Akademika 96(2), July 2026
https://doi.org/10.17576/akad-2026-9602-11

A Validated Big Data Management Model for Malaysian District Education Offices

Model Pengurusan Data Raya Pejabat Pendidikan Daerah di Malaysia

NOR KAMALIAH MOHAMAD, NORFARIZA MOHD RADZI*
& ZURAIDAH ABDULLAH

Received: 6-10-2025 /Accepted: 20-05-2026

ABSTRACT

Despite the rapid proliferation of Big Data within the Fourth Industrial Revolution (IR 4.0), existing management models frequently overlook the specific governance and cyclical complexities required at the administrative meso-level, often focusing narrowly on high-level analytics. This study addresses this critical gap by developing a comprehensive Big Data Management (BDM) model specifically tailored for District Education Offices (PPD) in Malaysia. Adopting a cross-sectional quantitative research design involving 156 data managers, the proposed model was empirically tested and validated using Partial Least Squares Structural Equation Modelling (PLS-SEM). The results confirm a robust, sequential management lifecycle spanning across acquisition, representation, dissemination, and preservation stages, characterized by strong model explanatory power. Specifically, the findings demonstrate a high coefficient of determination ($R^2 > 0.5$; medium to large effect) and substantial predictive relevance ($Q^2 > 0$), confirming the structural stability of the framework. Unlike generic frameworks, this model integrates four core elements and 59 validated dimensions uniquely aligned with the regulatory standards of educational administration. By providing a standardized blueprint to enhance the District Education Office Excellence Evaluation (PKPPD), these findings offer a strategic roadmap for policymakers to institutionalize data-driven decision-making, mitigate administrative fragmentation, and optimize informational assets within the Malaysian and broader Southeast Asian public sectors.

Keywords: Big data management; District Education Office; PLS SEM Technique; Educational Administration; Educational Management

ABSTRAK

Tatkala penggunaan data raya semakin pesat dan berlaku peningkatan penggunaan dalam Revolusi Perindustrian 4.0, model pengurusan sedia ada sering kali mengabaikan aspek tadbir urus dan proses kitaran yang diperlukan di peringkat daerah dalam menguruskan data raya. Kecenderungan pengurusan data raya memberikan fokus kepada analisis data semata-mata. Kajian ini menangani jurang tersebut dengan membangunkan satu model Pengurusan Data Raya yang komprehensif dan khusus bagi kegunaan Pejabat Pendidikan Daerah (PPD) di Malaysia. Melalui penggunaan reka bentuk kuantitatif yang melibatkan 156 pengurus data, model ini telah disahkan melalui kaedah Partial Least Squares Structural Equation Modeling (PLS-SEM). Dapatan kajian menunjukkan kitaran pengurusan yang teguh merangkumi peringkat pemerolehan data, pemprosesan data dan jaminan kualiti data, keterangan dan perwakilan data, pemeliharaan data dan perkhidmatan repositori serta penyebaran data ($a \rightarrow b \rightarrow c \rightarrow d$) yang dicirikan oleh kuasa penjelasan model yang tinggi. Secara khususnya, ia menunjukkan nilai pekali penentuan yang tinggi $R^2 > 0.5$; kesan sederhana hingga besar) serta kuasa ramalan model yang sesuai ($Q^2 > 0$), sekali gus mengukuhkan kestabilan struktur rangka model tersebut. Model ini menyepadukan empat elemen teras dan 59 dimensi yang telah disahkan, selaras dengan piawaian kawal selia pentadbiran pendidikan. Dengan menyediakan pelan tindakan untuk memperkukuh Penarafan Kecemerlangan Pejabat Pendidikan Daerah (PKPPD), dapatan ini menawarkan hala tuju strategik buat penggubal dasar dalam menginstitusikan pembuatan keputusan berasaskan data, mengurangkan fragmentasi pentadbiran, serta mengoptimumkan aset maklumat dalam sektor awam Malaysia dan rantau Asia Tenggara secara amnya.

Kata kunci: Pengurusan data raya; Pejabat Pendidikan Daerah; Teknik PLS SEM; Pengurusan Pentadbiran, Pengurusan Pendidikan

INTRODUCTION

Since its emergence in 2016, the Fourth Industrial Revolution (IR 4.0) has catalysed a paradigm shift across global sectors, fundamentally altering service delivery through advanced technological integration (Zhong et al., 2017). In line with the UNESCO and SEAMEO regional report (2023), the digital evolution in Southeast Asia now demands technology not just as a supporting tool, but as an instrument that supports equity and efficiency of the overall system. There are several elements that underlie Industrial Revolution 4.0, including network objects (internet of things), cloud computing, augmented reality, simulation, robot automation, system integration, additive manufacturing, big data and cyber security (Bruno & Antonelli, 2018).

Now, various industries and fields have started to use the elements of Industrial Revolution 4.0 in improving the quality of services including the field of education (Ibrahim et al., 2024; Sukhodolov, 2018). The digital system technology that has been widely used and the full use of the internet in the field of education has caused massive data dumping (Bollier, 2010). The use of big data among educators to analyse information and make focused decisions has increased organizational efficiency (Shamim et al., 2019). It is also in line with the contemporary need of educator towards digital competencies and utilizing online resources and technological tools in education (Ma et al. (2024). Systematic big data management has reduced the operating costs of an organization indirectly as it can optimize resources (Kirmse, F., & Hoffmann, 2019; Poulovassilis, 2016).

BACKGROUND OF STUDY

The advent of the Fourth Industrial Revolution (IR 4.0) has fundamentally reshaped educational landscapes, where the massive influx of big data necessitates a shift toward more systematic and agile management frameworks. In the Malaysian context, this digital transformation was initially spearheaded by the National 4IR Policy and the Malaysia Education Blueprint 2013-2025, both of which emphasized the integration of big data analytics to enhance institutional quality. At the meso-level, the District Transformation Program (DTP 3.0), through the District Education Office Excellence Evaluation (PKPPD), identified big data management as a core competency for District Education Offices (PPD). However, while the potential of big data to provide evidence-based solutions is well-recognized (Kamal & Dave, 2019), its practical implementation at the district level remains a significant challenge. This urgency is further underscored by the transition towards the National Education Plan (Rancangan Pendidikan Malaysia) 2026–2035, which signifies a pivotal shift in the nation's educational governance. Under this new plan, the integration of an Educational Data Warehouse (EDW) and 'EMIS Pintar' are prioritized to ensure robust, data-driven decision-making across all administrative levels (Ministry of Education Malaysia, 2026).

