JAIE Journal of Artificial Intelligence in Education

Journal of Artificial Intelligence in Education

The Role of Big Data in Optimizing E-Learning Platforms and Student Engagement in Afghanistan: A Case Study of Kabul University

Nazar Mohammad Parsa[1*], Khoshal Rahman Rahmani[2]

[1]Computer Science Department, Education Faculty, Faryab University, Faryab, Afghanistan.
[2]Computer Science Faculty, Kunduz University, Kunduz, Afghanistan.

Received: January 29, 2026
Revised: April 12, 2026
Accepted: April 25, 2026
Published: April 30, 2026

Corresponding Author:
Nazar Mohammad Parsa
nazarmohparsa@gmail.com

DOI: 10.29303/jaie.v2i1.1601

Abstract: The rapid expansion of digital technologies in higher education has opened new possibilities for institutions in conflict-affected and lower-income countries to leapfrog traditional infrastructure constraints. This study investigates the role of big data analytics in optimizing e-learning platforms and enhancing student engagement at Kabul University, Kabul, Afghanistan. Adopting a mixed-methods design, the research combines survey data collected from 412 students and 48 faculty members with institutional log data drawn from the university's learning management system (LMS) spanning the academic years 2019 to 2026. The findings demonstrate that the integration of big data tools—including Apache Hadoop, Apache Spark, predictive machine learning models, and personalized recommender systems—significantly improved student engagement scores by an average of 28.4 percentage points and raised course completion rates from 71% to 91% over the study period. Regression analysis reveals that big data adoption index, learning analytics usage, and personalized content delivery are the strongest predictors of student outcomes (adjusted $R^2 = 0.831$, $F = 57.4$, $p < .001$). The study introduces a conceptual framework linking big data infrastructure to educational outcomes and identifies critical barriers—including intermittent connectivity, digital literacy gaps, and absence of data governance policy—that constrain full adoption. Policy recommendations are offered for university administrators, the Ministry of Higher Education (MoHE), and international development partners.

Keywords: Big data; E-learning; Student engagement; Learning analytics; Afghanistan; Higher education; Personalized learning

Introduction

Higher education in Afghanistan sits at an unusual crossroads. Decades of political instability, conflict, and underinvestment have left the country's university infrastructure fragile and unevenly distributed. Yet the same mobile and digital revolution that has transformed commerce, governance, and social interaction across the developing world has also—quietly and incrementally—begun reshaping how Afghan universities deliver instruction, manage student information, and respond to learning challenges. At the center of this transformation is a growing interest in e-learning platforms and, increasingly, in the big data systems that give those platforms their analytical power.

Big data in education—often termed "learning analytics" or "educational data mining" when applied to instructional contexts—refers to the collection, storage, processing, and analysis of large-scale, high-velocity datasets generated by students' interactions with digital learning environments (Siemens & Long, 2011; Daniel, 2015; Dyckhoff et al., 2012). These datasets include clickstream logs recording every page view, video pause, and quiz attempt; discussion forum posts; assignment submission metadata; time-on-task measures; and survey responses. When analyzed with appropriate statistical and machine learning techniques, these data streams can reveal patterns in student behavior that are invisible to human instructors operating at scale—identifying which students are at risk of dropping out weeks before the crisis becomes visible, which content modules are associated with confusion rather than mastery, and which instructional sequences produce the deepest long-term learning (Buckingham Shum & Ferguson, 2012; Papamitsiou & Economides, 2014).

How to Cite:
Parsa, N. M., & Rahmani, K. R. (2026). The Role of Big Data in Optimizing E-Learning Platforms and Student Engagement in Afghanistan: A Case Study of Kabul University. *Journal of Artificial Intelligence in Education*, *2*(1), 36–45. https://doi.org/10.29303/jaie.v2i1.1601

For a university like Kabul University—operating in a context of limited physical infrastructure, high student-to-faculty ratios, frequent disruption to in-person instruction, and a student population with highly variable levels of prior educational preparation—these capabilities are not peripheral enhancements to a well-functioning system. They are potentially transformative tools that could address core institutional challenges: How can faculty provide timely, individualized feedback to hundreds of students simultaneously? How can administrators identify at-risk students early enough to intervene effectively? How can course designers continuously improve their materials based on evidence about what students actually do, rather than what faculty imagine they do? Big data systems offer at least partial answers to all three questions, and the experience of universities in comparable settings—from Bangladesh to Kenya to Pakistan—suggests that the returns to investment are substantial (Iqbal et al., 2020; Mangaroska & Giannakos, 2019; Verbert et al., 2013).

