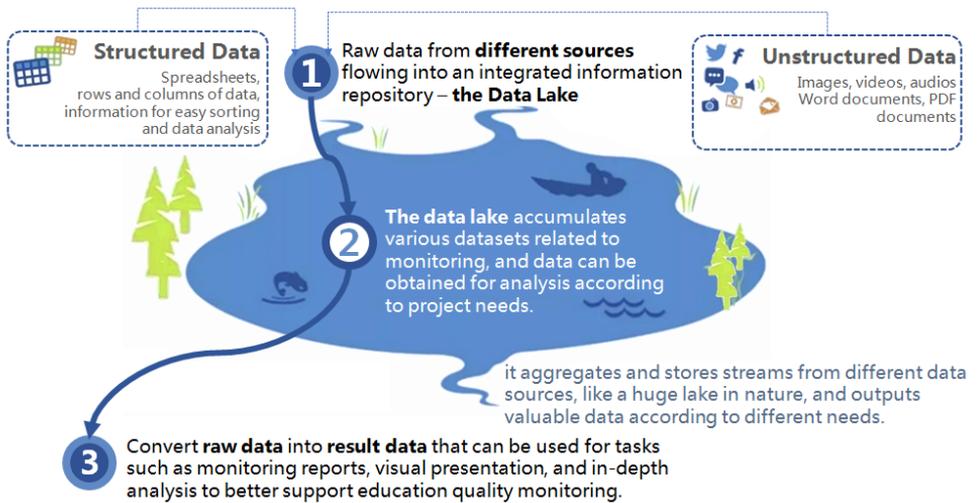


Figure 3. A Schematic Diagram of the Data Lake of Suzhou EQM Data Analysis System.

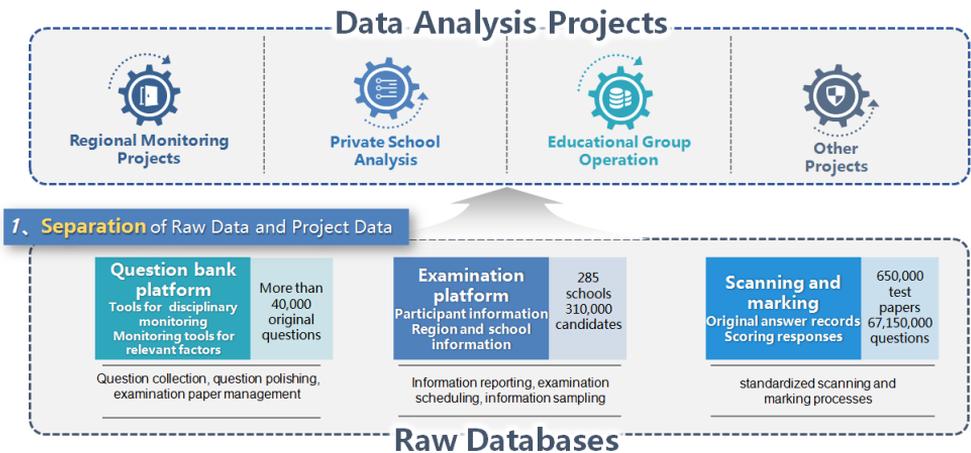


Figure 4. A Schematic Diagram of Raw Databases and Data Analysis Projects of Suzhou EQM 2021.

data of the same subject for consecutive years in the tracking analysis. Therefore, the raw database and the project database should be maintained separately in the data lake to meet the various calling logics of various data analysis projects (Figure 4).

Virtual Stratification in a Project-Based Database

The data is frequently converged and displayed by various school attributes in various data analysis initiatives. For instance, a three-level structure of “Suzhou city, district, and school” is necessary for calculation and report creation in the regional analysis project, whereas a three-level structure of “District, Group, and School” is necessary in the educational group project.

A distinct virtual stratification module is added to the system function module to construct a different database for each project based on the logic of virtual stratification to satisfy the needs of the project. According to the specifications of the data analysis project, a three-tier structure based on school attributes can be created, as illustrated in **Figure 5**, supporting a range of monitoring reports.