PROBLEM STATEMENT

Recent studies have introduced various big data models; however, a critical gap persists in their application within the specific context of district education administration. Most existing frameworks are either overly focused on high-level data analytics for higher education (Fairuzullah et al., 2020) or technical school-based applications (Kamal & Dave, 2019), often neglecting the integrated governance and complete lifecycle processes required by district-level administrators. Current PKPPD reports reveal a persistent disparity in data management standards among PPDs, exacerbated by a lack of skilled personnel, inadequate infrastructure, and fragmented security protocols (Othman & Yasin, 2015; Bahagian Pengurusan Sekolah Harian, 2019).

Furthermore, while general models acknowledge data security, they often fail to provide a sequential, end-to-end management cycle from acquisition to dissemination that aligns with the unique regulatory and operational needs of the Malaysian public service. As underscored by UNICEF Innocenti (2025), effective data governance in education necessitates a paradigm shift from passive data collection to active stewardship that safeguards privacy while maximizing the public value of educational data. This study, therefore, addresses this insufficiency by developing and empirically validating a context-specific Big Data Management Model. Unlike generic models, this proposed framework integrates governance, preservation, and representation into a cohesive cycle, ensuring that data integrity is maintained from the initial collection point to its final strategic use.

AIM OF STUDY

This study was conducted for the purpose of developing a design model for PPD's big data management in Malaysia to be used as a guide for big data managers in PPD throughout Malaysia. To achieve this, the following specific objectives have been formulated:

1. To develop a big data management model for Malaysian District Education Offices (PPD) specific to the needs of district-level educational administration.
2. To validate the causal relationships and predictive relevance of the big data management model for PPD.

LITERATURE REVIEW

BIG DATA DYNAMICS IN EDUCATIONAL MANAGEMENT

Big Data in education is characterized by the massive accumulation of data generated through various digital systems and applications (Drigas & Leliopoulos, 2014; Eynon, 2013). Scholars conceptualize Big Data Management (BDM) as a strategic orchestration of raw informational assets, aimed at generating actionable insights for organizational governance (Prashant Gokul et al., 2019). While studies have shown BDM's success in areas such as ideological teaching (Yujia, 2020) and foreign language pedagogy (Lijuan, 2020), its broader application in providing integrated databases and future-ready strategies for educational sectors remains a critical frontier (Shan, 2020).

According to global trends highlighted in the OECD Digital Education Outlook (2026), big data management is rapidly evolving beyond mere storage, shifting its focus toward enhancing institutional agency and mitigating administrative burdens through automated governance frameworks. This global shift finds a critical parallel within the Malaysian education ecosystem, where the utilization of big data holds immense transformative potential. Specifically, by leveraging deep data insights, Malaysian educational institutions particularly at the district level can significantly strengthen institutional accountability and align local practices with these emerging international standards (Jin et al., 2022). This critical need for robust data governance is further compounded by emerging challenges in educational leadership and management (EDLM)

https://doi.org/10.17576/akad-2026-9602-11

research. As observed in a recent systematic assessment by Zhang et al. (2026), contemporary educational settings are increasingly overwhelmed by complex data structures that traditional administrative frameworks fail to process efficiently. Without a systematic Big Data Management (BDM) model, the efficacy of data-driven leadership is compromised, leading to structural fragmentation between data acquisition and strategic action. Therefore, establishing a contextualized management sequence is not merely a technical upgrade, but a necessary analytical prerequisite to support the causal-predictive goals of modern educational institutions (Zhang et al., 2026).

COMPARATIVE ANALYSIS OF EXISTING MODELS: A DIMENSIONAL PERSPECTIVE

A critical synthesis of existing BDM models reveals two distinct orientations: technical processing and university-centric governance.

Technical and Analytics Focus: Early models by Sharma et al. (2014) and Wassan (2015) were instrumental in establishing the technical layer, focusing on storage and visualization via tools like MapReduce and MongoDB. However, these models were primarily technical and lacked holistic governance. Vincent (2015a) and Fairuzullah et al. (2020) advanced this by introducing analytical taxonomies (descriptive, diagnostic, predictive, and prescriptive). While effective in predicting student success, these frameworks are heavily reliant on high-level staff expertise to manage complex NoSQL (MongoDB) or Power BI environments, which are often not present at the district administrative level.

Administrative and Lifecycle Governance: More recently, Yao (2022) introduced a more comprehensive lifecycle including governance and technology layers. However, a significant distinction remains: Yao's model is tailored for the vast, multi-sectoral environment of Higher Education Institutions. In contrast, District Education Offices (PPD) operate within a more standardized yet resource-constrained regulatory framework that necessitates a more streamlined, governance-heavy lifecycle. From a methodological standpoint, the limitations of these existing models often stem from their inability to capture the empirical complexities inherent in educational administration.

While higher education models prioritize academic learning analytics, administrative meso-levels like the PPD demand frameworks that can withstand rigorous statistical validation. According to Zhang et al. (2026), models in educational management must transition from purely descriptive orientations to robust causal-predictive cycles. By incorporating standardized data preservation and representation constructs, the BDM model overcomes the structural vulnerabilities of generic frameworks, offering a theoretically grounded and empirically testable lifecycle specifically optimized for public service parameters.

As shown in table 1.0, while existing models prioritize analytics, the proposed PPD model addresses the critical pre-analytical governance gaps, specifically in how data is represented and preserved within the public service hierarchy.

TABLE 1. Comparison between existing BDM models and the proposed PPD model

Model / Author	Context	Main Focus	Missing Element/Gap
Sharma et al. (2014) / Wassan (2015)	Technical / General	Technical processing, MapReduce, MongoDB.	Lacks administrative governance and security protocols.
Vincent (2015a) / Fairuzullah et al. (2020)	Higher Education (Uni)	Data analytics taxonomy (Predictive/Prescriptive).	Overly complex for district-level resources; requires high technical expertise.
Yao (2022)	Higher Education (Uni)	Data lifecycle and smart governance.	Broad sectoral focus; not tailored to PPD's specific regulatory environment.
Proposed Model (PPD-BDM)	District Education (PPD)	Integrated Lifecycle (Acquisition to Dissemination).	Addresses district-specific governance and universal data coding (PKPPD).

CONCEPTUAL FRAMEWORK

The theoretical foundation of this study is anchored in the Scientific Data Management Model by Crowston and Qin (2011). Although originally designed for scientific research datasets, this model provides a highly resilient four-stage architecture that is uniquely suited for the administrative data of PPDs. The choice of this framework is justified by the need for a Capability Maturity Model approach in PPDs, where data must be managed through a rigorous, standardized cycle rather than ad-hoc processing. The framework is structured into four sequential levels:

i. Data Acquisition, Processing, and Quality Assurance: Unlike generic models, this stage emphasizes filtering raw data at the entry point to ensure veracity.
ii. Description and Representation of Data: This is a critical addition for the PPD context. Given the complexity of teacher employment data, this stage ensures that all data classifications are coded uniformly across all PPDs in Malaysia, creating a universal language for district data.
iii. Data Dissemination: This stage facilitates the interaction between coded data and end-users through strategic applications, ensuring that data is not just stored, but accessible for decision-making.
iv. Data Preservation and Repository Services: Addressing the security concerns raised by Jee and Kim (2013), this stage ensures the long-term control and safety of processed data.