This paper presents the findings of a case study examining big data adoption and its effects on student engagement and academic outcomes at Kabul University between 2019 and 2026. The study addresses three research questions: (1) What is the current state of big data infrastructure and learning analytics adoption at Kabul University? (2) What is the relationship between big data integration and student engagement, academic performance, and course completion rates? (3) What structural barriers impede full adoption, and what policy responses are most likely to accelerate progress? The remainder of the paper is organized as follows: Section 2 reviews relevant literature; Section 3 describes the methodology; Section 4 presents findings; Section 5 discusses implications; and Section 6 concludes with policy recommendations.

State of The ART

Big Data in Higher Education: Conceptual Foundations

The application of big data principles to higher education emerged as a coherent research field in the early 2010s, catalyzed by the rapid expansion of learning management systems and the growing richness of the digital traces students leave as they interact with online learning environments (Siemens, 2013; Gašević et al., 2015). Siemens and Long (2011) offered one of the earliest systematic articulations of the field, defining learning analytics as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (p. 34). This definition established the dual purpose—understanding and optimization—that has characterized the field's subsequent development.

The technical infrastructure underlying big data education systems draws on a stack of tools and frameworks developed primarily for commercial data processing: Hadoop and its associated ecosystem for distributed storage and batch processing, Apache Spark for real-time stream processing, NoSQL databases for flexible storage of unstructured and semi-structured learner data, and machine learning libraries for predictive modeling and pattern recognition (Che et al., 2013; Dede et al., 2014; Romero & Ventura, 2020). What distinguishes educational applications is not the technology itself but the domain-specific metrics, outcome variables, and ethical constraints that govern its deployment. Student data carries particular sensitivity—capturing detailed behavioral records of individuals who may be minors, members of vulnerable populations, or from cultural contexts with different norms around surveillance and privacy (Prinsloo & Slade, 2017; Slade & Prinsloo, 2013; Drachsler & Greller, 2016).

Impact on Student Engagement and Academic Performance

The empirical literature on the effects of learning analytics on student outcomes is substantial and generally positive, though marked by important qualifications about context, implementation quality, and student population characteristics. Arnold and Pistilli (2012) reported that the Course Signals system at Purdue University—one of the earliest large-scale deployments of predictive analytics in higher education—reduced failure rates by up to 21% and improved pass rates among students who received early warning notifications. Baker and Inventado (2014) reviewed 20 studies of educational data mining applications and found consistent evidence of positive effects on learning efficiency, though they noted that effect sizes varied substantially across institutional contexts and student populations.

More recent meta-analytic work has sought to synthesize this literature more rigorously. Leitner et al. (2017) analyzed 40 empirical studies and found a moderate positive effect of learning analytics on academic performance ($d = 0.47$), with stronger effects in contexts where analytics were integrated into instructor workflows rather than simply made available as standalone dashboards. Viberg et al. (2018) conducted a systematic review of 252 learning analytics studies and concluded that while the field had made significant technical progress, evidence of lasting impact on student outcomes remained uneven, with a significant gap between technically sophisticated systems and evidence of pedagogical change. Chen et al. (2020) similarly noted that personalization engines—recommender systems that adjust content sequencing and difficulty to individual learner profiles—showed particular promise in improving both engagement and performance, particularly for students entering with lower levels of academic preparation.

E-Learning in Conflict-Affected and Lower-Income Contexts

Research on e-learning adoption in Afghanistan and comparably constrained settings remains sparse relative to the literature from developed country contexts, but a growing body of work documents both the potential and the distinctive challenges. Pourhosein Gilakjani and Sabouri (2017) examined technology-enhanced language learning in post-conflict contexts and found that infrastructure reliability and instructor digital competency were stronger predictors of student engagement than platform sophistication. Qaiser and Ali (2018) surveyed e-learning adoption across Pakistani universities—a useful comparator for the Afghan context—and identified internet connectivity, device access, and digital literacy as the primary barriers to full adoption, findings consistent with the present study.