Database Design Based on Logical Structure

Hierarchical Management of Classified Databases

The database faces more internal security concerns because of its growing value and accessibility. For instance, unlawful overstepping operations and hostile infiltration result in the theft and disclosure of confidential information, but it is unable to adequately trace and audit the events that caused these events. Strengthening data security management is especially crucial for databases that contain a lot of personal data. Information leakage must be prevented at all costs, in addition to rapid improvements and effective vulnerability management. Database security protection techniques that are flexible and targeted are the most effective way to do this.

A crucial security measure for databases is database access control, and Suzhou EQM Center now mostly uses role-based access control (RBAC). Users and operation rights are intertwined with the idea of roles. To achieve hierarchical management, various roles are first developed in accordance with the organizational functions, each of which corresponds to a distinct level of operation privileges. RBAC must assign the user’s account a role and associated rights in accordance with the information in the role rights database, in addition to verifying the user’s identity and password when they log in to the system (**Figure 6**). The application system can be modified to meet the new access control requirements in the case of a functional shift in the organization by merely reassigning permissions to the roles.

Suzhou EQM Center has also implemented four different types of security protection measures for the database. In order to prevent operational paralysis in advance and ensure the ongoing availability of operational sys-

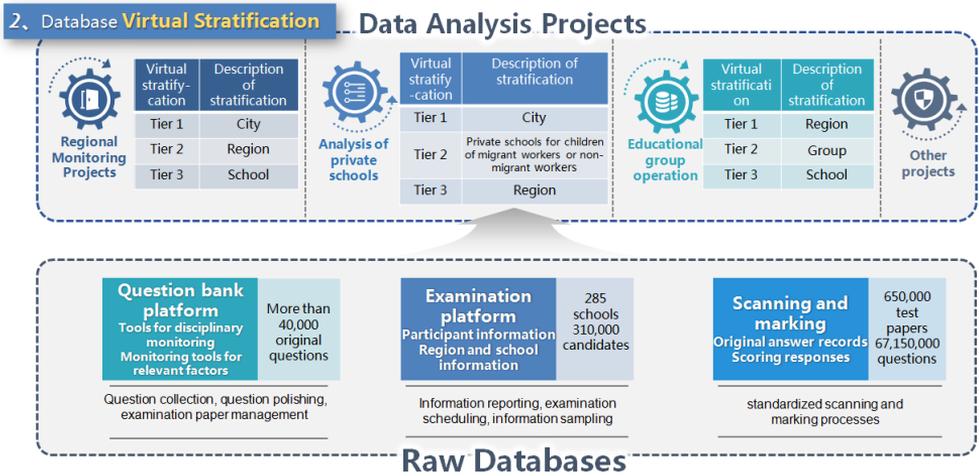


Figure 5. A Schematic Diagram of the Data Virtual Stratification Logic in Data Analysis Projects.

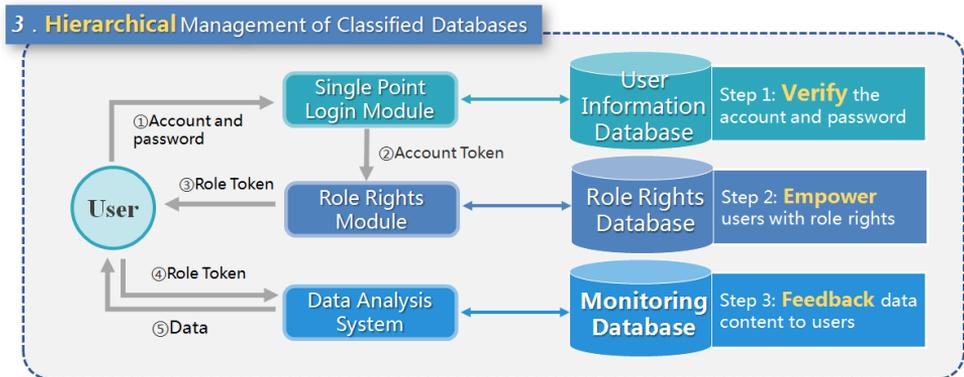


Figure 6. The Process of Account Authentication.

tems, it must be able to monitor the database operation state in real time and provide early warning when the condition is abnormal. The second is to be able to evaluate the database system’s risk, including weak password detection, system vulnerability, configuration risk, etc., to identify user and system access behavior patterns to the data and produce access rules with various levels of strength. Thirdly, it must be able to keep track of data activities in real time, create data access models, assess access risks promptly, find and prevent unauthorized access, and encrypt sensitive data to improve data se-

curity protection. The fourth is to use audit logs to carry out thorough user access behavior monitoring. Finding the risk source at the outset and following the individual responsible for the risk become required actions to close the security gap once the security risk has materialized. Using the audit log query tool, this issue can be fixed.