By integrating these four levels, the proposed framework extends Crowston and Qin's (2011) model into the domain of educational administration, bridging the gap between high-level technical processing and the practical governance needs of PPD data managers.

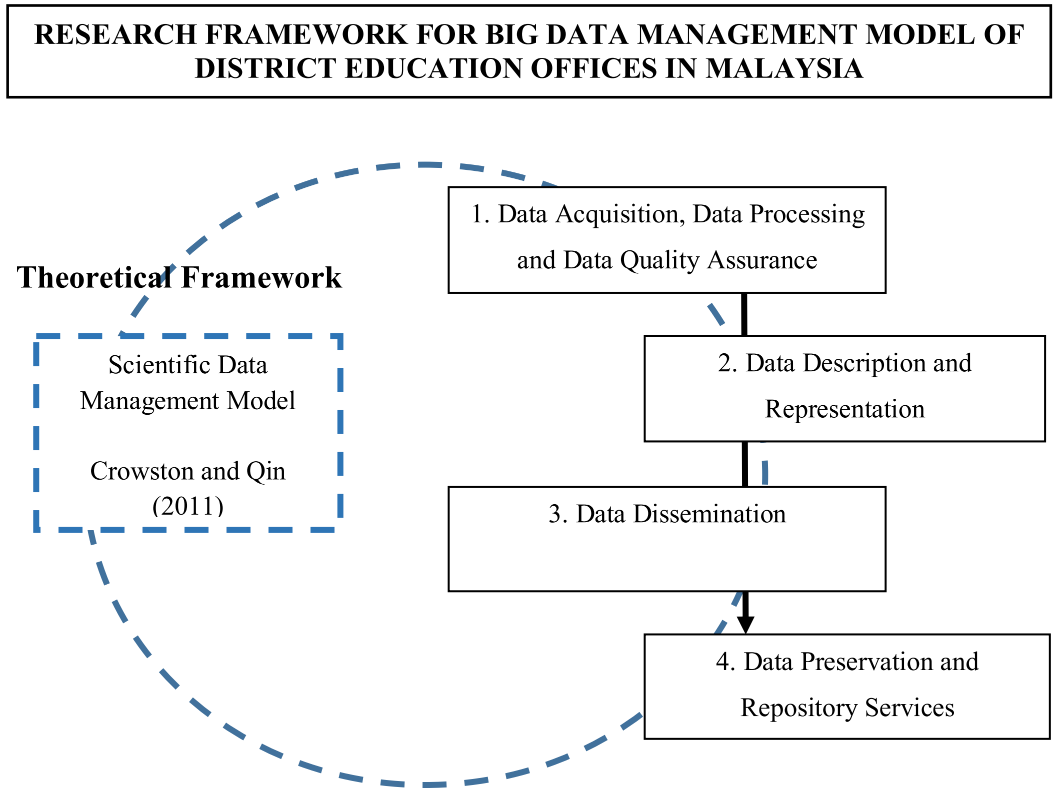

FIGURE 1. Research framework for big data management model of district education offices in Malaysia

METHODOLOGY

RESEARCH DESIGN

This study employs a cross-sectional quantitative research design, utilizing Partial Least Squares Structural Equation Modeling (PLS-SEM) for model development and validation. The selection of PLS-SEM as the primary analytical tool is justified by its suitability for theory development, particularly in adapting the Scientific Data Management Model by Crowston and Qin (2011) into a new administrative context. Furthermore, PLS-SEM is highly robust in handling complex multi-stage linear sequences without requiring strict distributional assumptions and is particularly effective for niche populations, such as the 156 data managers in this study, where high statistical power is required for smaller sample sizes (Hair et al., 2022). The adoption of PLS-SEM is further supported by Zhang et al. (2026), who recognize it as a dominant analytical instrument in educational management research due to its superior ability to establish causal-predictive evidence for complex structural models.

The research instrument was developed based on the four-stage data processing cycle—data acquisition, representation, dissemination and preservation with items adapted from established data governance literature. To ensure content validity, the initial item pool underwent a rigorous expert review process using Fuzzy Delphi Method (FDM) two cycled technique involving three subject matter specialists in educational management and information technology. A pilot study conducted with 30 respondents yielded exceptionally high Cronbach's Alpha values, ranging from .947 to .979. While these results indicate superior internal consistency, the researcher performed a redundancy check to ensure that the indicators remained distinct and captured unique dimensions of district-level management without excessive conceptual overlap.

TABLE 2. Reliability test results; cronbach's alpha

No	Item	Cronbach's Alpha Value
	Overall Dimension Value	
1	PPD Big Data Management Dimensions in Malaysia	.990
	Value by dimension	
2	Data acquisition, data processing and data quality assurance	.979
3	Description and representation of data	.97
4	Data preservation & repository services	.957
5	Data dissemination	.947

STUDY RESPONDENTS

The study utilized a total population sampling technique, targeting 156 public data managers across 140 District Education Offices (PPD) and 16 State Education Departments (JPN) in Malaysia. Data collection was carried out over a three-month period via an online platform, adhering strictly to ethical research procedures. Participation was voluntary, and an informed consent form was provided at the outset to guarantee anonymity and the confidentiality of responses. From the targeted population, 105 completed responses were received, representing a 67.3% response rate, which is deemed acceptable and sufficient for organizational-level research (Fincham, 2008).

TABLE 3. Respondent demographics

Item		No.	Percentage (%)	Total
Designation	Head Deputy Director	3	2.9	105
	Deputy Director	5	4.7	
	Deputy PPD	3	2.9	
	Assistant PPD	94	89.5	
Gender	Male	74	70.5	105
	Female	31	29.5	
Centre of Responsibility	JPN	16	15.2	105
	PPD	89	84.8	
Race	Malay	74	70.5	105
	Chinese	7	6.6	
	Others	24	22.9	
Age	26 – 35	2	1.9	105
	36 – 45	55	52.4	
	46 and above	48	45.7	
Academic Qualifications	PhD	3	2.9	105
	Master	31	29.5	
	Bachelor's Degree	71	67.6	
Experience (year)	5 and below	55	52.4	105
	6 – 10	20	19	
	11 – 15	15	14.3	
	16 and above	15	14.3	

https://doi.org/10.17576/akad-2026-9602-11

VALIDITY AND RELIABILITY

Data analysis followed a two-stage evaluation process involving the measurement model and the structural model. In the measurement model stage, convergent validity and reliability were assessed using Composite Reliability (CR) and Average Variance Extracted (AVE), with thresholds set at 0.70 and 0.50, respectively. During this refinement phase, 69 items were systematically removed to achieve discriminant validity and ensure a parsimonious model. This substantial reduction was necessary to meet the Heterotrait-Monotrait Ratio (HTMT) criterion of below 0.90, as many initial items exhibited cross-loadings between the acquisition and processing phases. This rigorous refinement ultimately resulted in a robust model consisting of 59 high-quality indicators, which provided a stable foundation for the subsequent structural path analysis.