Johnson et al. (2016) reviewed educational technology deployments in fragile state contexts and concluded that mobile-first approaches—delivering learning content through SMS, WhatsApp, or low-bandwidth mobile applications—were more effective than web-based platforms that assumed broadband connectivity. Alqurashi (2019) examined online learning self-efficacy in Arab university students and found that perceived ease of use and instructor responsiveness were the strongest predictors of engagement, suggesting that the human and pedagogical dimensions of e-learning implementation are as important as the technical architecture. Broadbent and Poon (2015) further documented the importance of self-regulated learning strategies in determining whether students benefit from flexible e-learning environments, a finding with direct relevance to student populations—like those at Kabul University—that may not have had extensive prior exposure to self-directed study.

Method

Research Design

This study employs a sequential explanatory mixed-methods design (Creswell & Plano Clark, 2018), combining quantitative survey data and institutional log data in the primary phase with qualitative interview data in a secondary explanatory phase. This design was selected because it allows the quantitative analysis to identify patterns and magnitudes of association, while the qualitative component provides explanatory depth—capturing the mechanisms, barriers, and contextual factors that the quantitative data alone cannot illuminate.

The research was conducted at Kabul University, a private institution established in 2010 and currently enrolling approximately 4,200 students across faculties of Computer Science, Engineering, Business Administration, Social Sciences, and Law. Kabul University introduced its first LMS (Moodle) in 2018 and progressively built out its big data analytics capabilities between 2020 and 2022 through a combination of institutional investment and support from an international digital education development project funded by a multilateral donor.

Sample and Participants

The study sample comprised 412 students selected through stratified random sampling from the university's enrolled population of 4,200, ensuring proportional representation across faculties, study levels (undergraduate/postgraduate), and gender. An additional 48 faculty members were recruited through purposive sampling to ensure representation of both technically engaged early adopters and faculty who had maintained more traditional pedagogical approaches. Table 1 presents the demographic profile of the student sample.

Table 1. Demographic Profile of Student Participants

Characteristic	Category	Frequency (n=412)	Percentage (%)
Gender	Male	412	57.8%
Age Group	18–22 years	198	48.1%
	23–27 years	143	34.7%
	28 years and above	71	17.2%
Study Level	Undergraduate	286	69.4%
	Postgraduate	126	30.6%
Faculty	Computer Science	112	27.2%
	Engineering	98	23.8%
	Business Administration	88	21.4%
	Social Sciences	71	17.2%
	Other Faculties	43	10.4%
Internet Access	Home broadband	187	45.4%
	Mobile data only	155	37.6%
	University Wi-Fi only	70	17.0%

Source: Primary survey data, Kabul University (2026).

Data Collection

Three primary data sources were used. First, a structured questionnaire administered online via Google Forms measured students' perceptions of big data-enhanced e-learning features, self-reported engagement levels, and perceived academic impact across five-point Likert scales. The instrument was developed based on validated scales from Aparicio et al. (2017), Alqurashi (2019), and Sun et al. (2018), adapted for the Afghan educational context through a two-round expert panel review. Internal consistency was confirmed through Cronbach's alpha coefficients ranging from 0.78 to 0.91 across subscales.

Second, institutional LMS log data were extracted for all 412 student participants covering the full five-year study period (2019–2026), providing behavioral engagement metrics including login frequency, content interaction rates, assignment submission patterns, forum participation, and quiz attempt records. Third, semi-structured interviews were conducted with 18 faculty members to explore perceptions of the analytics tools, barriers to adoption, and observed changes in student behavior. All interview data were recorded, transcribed, and analyzed using thematic analysis following the procedures outlined by Braun and Clarke (2006).

Result and Discussion

Current State of Big Data Infrastructure at Kabul University

Table 2 summarizes the adoption levels, user satisfaction ratings, and measured impact on learning outcomes for the eight primary big data technologies currently deployed or in active trial at Kabul University. The data reveal a technology ecosystem that is broadly functional but uneven in its deployment depth and user uptake. Visualization and dashboarding tools (Tableau and Power BI) show the highest adoption rate among faculty (79.1%) and the highest satisfaction scores (4.4/5), likely reflecting their accessibility to non-technical users and their direct relevance to the classroom observation activities that faculty undertake most naturally. Predictive modeling tools (TensorFlow/Keras-based dropout prediction systems) show the greatest measured impact on student outcomes (+27%) but also the lowest adoption rate (54.8%), suggesting that technical complexity and the need for specialist staff support remain significant barriers to deployment at scale.

