Sustainable Social Development 2026, 4(4), 8735.
https://doi.org/10.23812/ssd8735

Article

Digital media technology in higher education for SDG 4: Advancing sustainable learning environments

Hui Li[*], Mohamad Nizam Nazarudin[*], Faridah Mydin Kutty[*]

College of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
*** Corresponding author:** Hui Li, P131094@siswa.ukm.edu.my; Mohamad NizamNazarudin, mohdnizam@ukm.edu.my; Faridah Mydin Kutty, faridah_mydin@ukm.edu.my

CITATION

Li H, Nazarudin MN, Mydin Kutty, M. Digital media technology in higher education for SDG 4: Advancing sustainable learning environments. Sustainable Social Development. 2026; 4(4): 8735. https://doi.org/10.23812/ssd8735

ARTICLE INFO

Received: 2 June 2026
Revised: 25 June 2026
Accepted: 15 July 2026
Available online: 7 August 2026

Abstract: Digital media technologies have become central to higher-education delivery, raising important questions about how online learning environments can support inclusion, equity, and sustainability under Sustainable Development Goal 4 (SDG 4). This study examines associations between recorded digital engagement and academic outcomes using the Open University Learning Analytics Dataset, which contains 32,593 enrolments from 28,785 learners across 22 course-presentations and 10,655,280 timestamped learner–VLE records. Descriptive statistics, Pearson correlation, and multivariable logistic regression were used to examine engagement–outcome associations and demographic inequities. Total VLE interactions were aggregated by learner-enrolment, transformed as $\log(x+1)$, and standardised before modelling. Median interactions increased from 89 among withdrawn learners to 1896 among distinction learners, and log-engagement was positively correlated with success ($r = 0.477$). In the complete-case model ($n = 31,482$), a one-standard-deviation increase in log-engagement was associated with higher odds of success (OR = 27.074, 95 % CI [24.918, 29.415], $p < 0.001$) after adjustment for socioeconomic deprivation, gender, disability, age, prior education, previous attempts, and studied credits. Model discrimination was high (ROC-AUC = 0.891; accuracy = 81.0 %), although a sensitivity analysis excluding withdrawn enrolments produced a smaller association (OR = 7.036), indicating that withdrawal-related truncation explains part of the estimate. Success rates also varied from 35.2 % in the most deprived group to 57.5 % in the least deprived group and from 38.1 % among disabled learners to 48.2 % among non-disabled learners. Because the data are observational, institution-specific, and from 2013–2014, the results should be interpreted as historical associations rather than causal or universally contemporary effects. The findings support further prospective evaluation of learning-analytics systems, accessible course design, and equity-focused student support.

Keywords: sustainable education; educational technology; learning analytics; digital engagement; higher education inclusion; SDG 4 implementation; technology-enhanced pedagogy; equitable learning environments

1. Introduction

1.1. Background of the study

The world of higher education is currently experiencing rapid structural change as digital media technologies become an integral part of higher education teaching and learning. Virtual Learning Environments (VLEs), learning management systems (LMS), mobile learning platforms, AI-supported adaptive learning tools, video-based instruction and digital collaborative workspaces are no longer just add-on features but now form the backbone of learning. UNESCO estimates that in the COVID-19 pandemic, some 800 million people were impacted by emergency remote education, underscoring the critical importance of digital infrastructure for higher education as a practical need and a policy priority [1]. This has led to a growing institutional investment in educational technology, driving the structural changes that were already underway: the learning process is increasingly measurable, the learning process has become "data rich", and the course of learning is increasingly detached from fixed space and fixed time. A key feature of this shift is the capacity of digital platforms to record learners' interactions with course resources, providing an empirical basis for examining associations between recorded online activity and educational outcomes. Such records do not, by themselves, identify the causal effects of technology or engagement [2,3].