RESULTS

PLS-SEM ANALYSIS

The assessment of the reflective measurement model was conducted to ensure the reliability and validity of the constructs before testing the structural relationships. Unlike traditional regression methods, PLS-SEM was selected for its suitability in model development and its robustness in prediction-oriented studies involving moderate sample sizes (Hair et al., 2022). Below is the code replacement for elements of big data management.

TABLE 4. Code replacement for constructs

Code	Constructs/Elements of Big Data Management
a	Data acquisition, data processing and data quality assurance
b	Description and representation of data,
c	Data preservation & repository services
d	Data dissemination

REFLECTIVE MEASUREMENT MODEL

To achieve a refined model, the researcher evaluated internal consistency through Cronbach’s Alpha and Composite Reliability (CR). As shown in Table 1.1, all constructs (a, b, c, and d) achieved CR values between 0.945 and 0.974, significantly exceeding the required threshold of 0.70. Convergent validity was confirmed as the Average Variance Extracted (AVE) for all dimensions ranged from 0.559 to 0.743, surpassing the 0.50 benchmark (table 1.3). Discriminant validity was rigorously tested using the Fornell-Larcker criterion and the Heterotrait-Monotrait (HTMT) ratio. All HTMT values remained ≤ 0.90, confirming that each stage of the big data lifecycle is empirically distinct. Notably, the initial instrument was refined by removing 69 indicators that exhibited low factor loadings (<0.60) or high cross-loadings, resulting in a final, parsimonious model of 59 high-quality items optimized for the PPD context.

TABLE 5. Reflective measurement model analysis summary

Construct	**Convergent Validity**		**Internal Consistency Reliability**			**Discriminant Validity**	
	Loadings/ Indicator Reliability	***AVE***	***Cronbach's Alpha***	***rho_A***	***Composite Reliability***	***Fornell-Larcker Criterion***	***HTMT***
	> 0.6	> 0.5	> 0.9	> 0.9	> 0.9	≤ 0.90	≤ 9.0
a	Yes (30 items)	0.559	0.972	0.975	0.974	a-a (0.748) b-a (0.736) c-a (0.744) d-a (0.728)	Yes
b	Yes (10 items)	0.743	0.962	0.962	0.967	b-b (0.862) c-b (0.825) d-b (0.861)	Yes
c	Yes (11 items)	0.679	0.952	0.955	0.959	c-c (0.824) d-c (0.821)	Yes
d	Yes (8 items)	0.682	0.933	0.944	0.945	d-d (0.826)	Yes

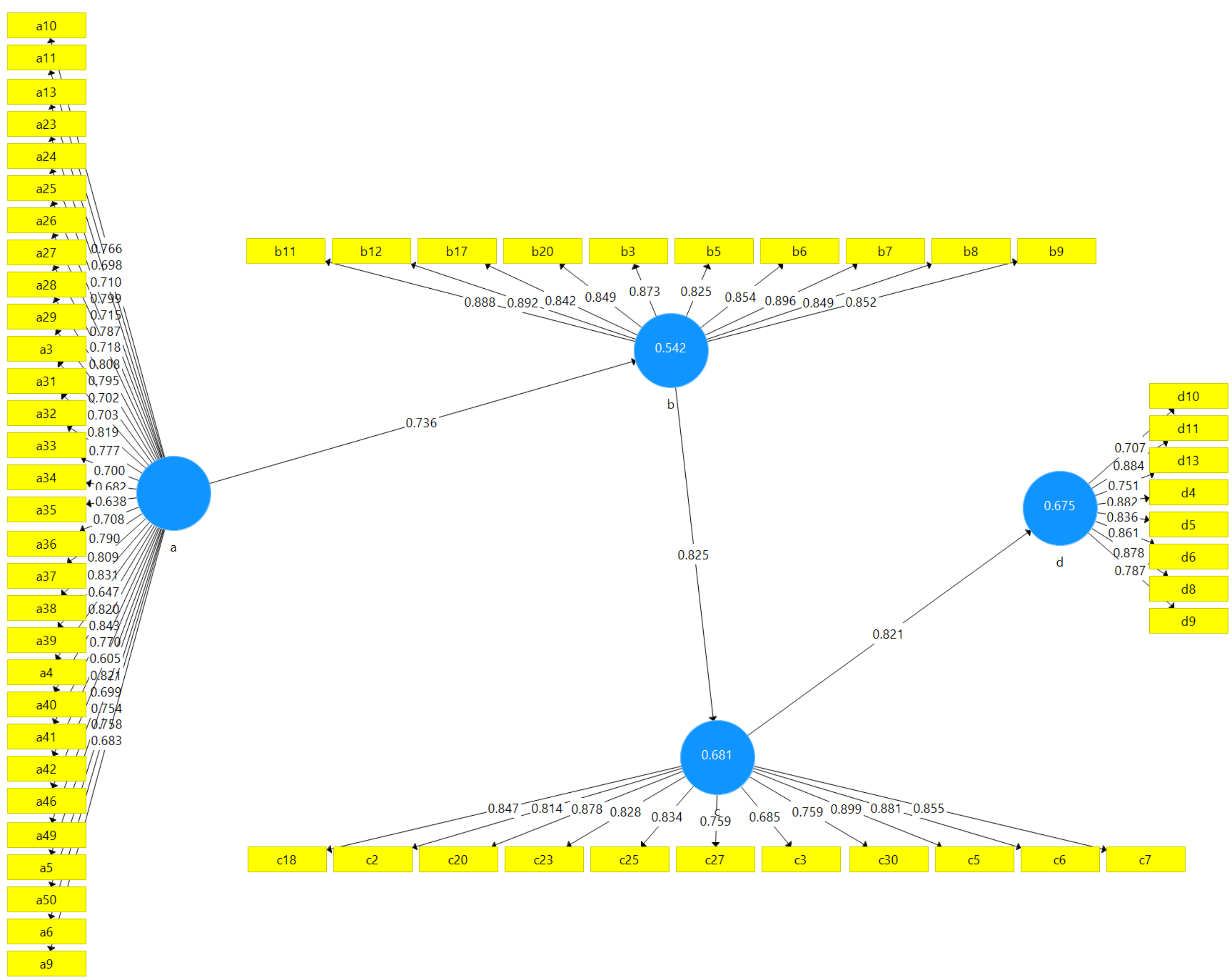

FIGURE 2. Reflective Measurement Model

STRUCTURAL MODEL

The structural model evaluation in PLS-SEM can identify the model's capability and model fit that has been developed (Hair et al., 2022; Hair, Ringle, & Sarstedt, 2011). There are five steps in the structural model evaluation procedure and the fifth step is a comparison of the model with previous models or models that are related to the model being studied (Hair et al., 2022). Model comparison does not meet the needs of the study because the model developed for this study is a new model and does not have any previous models that could be compared. Therefore, in order to complete the development of PPD big data management model design in Malaysia, the researcher will carry out the evaluation procedure from step one to step four only. The structural model was evaluated using standard PLS-SEM procedures including collinearity assessment (VIF), path coefficients, coefficient of determination (R^2), effect size (f^2), and predictive relevance (Q^2).