Desensitization and Recovery of Sensitive Data

For the purpose of scheduling exams, information on the monitoring subjects, such as student, teacher, and other personnel personal identifying information, must be gathered during the EQM process. A desensitization and recovery module should be set up in the system to centrally desensitize, encode, and save sensitive information in basic data in order to guarantee data security.

It is also necessary to establish an administrator position for the desensitization module, whose primary responsibility is to grant permissions to data in the desensitization and restoration module, because data analysis is a collaborative task that requires the participation of numerous people. This functional module and any associated forms and fields are not accessible to other roles in the system.

The coding rules for the desensitization and recovery of each datum are mostly stored in the desensitization rule algorithm library in the Suzhou EQM databases. These guidelines are used to desensitize the sensitive database used for data collection in order to obtain code-version basic data. To create the analysis result database with real names, the calculation and analysis module recovers the code-version result database. In order to offer data support for later data presentation and report creation, the desensitized code-version result database and the recovered name-version result database are finally pushed to the data presentation link through a data interface (see **Figure 7**).

Establishment of an Algorithm Rule Repository

The term “algorithm rule repository” primarily refers to the storage of data analysis algorithms as a new type of data asset and the formulation of pertinent specification processes in accordance with the guidelines and specifications of data asset management.

By building unique algorithm repositories for each functional module, the data analysis system controls them. Data production, data processing, data analysis, data preservation, data access, and data reuse are the six nodes that make up the loop structure model that the UK Data Achieve (UKDA) uses to characterize the cycle of research data (UK Data Service, n. d.). Su

and creative demands of data analysis for various monitoring projects, applicable analysis algorithms can be chosen from the algorithm library during the process of specific application for various monitoring projects. Additionally, new data analysis algorithms, various indicators, and charts can be customized in accordance with various analysis models.

The algorithm rule repository contains a cleaning rule algorithm library, a quality analysis index algorithm library, a desensitization rule library, and a data analysis algorithm library, which provide algorithm support for the four processes of data cleaning, quality analysis, data desensitization, and calculation of analysis, respectively. Each algorithm library has amassed a specific number of algorithms as a technological reserve and is regularly generating new algorithms as the EQM project advances. Taking the cleaning rule library as an example, to adapt to the needs of diverse projects, it has so far accumulated 106 cleaning methods for a variety of situations, such as missing tests, contradictory alternatives, contradictory logic between questions, invalid answers, etc. Also, using the data analysis algorithms library as an illustration, there are multiple algorithms for a single indicator of “percentile grade,” as well as algorithms that encapsulate multidimensional characteristics of the educational ecology, such as the balance of education, ecological health, etc.

Practical Application of EQM Databases in Suzhou

Although it only makes up a small portion of Suzhou’s overall EQM process, the development of databases is a key technology that underpins high-quality monitoring. Through these databases, the effective and reliable operation of each functional module of the Suzhou EQM data analysis system is ensured (**Figure 1**). This increases the effectiveness of the analysis and evaluation of monitoring data and satisfies the specific requirements of data analysis for the various Suzhou monitoring projects.

Using various quality analysis algorithms from the algorithm rule resource library, the monitoring tool quality analysis module in the data analysis system, for instance, provides quantitative indicators for analyzing test questions, improving tool quality, and improving teaching. It is based on classical measurement theory and structural equation modeling. As well as indicator algorithms used to verify the validity and reliability of each dimension of the pertinent factor instruments, there are algorithms for indicators of disciplinary tools, such as reliability, difficulty, differentiation, percentage of score bands, and ability value of score points. During the whole EQM 2021 process, this functional module offers more than 100 high-quality analysis reports for the monitoring tools at all grade levels and for all subjects,