Digital technologies used in modern higher education include content delivery applications (asynchronous videos, electronic textbooks, open courses), assessment tools (on-line tests, automated feedback systems), social learning environments (forums, collaborative wikis, peer review systems), and increasingly powerful AI-based personalisation systems that adapt content delivery and feedback to the profile of each learner [4]. Wherever this terrain is, access to digital resources has become one of the most consistent predictors of academic performance, but it is not automatic, and not necessarily equitable: A platform that can be flexibly accessed for one learner to learn at their own pace may be an impediment for another learner who does not have the right devices, or a strong broadband connection, or the literacy skills for navigating new digital interfaces [5,6].

1.2. Research problem

Although digital media technology has found great adoption within higher education systems, significant difficulties still need to be overcome in creating an inclusive, equitable, and sustainable learning experience as required by various policy frameworks. Three major obstacles should be noted. First, digital inequities regarding access to equipment and broadband, as well as spatial arrangements for studying, create a divide among learners on social, geographical, and disability bases [7]. Second, having access to modern technologies alone is not sufficient for making productive use of them, due to the prevalence of platform fatigue, lack of effective instruction, and support for developing digital skills [8]. Third, the quality-assurance frameworks and student-support architectures of most institutions remain poorly calibrated to the real-time engagement data that digital platforms continuously generate, leaving an enormous and actionable data resource largely unexploited [9]. The aggregate result is that the expansion of digital higher education has not reliably narrowed pre-existing educational inequalities; in some documented contexts it has reproduced or widened them [10].

1.3. Link with SDG 4

Goal 4 of the United Nations 2030 Agenda for Sustainable Development is to ensure inclusive and equitable quality education and promote lifelong learning for all [11]. Digital higher education is directly relevant to three SDG 4 targets : Target 4.3 (equal access to affordable and quality tertiary education), Target 4.4 (relevant technical and vocational skills for employment, including digital skills), and Target

4.5 (elimination of disparities and equal access to all levels of education for the vulnerable, including persons with disabilities) [12]. Digital media technology is a favourite option as an SDG 4 lever, because it helps to remove the physical limits of capacity, geographic closeness, and time inflexibility, that historically have limited access to higher education. Its contribution to SDG 4, however, is dependent: technology helps to advance the goal only when access leads to real engagement, and real engagement leads to equitable, measurable learning outcomes [13]. A platform that structurally disadvantaged groups can't or don't use just moves inequality around instead of taking it away.

1.4. Research purpose

This study examines how recorded engagement with digital media is associated with academic outcomes and outcome disparities relevant to SDG 4. It does not estimate causal effects or statistical moderation. Specifically, the study evaluates associations involving socioeconomic deprivation, disability, gender, age, and prior educational attainment while accounting for previous course attempts and current study load.

1.5. Research questions

RQ1. How are learning outcomes distributed across a large population of digitally enrolled higher-education learners, and how severe is the retention challenge?

RQ2. To what extent is the intensity of recorded engagement with the Virtual Learning Environment associated with academic success?

RQ3. Do socio-economic status, prior education, gender, and disability create systematic disparities in outcomes — the equity dimension of SDG 4?

RQ4. When engagement and learner characteristics are modelled jointly, which variables show the largest adjusted associations with academic success?

2. Materials and methods

2.1. Research design

This study employs a quantitative, learning-analytics research design grounded in secondary analysis of large-scale administrative and behavioural data. The design is appropriate for the research objectives for three reasons. First, the research questions require statistical estimation of associations between continuous predictors (engagement intensity) and a binary outcome (academic success), which is amenable to logistic regression [14]. Second, the large sample size (n > 32,000 enrolments) supports the disaggregated equity analyses required by the SDG 4 framework without loss of statistical power in minority sub-groups. Third, the use of an anonymised, publicly available dataset eliminates participant-level risk and enables full replication, consistent with open-science principles increasingly mandated by international journals [15]. The analytical pipeline progresses through data integration, cleaning, feature engineering, descriptive analysis, correlational analysis, and inferential modelling.

2.2. Data source and dataset description

Dataset name: Open University Learning Analytics Dataset (OULAD), summarised in **Table 1.**

Table 1. Summary of the OULAD dataset used in this study.