COLLINEARITY (VIF)

The threshold value for the accepted VIF is ≤ 3.3 and if the value is ≥ 5.0, it indicates that there is a potential collinearity problem with the indicator (Hair et al., 2022; Hair et al., 2011). The VIF values achieved a perfect score of 1.000, significantly below the conservative threshold of 3.0, indicating total absence of multi-collinearity issues.

TABLE 6. Variance inflated factor (VIF)

Construct	**Internal VIF Value**				**≤ 3.3 ?**
	a	**b**	**c**	**d**	
a		1.000			Yes
b			1.000		Yes
c				1.000	Yes

PATH COEFFICIENTS

The assessment of path coefficients confirms a strong and significant sequential dependency across the big data management lifecycle. Results indicate that the path from data acquisition (a) to data description and representation (b) is significantly positive ($t = 15.524$, $\beta = 0.736$, $p < 0.05$). Furthermore, data description and representation (b) exerts a substantial influence on data preservation and repository services (c) ($t = 16.582$, $\beta = 0.825$, $p < 0.05$), which in turn significantly predicts the effectiveness of data dissemination (d) ($t = 14.820$, $\beta = 0.821$, $p < 0.05$).

To address the theoretical rationale for utilizing SEM in this sequential framework, the model posits that these stages do not merely follow each other in time, but rather maintain a causal-linkage relationship. Within the PPD context, the accuracy of data representation (b) is a critical determinant of how effectively data can be preserved (c); without high-quality representation, preservation efforts become redundant. Similarly, the dissemination of data (d) is causally dependent on the integrity of the preservation and repository services (c). Therefore, the high path coefficients and t-statistics validate that the quality of each preceding phase is a causal driver for the successful outcome of the subsequent stage, justifying the use of SEM to evaluate the strength and stability of this integrated management cycle. This structure confirms that this model functions as a cohesive system where excellence in early-stage governance is mandatory for achieving overall organizational data dissemination.

TABLE 7. Structural model path coefficient evaluation results

Path Coefficient	**T Statistic**	**Critical *t* value**	***Statistically significant?***	***β, p values***	**Critical *P value***	***Statistically significant?***
a -> b	15.524	1.96	Yes	0.736, 0.000	0.05	Yes
b -> c	16.582	1.96	Yes	0.825, 0.000	0.05	Yes
c -> d	14.820	1.96	Yes	0.821, 0.000	0.05	Yes

MODEL'S EXPLANATORY POWER

In evaluating the explanatory power of a model, there are two values that can show the explanation of a model, namely the coefficient of determination value, R^2 and the effect size value, f^2. The value of the coefficient of determination, R^2 shows that changes in the construct can be explained by the indicators. Acceptable values for R^2 range from 0 to 1, however according to Hair et al. (2022) and Hair, Hult, Ringle, and Sarstedt (2014), the common rule in determining the level of R^2 value is as in the following table. The higher the R^2 value, the greater the explanatory power of the construct.

Based on the value of the coefficient of determination (R^2) b, 54.2% of the change (moderate) in construct b can be explained by the indicators that are under construct b. While for construct c, 68.1% of the change (moderate) in construct c can be explained by the indicators under construct c and 67.5% of the change (moderate) in construct d can be explained by the indicators under construct d. Therefore the R^2 value for the structural model is appropriate and acceptable.

TABLE 8. Results of coefficient of determination R^2

Construct	***R^2***	***R^2 Adjusted***	***Accepted R^2 Value***
b	0.542	0.537	Medium (0.5)
c	0.681	0.678	Medium (0.5)
d	0.675	0.672	Medium (0.5)

Next is the effect size value, f^2. Effect size assessment allows the researcher to observe the effect that occurs on each variable or construct. If there is a change in the R^2 value when a variable or construct is dropped or removed from the model, the effect size value will show the impact. According to Cohen (1988), there are three levels that describe the effect size value as follows;

TABLE 9. Effect size range f^2

Value	≥ 0.02	≥ 0.15	≥ 0.35
Level	Small	Medium	Large

The effect size value shows the effect of the construct in the PPD big data management model in Malaysia. All three effect size values obtained show a large effect size. This shows that dropping or discarding variables or constructs a, b and c will have a significant impact on the value of the coefficient of determination R^2.

https://doi.org/10.17576/akad-2026-9602-11

TABLE 10. Effect size results (f^2)

Construct	a	b	c	d
a		1.183 (large)		
b			2.138 (large)	
c				2.075 (large)
d				

MODEL'S PREDICTIVE POWER

Predictive relevance or prediction accuracy is also called Stone Geisser's Q^2 (Stone, 1974; Geisser, 1974). It is a test to measure the developed model has a good predictive ability or vice versa. If the Q^2 value exceeds > 0, it indicates that the values have been well restructured and the developed model has predictive ability. In order to obtain the Q^2 value, the blindfolding analysis method must be performed.

The prediction accuracy results of the blindfolding analysis show that all values are > 0 and it can be concluded that the developed big data management model design has values that have been rearranged well and this model has the ability to predict, so the Q^2 value shows that the structural model is suitable and acceptable.

TABLE 11. Results of blindfolding prediction accuracy Q^2

Construct	Q^2	> 0 ?
a		
b	0.398	Yes
c	0.458	Yes
d	0.448	Yes

Table 12. Summary of structural model evaluation results

VIF		**Path Coefficients**				R^2	f^2	Q^2
≤ 0.3		**T Value**	***Statistically Significant?***	**β, *p value***	***Statistically Significant?***	**> 0.5. medium**	**≥ 0.35, large**	**> 0**
		> 1.96		**Positive, < 0.05**				
a-b, 1.000	Yes	15.524	Yes	0.736, 0.000	Yes	b, 54.2%	Large effect size	Yes
b-c, 1.000	Yes	16.582	Yes	0.825, 0.000	Yes	c, 68.1%		Yes
c-d, 1.000	Yes	14.820	Yes	0.821, 0.000	Yes	d, 67.5%		Yes
fit		*fit*				*fit*		*fit*

The structural model evaluation results as shown in the table above show that the PPD big data management model in Malaysia is suitable because it does not have collinearity problems with the indicators, obtains a significant positive relationship between the constructs, has a moderate level of explanation by each construct and has a good forecasting ability. Therefore it can be concluded that the developed model is appropriate (fit).