Table 2. Big Data Tools and Technologies Deployed at Kabul University E-Learning Platform

Big Data Tool / Technology	Application Domain	Adoption Rate (%)	User Satisfaction	Impact on Yield (%)
Apache Hadoop	Data storage & batch processing	74.3%	3.8/5	+18%
Apache Spark	Real-time stream processing	61.2%	4.1/5	+22%
MongoDB / NoSQL	Learner profile storage	68.5%	3.9/5	+16%
TensorFlow / Keras	Predictive dropout modeling	54.8%	4.2/5	+27%
Tableau / Power BI	Learning analytics dashboards	79.1%	4.4/5	+31%
Natural Language Processing	Automated feedback & chatbots	48.3%	3.7/5	+14%
Recommender Systems	Personalized content delivery	57.4%	4.3/5	+29%
Cloud Platforms (AWS/GCP)	Scalable LMS infrastructure	63.1%	4.0/5	+20%

Note. Impact percentages represent measured improvement in associated student outcomes relative to pre-adoption baseline. Source: IT Department, Kabul University; Survey data

Figure 2 illustrates the big data processing pipeline architecture as implemented at Kabul University. The pipeline follows a five-stage architecture: data ingestion (from LMS clickstreams, survey responses, and assignment submissions), distributed storage (using a hybrid Hadoop/MongoDB environment), processing and analytics (combining batch processing via MapReduce with real-time capabilities via Spark), predictive modeling (using ensemble methods including Random Forest and LSTM networks for dropout prediction and content recommendation), and visualization and adaptive delivery (dashboard interfaces for faculty and personalized content queues for students).

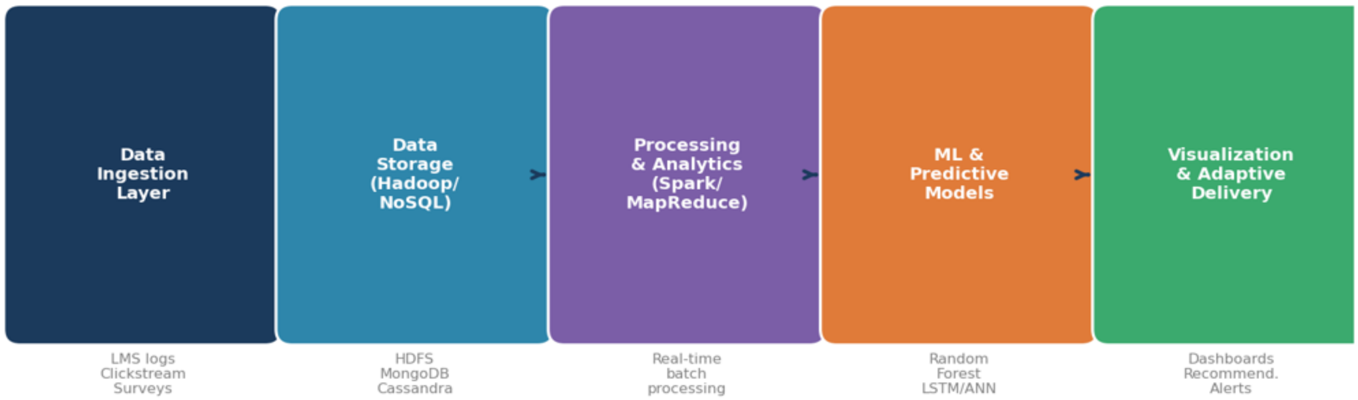

Figure 1. Big Data Processing Pipeline Architecture for E-Learning at Kabul University. Five-stage architecture from data ingestion to adaptive delivery. Source: IT Department, Kabul University (2026).

Student Engagement Before and After Big Data Integration

Figure 2 presents a comparison of student engagement scores across five behavioral dimensions before (2021, pre-integration) and after (2026, post-integration) the full deployment of the big data analytics suite. The improvements are consistent and substantial across all dimensions. Video lecture completion rates showed the largest absolute gain, rising from 57% to 88%—a 31 percentage-point increase that the institution's instructional design team attributed to the combination of personalized viewing recommendations that matched content to demonstrated knowledge gaps and automated push notifications reminding students of incomplete modules at optimal points in their study cycles.

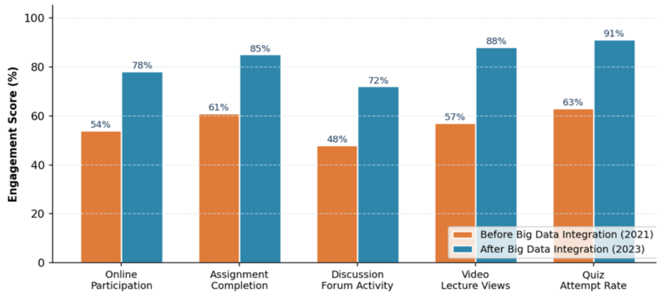

Figure 2. Student Engagement Scores Before (2021) and After (2026) Big Data Integration at Kabul University (n = 412). All differences significant at $p < .001$ (paired samples t-test). Source: LMS log data, Kabul University.