Attribute	Detail
Source platform	Kaggle / Open University Knowledge Media Institute, UK
Country / Context	United Kingdom — distance higher education
Time period	Academic presentations: 2013–2014 (B/J cohorts)
Unique students	28,785
Enrolment records	32,593
Course-presentations	22 (7 modules × B/J presentations)
Digital interactions	10,655,280 timestamped VLE interaction-day records
Access URL	https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad
Licence	Open University: Creative Commons Attribution 4.0

The Open University (OU) is among the largest providers of distance education worldwide, thus OULAD represents a rather unique example of how such digitally enabled, scalable, and lifelong higher education envisaged by SDG 4 can be delivered [16]. The database used in this study is relational and is comprised of seven related tables linked to each other via module ID, presentation ID, and student ID [17].

2.3. Variables and measurement

Table 2 Presents the operationalisation of the study's key variables.

Table 2. Variable operationalisation and roles in the analytical framework.

Variable	Operationalisation	Scale	Role
VLE engagement	Total clickstream interactions (log-transformed)	Continuous	Independent
Academic success	Pass or Distinction vs. Fail or Withdrawn	Binary	Dependent
IMD band	Index of Multiple Deprivation (1 = most deprived, 10 = least)	Ordinal	Equity/Covariate
Disability	Self-declared disability status	Dichotomous	Equity/Covariate
Gender	Male / Female	Dichotomous	Control
Prior education	Highest qualification on entry (6 categories)	Ordinal	Covariate
Age band	0–35 / 35–55 / 55+	Ordinal	Control
Prior attempts	Number of previous registrations for same module	Count	Control
Studied credits	Credit load in current presentation	Continuous	Control

2.4. Data preprocessing

Data preparation was conducted at the learner-enrolment level. The student information and VLE activity tables were linked using module code, presentation code, and student identifier. The 10,655,280 student–VLE records were aggregated by learner-enrolment using the sum of recorded clicks. A total of 3365 enrolments had no corresponding studentVle record; these enrolments were assigned zero recorded interactions because the absence of an activity record indicates no recorded VLE activity rather than an unknown value. Total engagement was transformed as ln(total clicks + 1) to reduce positive skewness.

Among the 32,593 enrolments, IMD band was missing for 1111 observations (3.41 %). No missing values were identified for age band, gender, disability status, highest education, previous attempts, studied credits, or final result. Enrolments with missing IMD were retained in descriptive and outcome-specific analyses but excluded from the primary complete-case regression, producing an analytical sample of 31,482 enrolments. A sensitivity analysis using median IMD imputation together with a missingness indicator was also conducted. Continuous regression predictors were standardised within the analytical sample to a mean of zero and a standard deviation of one. Academic success was coded as 1 for Pass or Distinction and 0 for Fail or Withdrawn. Descriptive statistics and the composition of the full enrolment sample are reported in **Table 3**.

2.5. Analytical methods

Analyses proceeded in four stages. First, final outcomes, learner characteristics, and VLE activity were summarised using frequencies, percentages, means, standard deviations, medians, and interquartile ranges. Second, subgroup-specific success and withdrawal rates were calculated for gender, disability, socioeconomic deprivation, age, and prior education. Third, the bivariate association between log-transformed engagement and binary academic success was estimated using the Pearson point-biserial correlation. Fourth, multivariable logistic regression was used to estimate adjusted associations with academic success.

Total clicks were transformed as ln(clicks + 1) because of substantial positive skewness. Continuous predictors were subsequently z-standardised within the complete-case analytical sample. Therefore, the engagement odds ratio represents a one-standard-deviation increase in log-transformed engagement and not one additional raw click. In this sample, one standard deviation was 2.456 log units, corresponding to an approximately 11.66-fold difference in clicks + 1. Age and prior education were represented using indicator variables, with age 0–35 and A Level or equivalent serving as reference categories.