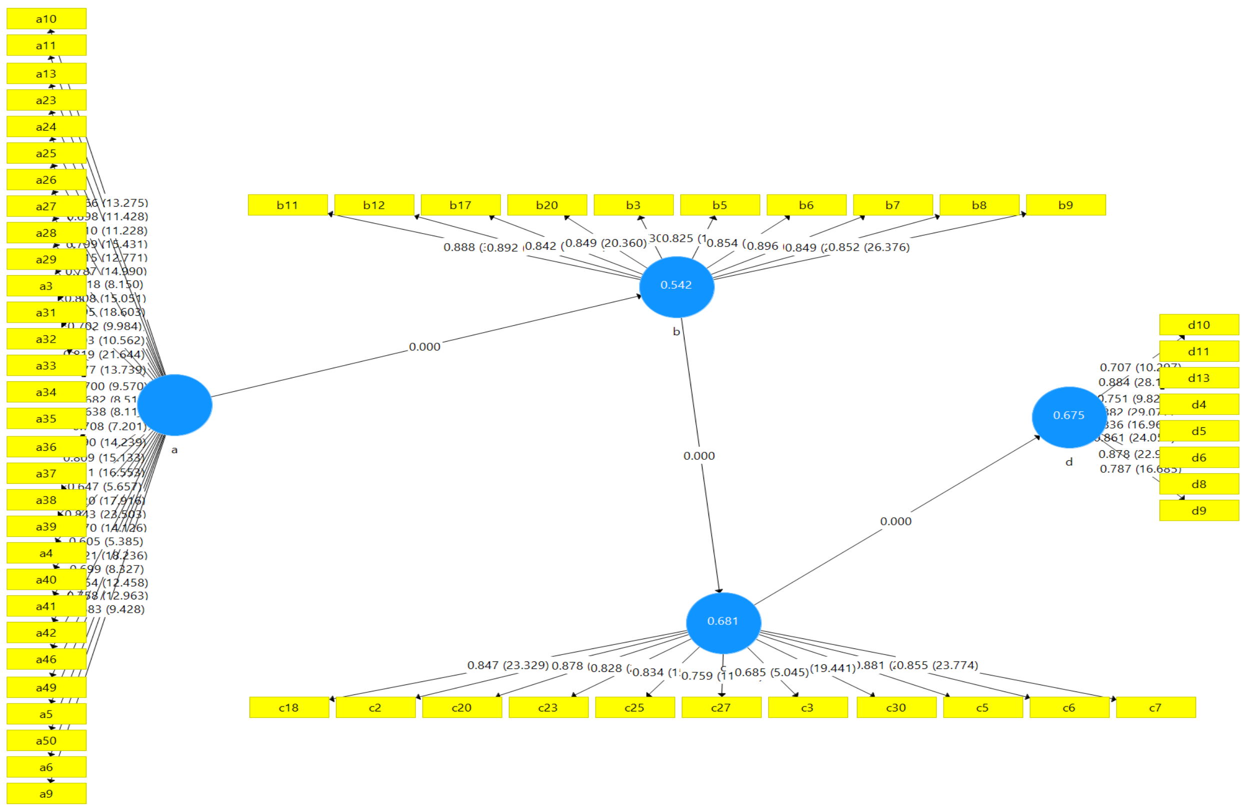

FIGURE 3. Structural model

TABLE 13. PLS SEM analysis summary

No	Element and Dimension	Amt. Original Items	Amt. Rejected Items	Amt. Retained Items
1	Data acquisition, data processing and data quality assurance			
	1.1 Data acquisition	20	12	**8**
	1.2 Security and Confidentiality of Data Acquisition	15	3	**12**
	1.3 Data Processing and Data Quality Assurance	25	15	**10**
2	Description and representation of data,	24	14	**10**
3	Data preservation & repository services	30	19	**11**
4	Data dissemination	14	6	**8**
	Total	**128**	**69**	**59**

The refinement of the research instrument, which resulted in the removal of 69 indicators, was a systematic necessity to ensure the structural integrity and parsimony of this model. Following established PLS-SEM protocols, items were eliminated based on two critical statistical criteria: outer loadings falling below the 0.60 threshold to guarantee indicator reliability, and the presence of high cross-loadings that compromised discriminant validity between the sequential stages of the data lifecycle. By distilling the original 128 items down to 59 high-quality indicators, the model transitioned from a broad conceptual pool to a fit-for-purpose administrative tool that eliminates redundant noise, thereby ensuring that the constructs of acquisition, representation, preservation, and dissemination remain empirically distinct and practically actionable for district-level governance.

DISCUSSION

The development of the PPD Big Data Management (PPD-BDM) model establishes a robust four-stage lifecycle; acquisition, representation, preservation and dissemination, which aligns with global architectural frameworks proposed by Faroukhi et al. (2020) and Sultana et al. (2021). However, this study uniquely extends these generic models by integrating specific security and confidentiality dimensions (Siddiqa et al., 2016; Nor Kamaliah et al., 2022) tailored specifically for the Malaysian public sector's centralized education hierarchy. The exceptionally high path coefficients and predictive relevance (Q^2) observed in this study suggest that in a highly regulated administrative environment such as the District Education Office (PPD), data management is not merely a series of technical steps but a causal chain of governance. For instance, the significant relationship between data representation and preservation underlines that standardized coding serves as the primary driver for long-term data integrity within the Malaysian context.

This emphasis on standardized data sequences is particularly crucial when considering internal customer satisfaction within the public sector. As highlighted by Ibrahim et al. (2024), service quality in Malaysian public organizations is heavily dependent on the efficiency of internal processes, especially during periods of administrative transition. By institutionalizing the PPD-BDM model, district offices can minimize data-related bottlenecks, thereby ensuring that the internal customers; namely school administrators and teachers, receive accurate and timely informational support. This efficiency directly correlates with higher organizational satisfaction, as a streamlined data lifecycle reduces the ambiguity often associated with manual or ad-hoc data handling in centralized bureaucracies.

Furthermore, the PPD-BDM model significantly advances existing data governance frameworks by bridging the critical gap between high-level policy and grassroots implementation. This alignment is vital, as technology-driven reforms are most effective when they resolve the disparity between macro-level digital policies and the practical readiness of local administrative units, a challenge notably observed in the Indonesian educational reform context (Wang et al., 2023). Unlike previous models that focused predominantly on higher education (Yao, 2022) or purely technical processing (Sharma et al., 2014), the PPD-BDM model prioritizes the veracity of administrative data through a validated sequence of processing.