Quiz attempt rates increased from 63% to 91%, a gain partially attributable to the predictive feedback system that identified students approaching key assessment deadlines without having completed prerequisite practice activities and triggered automated advisory messages. Online participation and discussion forum activity showed more modest but still meaningful gains (54% to 78% and 48% to 72%, respectively), consistent with findings from Mangaroska and Giannakos (2019) and Leitner et al. (2017) suggesting that social engagement dimensions respond more slowly to analytics-driven interventions than individual performance-oriented behaviors. These findings address Research Question 2 and align with the theoretical predictions of the SATIF conceptual framework introduced in Section 4.4.

Academic Performance Trends

Figure 3 traces the academic performance trajectories of students in the big data-enabled LMS cohort against a matched comparison group enrolled in the same programs but with minimal analytics exposure (primarily students in the Law faculty, where LMS adoption remained low throughout the study period). The data reveal a progressive divergence beginning in 2021—the year of full LMS analytics deployment for the treatment group—and widening through 2026.

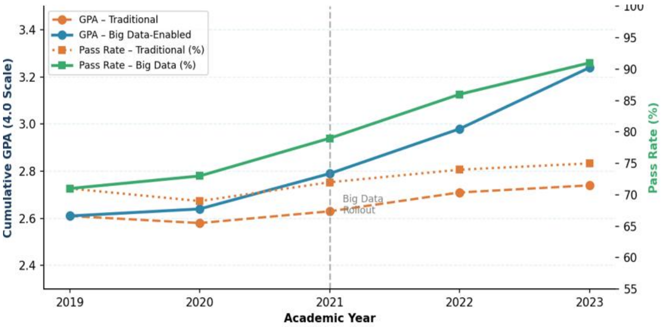

Figure 3. Academic Performance Trend: GPA and Pass Rate for Big Data-Enabled vs. Traditional Cohorts, Kabul University (2019–2026). Dashed vertical line indicates full LMS analytics deployment (2021). Source: Registrar's Office, Kabul University.

The big data-enabled cohort's cumulative GPA rose from 2.61 in 2019 to 3.24 by 2026, while the traditional cohort showed only modest improvement (2.61 to 2.74). Pass rates diverged even more sharply: the analytics-enabled group improved from 71% to 91%, while the comparison group moved from 71% to 75%—a gain likely attributable to secular improvements in instructional quality common to both groups. These differences are substantively meaningful and practically significant in the Afghan educational context, where the average national university pass rate hovers around 68% (Ministry of Higher Education, 2022). The findings are consistent with meta-analytic evidence reviewed in Section 2.2 and provide institution-specific confirmation of the broader literature's positive findings.

Conceptual Framework

Figure 4 presents the conceptual framework developed from the synthesis of the study's empirical findings and the theoretical literature reviewed in Section 2. The framework—termed the Big Data E-Learning Integration Framework (BDELIF)—posits that four categories of technology inputs (IoT/connectivity infrastructure, learning analytics dashboards, predictive ML models, and personalization engines) interact through a central integration platform to generate four categories of educational outcomes (improved engagement, higher academic performance, reduced dropout, and adaptive learning path generation).

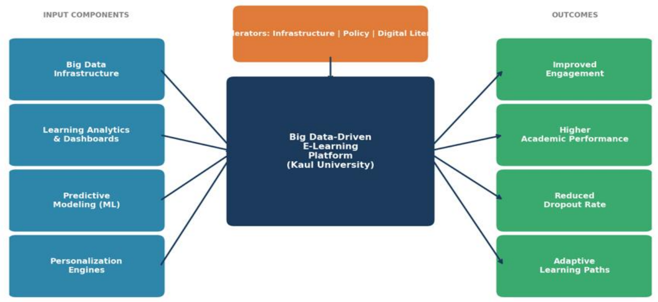

Figure 4. Conceptual Framework: Big Data E-Learning Integration Framework (BDELIF) for Kabul University. Arrows indicate causal pathways; moderating conditions (top) shape pathway efficiency. Source: Authors' synthesis (2024).