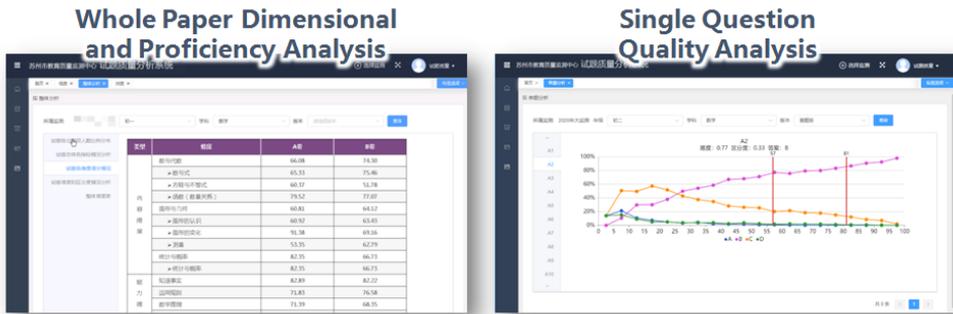


Figure 9. Screenshots of Some Functions of the Monitoring Tool Quality Analysis Module.

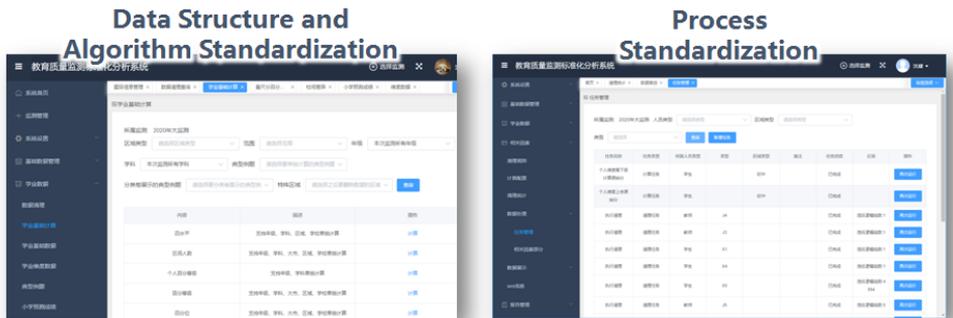


Figure 10. Screenshots of the Functions of Calculation and Analysis Related Modules.

providing a data foundation for the development and accumulation of monitoring tools (Figure 9).

In the “calculation and analysis” process, for instance, the virtual stratification database is utilized to ensure that the data required for each project can be reliably accessed in order to execute numerous projects simultaneously. Through the algorithms of various data analyses in the algorithm rule repository, a series of quality assurance measures, such as “double-track parallelism and double-blinded comparison; seamless docking and errorless flow; sampling verification and reverse verification,” are realized in the analysis. Using these databases and functional modules, the data processing procedure has been standardized, the effectiveness of data analysis has been enhanced, and the precision of data calculation has been ensured. During the

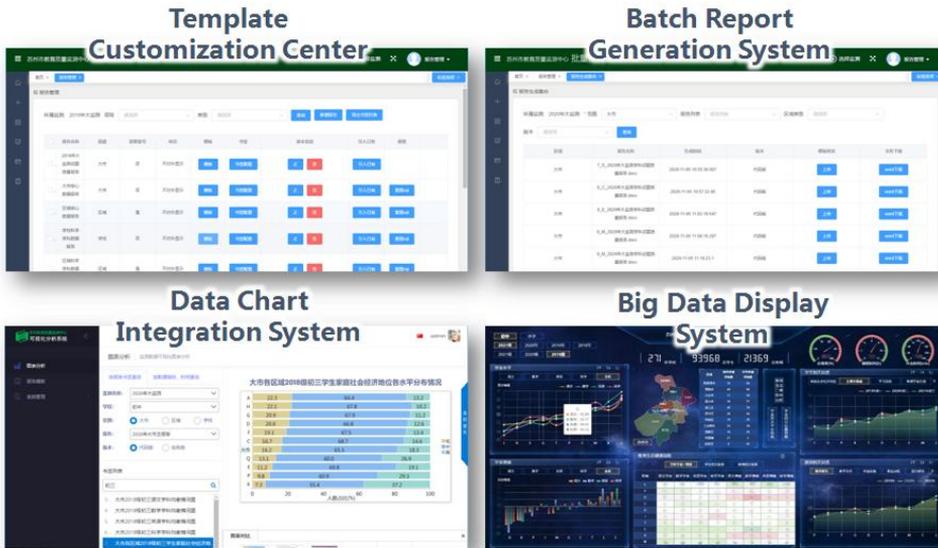


Figure 11. Screenshots of the Functions of Each System at the Data Presentation Stage.

data analysis in 2021, the Suzhou EQM Center completed the basic data processing in just two weeks with high efficiency and high quality and, on this basis, completed the projects of basic data of junior secondary schools, basic data of elementary schools, educational group research, and private school education quality, as well as a total of more than 3,600 reports of various monitoring results, which ensured the smooth implementation of Suzhou’s education reform (**Figure 10**).