Logistic regression coefficients were estimated by maximum likelihood. Logistic-regression specification, coefficient interpretation, and model assessment followed established guidance for applied logistic modelling [18]. Results are reported as coefficients, standard errors, Wald z-statistics, two-sided p-values, odds ratios, and 95 % confidence intervals. Model performance was evaluated using McFadden's pseudo-R^2, ROC-AUC, accuracy, sensitivity, specificity, precision, *F1*-score, Brier score, and confusion-matrix counts. Variance-inflation factors were calculated to assess multicollinearity. Robustness was examined through student-grouped five-fold cross-validation, cluster-robust standard errors by learner, 99th-percentile winsorisation of clicks, median IMD imputation with a missingness indicator, and exclusion of withdrawn enrolments.

Table 3. Descriptive statistics and sample composition for the OULAD enrolment records. IQR = interquartile range; IMD = Index of Multiple Deprivation; VLE = Virtual Learning Environment.

Variable	Valid *n*	Missing *n*	Mean	SD	Median	IQR	Range
Academic success, 1 = yes	32,593	0	0.472	0.499	0	0–1	0–1
Total VLE clicks	32,593	0	1,215.141	1692.604	602	142–1,585	0–24,139
ln(total clicks + 1)	32,593	0	5.706	2.450	6.402	4.963–7.369	0–10.092
IMD band, 1–10	31,482	1,111	5.212	2.816	5	3–8	1–10
Female, 1 = yes	32,593	0	0.452	0.498	0	0–1	0–1
Disabled, 1 = yes	32,593	0	0.097	0.296	0	0–0	0–1
Previous attempts	32,593	0	0.163	0.480	0	0–0	0–6
Studied credits	32,593	0	79.759	41.072	60	60–120	30–655
Variable/category			***n***			**%**	
Pass			12,361			37.93	
Distinction			3024			9.28	
Fail			7052			21.64	
Withdrawn			10,156			31.16	
Female			14,718			45.16	
Male			17,875			54.84	
Disabled			3164			9.71	
Non-disabled			29,429			90.29	
Age 0–35			22,944			70.40	
Age 35–55			9,433			28.94	
Age 55+			216			0.66	
A Level or equivalent			14,045			43.09	
Lower than A Level			13,158			40.37	
Higher-education qualification			4730			14.51	
No formal qualification			347			1.06	
Postgraduate qualification			313			0.96	

2.6. Ethical considerations

This research project is based on a secondary data set which is open access and completely anonymous that has been released for this research project under a Creative Commons License by the Open University. OULAD does not contain any personally identifying information, there is no involvement of any human subjects and there is no data collection for this purpose. There was therefore no need for ethical approval in the context of the ethical guidelines of the Universities affiliated to the authors. The study strictly adheres to the guidelines laid down by Committee on Publication Ethics (COPE) [19].

3. Results

3.1. Conceptual framework

Figure 1 presents a conceptual framework that organises potential relationships among digital infrastructure, learner characteristics, recorded engagement, institutional design, and SDG 4-relevant outcomes. The arrows represent theoretically proposed relationships rather than causal effects estimated in the present study. Recorded engagement is treated as an observable behavioural indicator that may reflect several underlying processes, including access, motivation, self-regulation, prior knowledge, and course design. Learning analytics is positioned as a potential institutional feedback mechanism whose effectiveness must be evaluated prospectively.

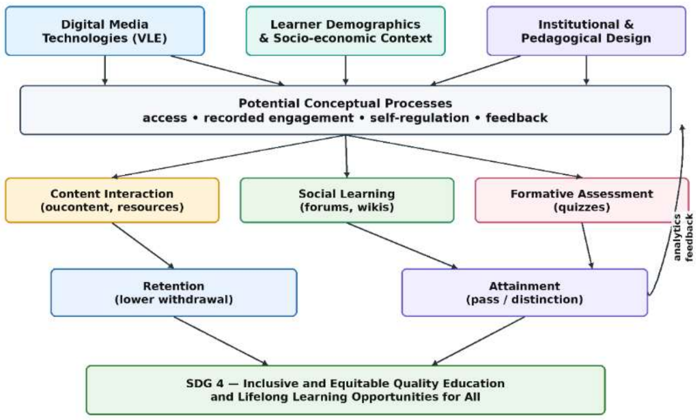

Figure 1. Conceptual framework of hypothesised relationships among digital media technologies, learner and institutional contexts, potential conceptual processes, digital activity domains, and SDG 4-relevant outcomes. Arrows represent conceptual relationships rather than causal effects estimated from OULAD.