This administrative readiness is also intrinsically linked to the digital competence of the workforce. While the PPD-BDM model provides the structural framework, its implementation success relies on the synergy between system design and personnel capability. Ma et al. (2024) argue that digital competence and targeted training are the cornerstones of effective technology adoption in the Malaysian educational landscape. In the context of PPDs, the 59 validated dimensions of this model serve as a roadmap for identifying specific competency gaps among data managers. Without such a structured management model, even advanced digital tools remain underutilized, as the interconnection between technical training and practical data governance must be explicitly defined to ensure long-term sustainability in remote and digital administrative operations.

Consequently, this study provides a strategic blueprint for Evidence-Based Decision Making (EBDM), enabling district officers to transition from ad-hoc data handling to strategic analytics. This shift directly enhances the precision of teacher placements, student performance tracking, and resource allocation which indirectly supporting the initiative for good governance practice for sustainable development generally in the public sector services in Malaysia (Zabidi et al., 2024). The impact of such analytics is maximized when it is deeply embedded into the administrative areas, facilitating proactive intervention rather than mere reactive reporting, as evidenced by successful initiatives in Singapore (Lee et al., 2023). In a broader context, these findings contribute to the ongoing discourse on digital transformation within Southeast Asian public administrations, where centralized systems often struggle with data silos and inconsistent reporting. The high predictive power of the PPD-BDM model aligns seamlessly with the 2026 digital continuity priorities in the region, which advocate for real-time audit readiness. As suggested by Creatrix Campus (2026), educational offices must transition towards live data governance to maintain institutional continuity in an increasingly digitalized public sector.

The integration of these elements ultimately fosters a culture of integrity and accountability. Zabidi et al. (2024) emphasize that good governance practices are the primary drivers for sustainable development within Malaysia's public sector services. The PPD-BDM model operationalizes these governance principles by providing a transparent and auditable data processing sequence. By aligning district-level data management with national sustainability goals, the model ensures that public resources are managed with high precision, fulfilling the mandate for excellent public service delivery. Ultimately, this ensures that the PPD-BDM model remains not only theoretically sound but also practically indispensable for future-ready educational administration in Malaysia and beyond.

CONCLUSION

This study successfully establishes the first specialized Big Data Management Model for District Education Offices in Malaysia using PLS-SEM validation. By synthesizing the experiences and requirements of data managers across the PPD, JPN, and KPM levels, the resulting model offers a high-fidelity representation of the actual administrative needs in the Malaysian education ecosystem. The theoretical contribution of this research lies in the successful adaptation of the Scientific Data Management Model (Crowston & Qin, 2011) into a functional administrative framework, proving that scientific data rigor is both applicable and necessary for public sector governance.

Practically, this model serves as a vital instrument for the District Education Office Excellence Assessment (PKPPD). It provides a standardized rubric for auditing data maturity, enabling the Ministry of Education to benchmark district-level digital readiness accurately. While this study is limited by its focus on the Malaysian Ministry of Education's internal centralized structure, making direct generalization to decentralized or private sectors difficult, it sets a foundational benchmark for public sector data governance. Future research should explore the integration of Artificial Intelligence (AI) within this four-stage model to automate the representation phase, further enhancing the speed and accuracy of educational analytics across the region.

Beyond its technical and administrative utility, the PPD-BDM model represents a transformative shift toward a more resilient and transparent educational governance culture. By bridging the gap between digital competence and institutional accountability, the model ensures that data is no longer treated as a static administrative byproduct, but as a dynamic strategic asset. This alignment is critical in fostering a sustainable digital ecosystem where district-level administrators are empowered to make evidence-based decisions that are both timely and ethically sound. Ultimately, the institutionalization of this framework not only strengthens the Malaysian public service delivery but also serves as a compelling archetype for other emerging economies in the region striving to navigate the complexities of the Fourth Industrial Revolution while maintaining high standards of organizational integrity.

ACKNOWLEDGEMENT

The author would like to express the deepest gratitude to the Ministry of Education Malaysia (MOE), particularly the District Education Offices (PPD) involved, for their cooperation and support in facilitating the data collection process. Special thanks are also extended to the data managers who participated in this study for their invaluable insights. Finally, I would like to thank my supervisors and colleagues for their continuous guidance and intellectual support throughout the development of this research.

REFERENCES

Bahagian Pengurusan Sekolah Harian, K. (2019). *Penilaian instrumen PKPPD.*

Bollier, D. (2010). *The Promise and Peril of Big Data*. Washington, DC: The Aspen Institute.

Bruno, G., & Antonelli, D. (2018). Ontology-based platform for sharing knowledge on industry 4.0. *Paper presented at the International Conference on Product Lifecycle Management, PLM 2018*, Turin; Italy.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Creatrix Campus (2026). D*igital continuity and institutional data governance in Southeast Asian public sectors: A 2026 perspective.* Creatrix Research Reports. https://www.creatrixcampus.com/blog/ai-competency-based-education-higher-ed

Crowston, K., & J., Q. (2011). A Capability Maturity Model for Scientific Data Management: Evidence from the Literature. *Proceedings of the ASIST Annual Meeting, 48*. doi:10.1002/meet.2011.14504801036

Drigas, A. S., & Leliopoulos, P. (2014). The Use of Big Data in Education. *International Journal of Computer Science IJCSI, 11*(5).

Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research? *Learning, Media and Technology, 38*(3), 237-240. doi:10.1080/17439884.2013.771783

Fairuzullah, A., Mohd Helmy, A. W., Noraziah, A., & Zarina, M. (2020). A Distributed Big Data Model for Education Sector. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(1), 414 - 417.

Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). An Adaptable Big Data Value Chain Framework for End-to-End Big Data Monetization. *Big Data and Cognitive Computing*, 4(4), 34. Retrieved from https://www.mdpi.com/2504-2289/4/4/34

Fincham, J. E. (2008). Response Rates and Responsiveness for Surveys, Standards, and the Journal. *American Journal of Pharmaceutical Education, 72*(43).

Global Education Monitoring Report Team & SEAMEO. (2023). *2023 Southeast Asia Report: Technology in Education – A Tool on Whose Terms?* https://unesdoc.unesco.org/ark:/48223/pf0000387214

Hair, Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–151. doi:10.2753/MTP1069-6679190202

Hair, Hult, G. T., Ringle, C. M., & Sarstedt, M. (2014). *A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM)*. United States of America: SAGE Publications, Inc.

Hair, Hult, G. T., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM)* (Third ed.). United States of America: SAGE Publications, Inc.