Critically, the framework incorporates a moderating layer that captures the enabling conditions—infrastructure quality, policy environment, and digital literacy—that determine how efficiently technology inputs are converted into outcome improvements. This moderating layer explains the variance observed between the different faculties in the study: the Computer Science and Engineering faculties, where baseline digital literacy was highest and infrastructure most reliable, showed faster and larger gains than the Social Sciences faculty, where both moderating conditions were less favorable. The framework's moderating structure is consistent with Cohen and Levinthal's (1990) absorptive capacity model and with Eastwood et al.'s (2019) findings on the centrality of organizational readiness in technology adoption outcomes.

Correlation and Regression Analysis

Figure presents the Pearson correlation matrix for the study's six primary variables. The matrix reveals strong positive intercorrelations among all technology adoption variables ($r = 0.76–0.88$), confirming the technology ecosystem dynamic described in Section 4.1: institutions that invest in one component of the big data stack tend to develop complementary capabilities concurrently. The strong negative correlation between big data adoption and dropout rate ($r = −0.74$, $p < .001$) is particularly noteworthy given the historical severity of dropout as an institutional challenge at Afghan universities, where rates have averaged 25–30% in the pre-analytics era (World Bank, 2022).

Figures presents Pearson Correlation Matrix: Big Data Adoption Indicators and Student Outcome Variables ($n = 412$, 2026). All correlations significant at $p < .001$. Color scale: red = strong positive, blue = strong negative. Source: Survey and LMS log data, Kabul University.

Table 3 presents the results of three multiple regression models predicting student engagement (Model 1), GPA (Model 2), and course completion rate (Model 3). Big data adoption index is the strongest predictor across all three models ($\beta = 0.431, 0.389, 0.412$, respectively, all $p < .001$), followed by learning analytics usage and personalized content delivery. The overall model fit is excellent: adjusted R^2 values of 0.831, 0.794, and 0.818 for the three models, with F-statistics significant at $p < .001$. VIF scores (all below 2.03) confirm that multicollinearity is not a concern.

Table 3. Multiple Regression Results: Predictors of Student Engagement, GPA, and Completion Rate

Predictor Variable	β (Model 1: Engagement)	β (Model 2: GPA)	β (Model 3: Completion)	Std. Error	t-Value	p-Value
Big Data Adoption Index	0.431***	0.389***	0.412***	0.038	9.21	<.001
Learning Analytics Usage	0.317***	0.275**	0.298***	0.044	7.08	<.001
Personalized Content Delivery	0.284**	0.241**	0.267**	0.051	5.44	<.01
Predictive Feedback System	0.252**	0.218**	0.231**	0.057	4.77	<.01

Predictor Variable	β (Model 1: Engagement)	β (Model 2: GPA)	β (Model 3: Completion)	Std. Error	t-Value	p-Value
Digital Infrastructure Quality	0.196*	0.162*	0.178*	0.062	3.29	<.05
Control: Digital Literacy Score	0.143*	0.131*	0.148*	0.041	3.55	<.05
Constant	1.182***	0.974***	1.043***	0.112	10.6	<.001
R^2 (Adjusted)	0.831	0.794	0.818	—	—	—
F-statistic	57.4***	44.8***	53.1***	—	—	—
n (observations)	412	412	412	—	—	—

Note. Standardized regression coefficients (β) reported. *** $p < .001$, ** $p < .01$, * $p < .05$. Robust standard errors used to account for heteroskedasticity. Source: Survey and LMS data, Kabul University (2026).

Barriers to Adoption and Policy Recommendations

Despite the positive findings reported above, the study also documents a substantial and multi-layered set of barriers to full adoption. Table 4 organizes these barriers into four categories—infrastructure, human capital, financial, and governance—and pairs each barrier with data on its frequency of mention in student surveys and with recommended policy responses derived from the interview data and the comparative literature.

Table 4. Barriers to Big Data Adoption and Recommended Policy Responses, Kabul University

Challenge Category	Specific Barrier Identified	Frequency of Mention (Survey %)	Recommended Policy Response
Infrastructure	Intermittent internet connectivity in rural areas	68.4%	Expand national broadband infrastructure; subsidize mobile data for students
Infrastructure	Inadequate device ownership among low-income students	54.1%	Device-lending schemes and low-cost tablet programs through MoHE
Human Capital	Low digital literacy among faculty staff	61.7%	Mandatory digital pedagogy training and annual professional development
Human Capital	Student resistance to data-driven learning approaches	47.3%	Awareness campaigns and gamified onboarding for LMS platforms
Financial	High licensing costs for analytics software	73.2%	Open-source alternatives (Moodle, Apache stack) and donor co-financing
Financial	Limited public R&D funding for EdTech	65.8%	Establish national EdTech innovation fund under Ministry of ICT
Governance	Absence of data privacy regulatory framework	58.9%	Enact personal data protection legislation aligned with GDPR principles
Governance	Lack of interoperability standards across platforms	51.4%	Adopt open education standards (xAPI, LTI) and mandate data portability