3.2. Outcome landscape and the retention challenge (RQ1)

With respect to the 32,593 enrolments included in this analysis, 37.9 % concluded with a Pass, while another 9.3 % were awarded a Distinction, giving a total success rate of 47.2 %. A substantial proportion of enrolments (31.2 %) ended in withdrawal, making withdrawal the most common adverse outcome. A further 21.6 % of enrolments formally failed their courses. Together, these outcomes indicate that non-completion remains a major challenge for achieving the SDG 4 goal of inclusive and equitable lifelong higher education[20] (**Figure 2**).

3.3. Equity dimension (RQ3)

3.3.1. Gender

Outcome distributions were broadly similar by gender (**Figure 3**). Female learners had a slightly higher overall success rate than male learners (48.5% versus 46.2%), but the descriptive difference was small. Because the data are cross-sectional and institution-specific, they do not support conclusions about discrimination or temporal narrowing of gender disparities.

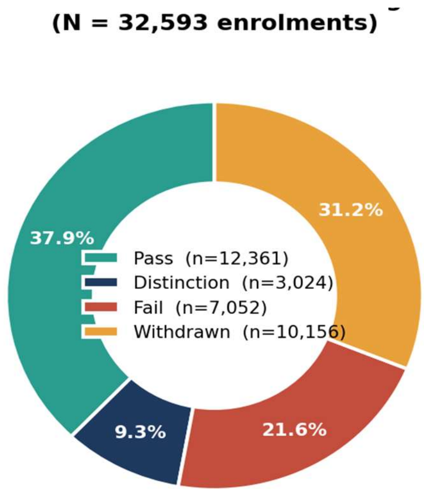

Figure 2. Distribution of final outcomes across all enrolments. Withdrawal is the largest single adverse category.

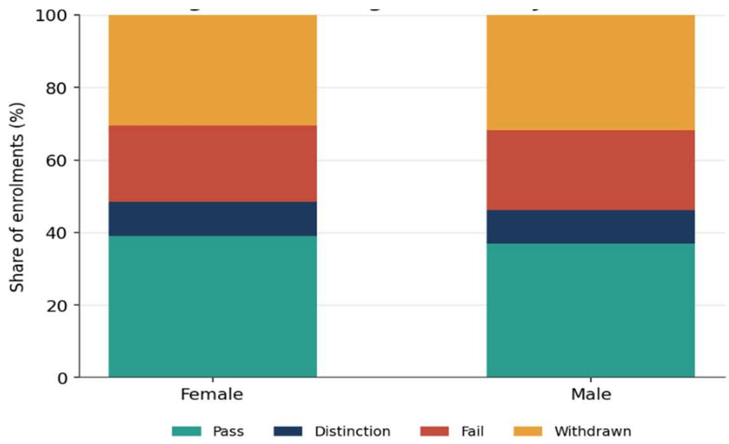

Figure 3. Outcome composition by gender. Distributions are closely comparable.

3.3.2. Prior educational attainment

Success rates increased across prior-education categories, from 29.7% among learners with no formal qualifications to 65.5% among those with postgraduate qualifications **(Figure 4)**. This pattern indicates an educational-capital gradient but does not establish a causal effect of prior education. Digital provision alone may therefore be insufficient to offset accumulated differences in prior educational preparation [21]. In line with human capital theory [22], the finding reinforces the importance of combining tertiary digital-learning opportunities with strong foundational education.