Ibrahim, N. H., Makhbul, Z. K., Kamarul, C. W., Ahmad, H. C., & Dzulkarnain, S. N. Z. H., (2024). Service Quality and Internal Customer Satisfaction Amidst the Pandemic. *Akademika*, 94(1), 2024: 53-66. https://doi.org/10.17576/akad-2024-9401-05

Jin, S. J., Abdullah, A. H., Mokhtar, M., & Abdul Kohar, U. H. (2022). The Potential of Big Data Application in Mathematics Education in Malaysia. *Sustainability,* 14(21), 13725. https://doi.org/10.3390/su142113725

Kamal, J., & Dave, M. (2019). A Framework for Managing and Analyzing Big Data in Indian School Education System with Reference to Jammu & Kashmir. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(9s).

Kementerian Pendidikan Malaysia (2026). *Rancangan Pendidikan Malaysia 2026-2035: Memacu Kecemerlangan Pendidikan Digital.* Putrajaya: Bahagian Perancangan dan Penyelidikan Dasar Pendidikan.

Kirmse, A., F., K., & Hoffmann, M. (2019). Industrial Big Data: From Data to Information to Actions. *Paper presented at the 4th International Conference on Internet of Things, Big Data and Security, Germany*.

Lee, A. V. Y., Koh, E., & Looi, C. K. (2023). *AI in Education and Learning Analytics in Singapore: An Overview of Key Projects and Initiatives. Information and Technology in Education and Learning.* https://www.jstage.jst.go.jp/article/itel/3/1/3_3.1.Inv.p001/_pdf/-char/en

Malaysia, K. P. (2020). *Panduan Pengurusan Program Transformasi Daerah Edisi 3.0.* . https://www.moe.gov.my/en/pemberitahuan/announcement/panduan-pengurusan-program-transformasi-daerah-dtp-3-0

OECD (2026), OECD Digital Education Outlook 2026: *Exploring Effective Uses of Generative AI in Education*, OECD Publishing, Paris, https://doi.org/10.1787/062a7394-en.

Othman, M. K., & Yasin, N. M. (2015). Essential Elements for the Successful Implementation of Public Information and Service Delivery System (e-Service) for the District Education Office in Malaysia. *International Arab Journal of e-Technology, 4*, 67-77.

PrashantGokul, K., Sundararajan, M., & Prantosh, K. P. (2019). Big Data Management, Data Science and Data Analytics: What is it and Where— An Educational in Indian Perspective. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(12).

Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information and Management, 56*(6). doi:10.1016/j.im.2018.12.003

Shan, W. (2020). Smart Education — The Necessity And Prospect Of Big Data Mining And Artificial Intelligence Technology In Art Education. *Journal of Physics: Conference Series, 1648*(4). doi:10.1088/1742-6596/1648/4/042060

Sharma, S., Tim, U., Wong, J., Gadia, S., & Sharma, S. (2014). A Brief Review on Leading Big Data Models. *Data Science Journal*. doi:10.2481/dsj.14-041

Siddiqa, A., Hashem, I., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A Survey of Big Data Management: Taxonomy and State-of-the-Art. *Journal of Network and Computer Applications,* 71. doi:10.1016/j.jnca.2016.04.008

Sukhodolov, Y. A. (2018). The Notion Essence and Peculiarities of Industry 4.0 as a sphere of industry. In Y. V. R. Elena G. Popkova, Aleksei V. Bogoviz (Ed.), *Industry 4.0: Industrial Revolution of the 21st Century* (Vol. 169, pp. 3-10). Moscow, Russia: Springer.

Sultana, S., Akter, S., Kyriazis, E., & Wamba, S. F. (2021). Architecting and developing big data-driven innovation (ddi) in the digital economy. *Journal of Global Information Management, 29*(3), 165-187. doi:10.4018/JGIM.2021050107

UNESCO. (2021). *Distance learning solutions.* https://en.unesco.org/covid19/educationresponse/solutions

UNICEF Innocenti. (2025). *Data Governance for EdTech.* https://www.unicef.org/innocenti/media/11611/file/UNICEF-Innocenti-Data-Governance-Education-Technology-2025.pdf.pdf

Vincent, K. O. (2015a). Big Data and its Research Implications for Higher Education: Cases from UK Higher Education Institutions. *Paper presented at the IIAI 4th International Congress on Advanced Applied Informatics.*

Vincent, K. O. (2015b). Big Data and its Research Implications for Higher Education: Cases from UK Higher Education Institutions. *4th International Congress on Advanced Applied Informatics*.

Wang, C. et al. (2023). *Technology-Driven Education Reform in Indonesia.* https://repositori.kemendikdasmen.go.id/30538/1/Indonesias-K-12-Education-QualityImprovement-English-05122023.pdf

Wassan, J. T. (2015). Discovering Big Data Modelling for Educational World. *Procedia - Social and Behavioral Sciences,* 176.

Ma, W., Mat, A., Rahman, G. Y., & Jaafar, R. (2024). Exploring the Interconnections: Digital Competence, Teacher Training and Remote Teaching. *Akademika,* 94(3), 2024: 108-125. https://doi.org/10.17576/akad-2024-9403-07108

Yao, Y. (2022). Design of University Data Governance Process System Under the Big Data Environment. *Paper presented at the Lecture Notes on Data Engineering and Communications Technologies.*

Yujia, Z. (2020). Fusion Development of Ideological and Political Teaching with Information Technology in the Big Data Era. *Journal of Physics: Conference Series, 1533*(4). doi:10.1088/1742-6596/1533/4/042013

Zabidi, N., Ismail, M.M., Adnan, Z.H., & Zain, M.I. (2024). Good Governance Practices for Sustainable Development in the Public Sector Services in Malaysia. *Akademika*, 94(2), 2024: 252-266. https://doi.org/10.17576/akad-2024-9402-14

Zhang, Y., Adams, D., Cheah, J.-H., Teeroovengadum, V., & Ramayah, T. (2026). An assessment of the use of Partial Least Squares Structural Equation Modelling (PLS-SEM) in educational leadership and management research. *European Journal of Education*, 61, e70600. https://doi.org/10.1111/ejed.70600

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Elsevier Engineering, 3*, 616-630. doi:http://dx.doi.org/10.1016/J.ENG.2017.05.015

https://doi.org/10.17576/akad-2026-9602-11

BIODATA OF AUTHORS:

Nor Kamaliah Mohamad
Faculty of Education,
Universiti Malaya
Email: kuya2307@gmail.com

Norfariza Mohd Radzi (Corresponding author)
Faculty of Education,
Universiti Malaya
Email: norfariza@um.edu.my

Zuraidah Abdullah
Faculty of Education,
Universiti Malaya
Email: zuraidahab@um.edu.my