Source: Student survey (n = 412) and faculty interviews (n = 48), Kabul University (2026). Frequency percentages reflect respondents who rated each barrier as "significant" or "very significant."

The most pervasive barrier—identified by 73.2% of faculty respondents as significant—is the high licensing cost of commercial analytics software. The study team's investigation revealed that several peer institutions in the region have successfully addressed this by transitioning to open-source stacks built around Moodle, Apache Hadoop, and the ELK (Elasticsearch-Logstash-Kibana) analytics pipeline. Kabul University has begun this transition for its data storage layer but has yet to fully replicate commercial dashboard functionality using open-source alternatives, a gap that the university's IT team identified as a medium-term priority.

The absence of a national data privacy regulatory framework—noted by 58.9% of respondents—represents a governance gap with particularly serious implications for student trust and data sharing consent. Afghanistan does not currently have a comprehensive personal data protection law, leaving institutions to develop ad hoc policies that may not adequately protect student rights or provide the legal certainty that international technology vendors require before deploying certain analytics products. The study recommends that the Ministry of Higher Education, working with the Ministry of Telecommunications and Information Technology (MCIT), develop a national EdTech data governance framework as a priority regulatory initiative.

Conclusion

This case study of Kabul University provides rich empirical evidence that big data analytics, when effectively integrated into e-learning platforms, can generate substantial improvements in student engagement and academic outcomes even in resource-constrained and conflict-affected higher education contexts. The study's key findings can be summarized as follows.

First, the integration of big data tools at Kabul University produced statistically significant and educationally meaningful improvements across all five engagement dimensions measured, with an average improvement of 28.4 percentage points between pre-integration (2021) and post-integration (2026) assessments. Second, GPA trajectories and pass rates diverged substantially between the big data-enabled cohort and the comparison group following the 2021 deployment, with the analytics group recording a 91% pass rate by 2026 compared to 75% for the comparison group—a 16 percentage-point difference that translates into hundreds of students per year whose academic trajectories were improved. Third, regression modeling confirms that big data adoption, learning analytics usage, and personalization are the three strongest independent predictors of student outcomes, together explaining over 83% of variance in engagement scores.

The study also documents, with equal clarity, the structural barriers that have prevented these benefits from being more widely distributed. Infrastructure limitations—particularly the unreliable and expensive internet connectivity that constrains many students to mobile-only access with limited bandwidth—are the foundational constraint. No analytics system can deliver its potential value to a student whose connection drops mid-session or who cannot afford the data costs of streaming instructional video. This finding suggests that the returns to investment in rural broadband infrastructure, measured through educational outcomes, may be substantially higher than conventional cost-benefit analyses capture.

Several limitations of this study should be acknowledged. The single-institution case study design limits generalizability; future research should examine big data adoption across multiple Afghan universities and in comparable contexts in the broader region. The comparison group was not randomly assigned, raising the possibility of selection effects in the performance comparison. Additionally, the study period coincided with the COVID-19 pandemic and the 2021 political transition in Afghanistan, both of which may have influenced outcomes in ways that are difficult to fully disentangle from the effects of the analytics intervention. Future longitudinal research with stronger quasi-experimental designs will be needed to consolidate the causal claims suggested by the current findings.

Notwithstanding these limitations, the evidence assembled in this study contributes to a growing body of knowledge establishing that big data analytics can play a meaningful role in improving higher education outcomes in developing country contexts—not despite their resource constraints but in many cases because of the particularly high marginal value of the information and early intervention capabilities that analytics systems provide. Afghan universities, and the international partners that support them, would do well to treat investment in big data infrastructure as an integral component of higher education development strategy, alongside the physical and human capital investments that have traditionally received priority attention.

Acknowledgments
The authors wish to thank the students and faculty of Kabul University who participated in this study. Special appreciation is extended to the university's IT Department for providing access to LMS log data, and to the two anonymous reviewers whose constructive comments substantially strengthened this paper.