3.3.3. Socio-economic deprivation — Core equity finding

The Index of Multiple Deprivation (IMD) analysis indicates the highest equity signal of all the data (**Figure 5**). The success rate increases almost in a linear fashion according to socio-economic advantage, ranging from 35.2 % for learners in the most deprived decile, to 57.5 % for those in the least deprived (a gradient of about 22 percentage points). The correlation between IMD band and success is bivariate, which is $r = 0.131$; this is not a strong correlation in absolute value, but as mentioned it is consistent and clear in direction. The key equity conclusion of the research is that digitally mediated higher education as it is commonly structured is consistent with the persistence of socioeconomic disparities, but does not demonstrate that digital provision caused the disparity. Deliberate compensating interventions are needed to advance SDG 4 Target 4.5, which must address this gradient [23].

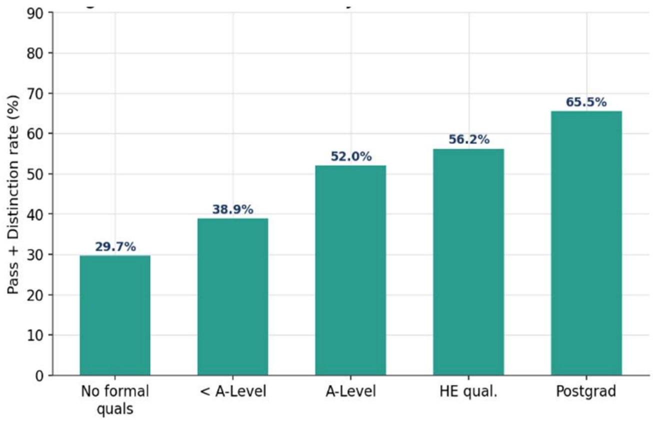

Figure 4. Success rate by highest prior qualification, illustrating a strong educational-capital gradient.

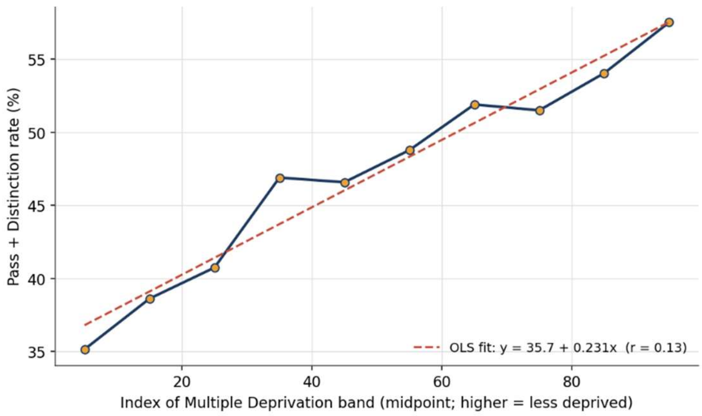

Figure 5. Socio-economic gradient in academic attainment across IMD deciles.

3.3.4. Disability status

The gap between learners' success rates for declaring a disability and those of non-disabled learners is around 10 percentage points, with those who declare a disability achieving 38.1 % compared with 48.2 % of non-disabled learners. Again the withdrawal rate for disabled learners (39.3 %) is significantly higher than for non-disabled learners (30.3 %) (see **Figure 6**). The gaps highlight that obstacles to learning in the digital learning space, including screen-reader support, captioning, interface complexity, and bandwidth requirements, remain a disadvantage to this population, which is explicitly mentioned in SDG 4 [24].

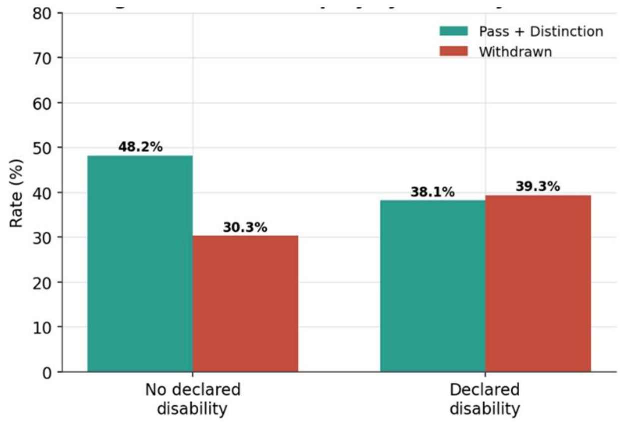

Figure 6. Outcome equity by disability status. Disabled learners face higher withdrawal and lower academic success.