Not only does the database play a crucial role in the data analysis process, but it also ensures the secure and efficient flow of data inside each system throughout the whole EQM process. By setting data requirements during the data gathering phase, the “Data Standardization” module completes the integration of data from disparate systems. Moreover, the database supplies data sources for subsequent data presentation, which are handled by the “Data Push Control” module and delivered on demand to the batch report generating system, the data chart integration system, the big data display system, and other platforms (**Figure 11**).

Conclusion

The development of databases for the data analysis system is a lengthy process. With the extensive and in-depth promotion of EQM, we will be con-

fronted with new issues and challenges, necessitating ongoing innovation and the study of new technologies in our research. In the era of big data, we believe that the building and maintenance of databases demands not only a reorganization of data structure but also a modernization of big data mentality and data asset management procedures. Together with our contemporaries in the field of EQM, we anticipate seizing these new chances and conquering these new obstacles.

References

- Campbell, C. (2017, 19 May). Top five differences between data lakes and data warehouses. Blue-Granite.com. Available at: <https://www.bluegranite.com/blog/bid/402596/top-five-differences-between-data-lakes-and-data-warehouses>
- Guo, D. H. & Wang, H. Y. (2016). The implementation of compulsory education quality monitoring: A study based on the survey of 16 districts and counties in Tianjin. *Journal of Tianjin Academy of Educational Sciences*, 2016(4):32-35. [Chinese]
- Li, L. Y. & Chen, H. J. (2020). Basic strategies and support mechanisms for promoting the construction of China's basic education quality monitoring system. *Journal of the Chinese Society of Education*, 2020(3):68-73. [Chinese]
- Li, L. Y., Ren, X. Q., & Jiang, Z. F. (2017). Regional compulsory education quality monitoring in China. *Journal of the Chinese Society of Education*, 2017(12):37-41. [Chinese]
- Research Data Lifecycle. (n. d.). UK Data Service. [EB/OL]. Available at: <http://www.data-archive.ac.uk/create-manage/life-cycle>
- Song, Y. D. & Luo, Q. (2021). Monitoring based on big data for more pertinent education for children: An analysis based on data from academic quality monitoring of 270816 junior middle school students in Suzhou. *Future Educators*, 2021(1):63-67. [Chinese]
- State Council of China. (2020). The Overall Plan for Deepening Educational Evaluation Reform in the New Era. [EB/OL]. Available at: http://www.gov.cn/xinwen/2020-10/13/content_5551032.htm
- Wang, Y. C. (2016). The research and development of provincial basic education quality monitoring tools. *Times Education*, 2016 (24): 33. [Chinese]
- Xin, T. & Zhao, Q. (2020). Orientation, structure and support for basic education quality monitoring and evaluation system. *Journal of the National Academy of Education Administration*, 2020 (9): 16-23, 43. [Chinese]
- Yang, Y. Y. (2019). How to utilize monitoring results of regional compulsory education quality: Reflections based on the use of monitoring results at the school level. *Theory and Practice of Education*, 39(26):3-6. [Chinese]
- Zhang, A. L. (2010). Building a government-supported independent mechanism for basic education quality monitoring. *Educational Measurement and Evaluation*, 2010(8):12-15. [Chinese]
- Zhang, M. Q., Fan, X. Z., Guo, K. Y., & Zhong, Y. (2016). The reform and development of education quality monitoring and evaluation under the concept of big data. *Modern Education Journal*, 2016 (3): 2-6. [Chinese]
- Zhou, J. R. (2016). The mechanism for the construction of basic education quality monitoring. *Shanghai Journal of Educational Evaluation*, 5(5):1-7. [Chinese]

Shen & Luo. (Jiangsu). Build Up Regional Education Quality Monitoring Databases.

Received: 15 September 2022

Revised: 05 October 2022

Accepted: 01 November 2022