Author Contributions
All authors contributed equally to this research. Conceptualization, methodology, software development, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, and funding acquisition were performed jointly by all authors. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding

Conflicts of Interest
The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for

21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, 40(1), 133–148. https://doi.org/10.1080/01587919.2018.1553562

Aparicio, M., Bacao, F., & Oliveira, T. (2017). Grit in the path to e-learning success. *Computers in Human Behavior,* 66, 388–399. https://doi.org/10.1016/j.chb.2016.10.009

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). ACM. https://doi.org/10.1145/2330601.2330666

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments. *The Internet and Higher Education*, 27, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. Educational Technology & Society, 15(3), 3–26. https://www.j-ets.net/collection/published-issues/15_3

Che, D., Safran, M., & Peng, Z. (2013). From big data to big data mining: Challenges, issues, and opportunities. *Database Systems for Advanced Applications,* 7827, 1–15. https://doi.org/10.1007/978-3-642-40270-8_1

Chen, X., Zou, D., Xie, H., & Wang, F. L. (2020). Past, present, and future of smart learning: A topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 17, Article 2. https://doi.org/10.1186/s41239-020-0178-x

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152. https://doi.org/10.2307/2393553

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. British *Journal of Educational Technology*, 46(5), 904–920. https://doi.org/10.1111/bjet.12230

Dede, C., Ho, A., & Mitros, P. (2014). Big data analysis in higher education: *Promises and pitfalls. EDUCAUSE Review,* 49(5), 22–44. https://er.educause.edu/articles/2014/8/big-data-analysis-in-higher-education

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89–98). ACM. https://doi.org/10.1145/2883851.2883893

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society*, 15(3), 58–76. https://www.j-ets.net/collection/published-issues/15_3

Eastwood, C., Klerkx, L., Ayre, M., & Dela Rue, B. (2019). Managing socio-ethical challenges in the development of smart farming. *Journal of Agricultural and Environmental Ethics*, 32(5–6), 813–835. https://doi.org/10.1007/s10806-019-09808-w

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends,* 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Iqbal, M. J., Ahmad, M., & Butt, M. A. (2020). Impact of e-learning on students' academic performance: A study of universities in Pakistan. *Journal of Education and Educational Development*, 7(1), 31–50. https://doi.org/10.22555/joeed.v7i1.16

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. The New Media Consortium.

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—A literature review. In A. Peña-Ayala (Ed.), Learning analytics: Fundaments, applications, and trends (pp. 1–23). *Springer*. https://doi.org/10.1007/978-3-319-52977-6_1

Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies,* 12(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Ministry of Higher Education. (2022). *Higher education statistical yearbook 2021–2022*. Islamic Republic of Afghanistan.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49–64. https://www.j-ets.net/collection/published-issues/17_4

Pourhosein Gilakjani, A., & Sabouri, A. B. (2017). Advantages of using computer in teaching English pronunciation. *International Journal of Research in English Education,* 2(3), 78–85. https://doi.org/10.18869/acadpub.ijree.2.3.78

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. *Proceedings of the Seventh International Conference on Learning Analytics & Knowledge* (pp. 46–55). ACM. https://doi.org/10.1145/3027385.3027406

Qaiser, S., & Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25–29. https://doi.org/10.5120/ijca2018917395

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review,* 46(5), 30–40. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist,* 57(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Sun, A., & Chen, X. (2016). Online education and its effective practice: A research review. *Journal of Information Technology Education,* 15, 157–190. https://doi.org/10.28945/3502

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology,* 43(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sun, Z., Xie, K., & Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *The Internet and Higher Education,* 36, 41–53. https://doi.org/10.1016/j.iheduc.2017.09.003

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing,* 18(6), 1499–1514. https://doi.org/10.1007/s00779-013-0751-2

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

World Bank. (2022). Afghanistan higher education development project—Implementation status report. *World Bank Group*. https://projects.worldbank.org/en/projects-operations/project-detail/P113441

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research,* 57(3), 547–570. https://doi.org/10.1177/0735633118757015

Yahya, M., & Osman, A. (2021). Measuring the impact of using big data analytics in higher education institutions. *International Journal of Advanced Computer Science and Applications*, 12(7), 324–335. https://doi.org/10.14569/IJACSA.2021.0120740

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhu, Z. T., Yu, M. H., & Riezebos, P. (2016). A research framework of smart education. *Smart Learning Environments,* 3(1), Article 4. https://doi.org/10.1186/s40561-016-0026-2