3.3.5. Age and regional variation

As shown in **Table 4**, the proportion of learners achieving success in the assessment increases with age from 45.0 % for those aged under 35 to 61.6 % for those aged 55 and over. This is reflective of the characteristics of mature lifelong learners, for which SDG 4 is designed, who are more self-regulated and motivated to learn. There is also an element of regional variation (**Figure 7**); attainment in the geographic areas, including those in Wales, deviates from the national average by up to 8 percentage points suggesting that local socio-economic and infrastructure influences the relationship between digital provision and outcome.

Table 4. Academic success rate by age band. Older learners achieve markedly higher success, consistent with SDG 4 lifelong-learning objectives.

Age band	Pass + distinction rate	Interpretation
0–35	45.0 %	Younger, often less self-directed
35–55	52.2 %	Mid-career lifelong learners
55 and over	61.6 %	Mature, self-directed learners

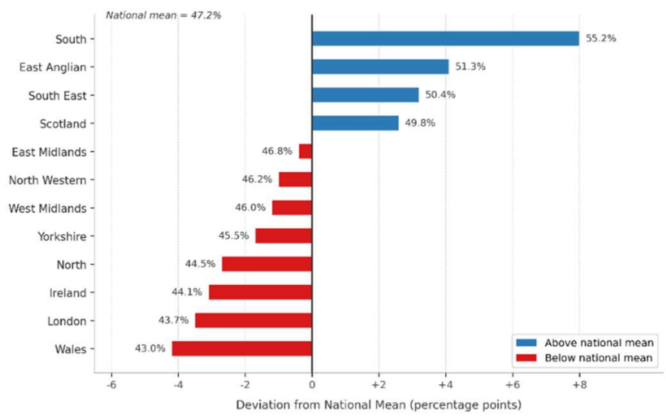

Figure 7. Regional variation in attainment relative to the national mean.

3.4. Digital engagement and academic success (RQ2)

The relationship between engagement with digital media technology and academic success is both large in magnitude and monotonic in form. Median VLE interactions rise from just 89 clicks among withdrawn learners and 317 among failing learners, to 1343 among those who pass and 1896 among distinction students (**Figure 8**). The contrast between the two extremes spans more than an order of magnitude, indicating that engaged learners are not merely marginally more active — They inhabit a qualitatively different mode of participation.

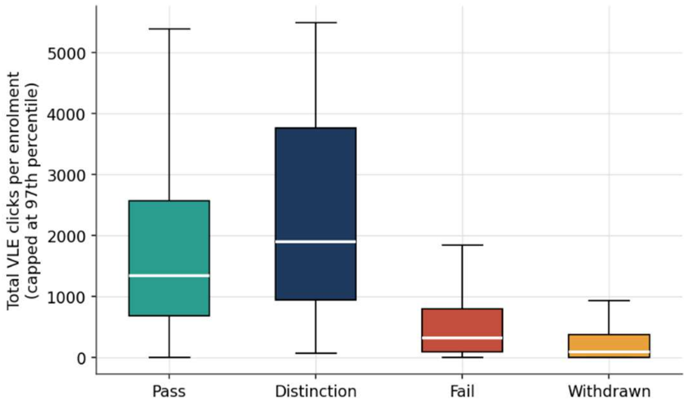

Figure 8. Distribution of total VLE interactions by final outcome. Engagement separates outcome groups sharply and monotonically.

Translating engagement into population deciles reveals a clean graded association across engagement deciles (**Figure 9**): the success rate rises from 0.1 % in the lowest engagement decile to 92.1 % in the highest, with the overall correlation r = 0.477 — strong by educational-data standards [25]. This graded pattern is a central descriptive finding; however, it does not demonstrate that greater VLE activity causes improved academic outcomes because engagement may also reflect motivation, prior understanding, course progression, and withdrawal timing.

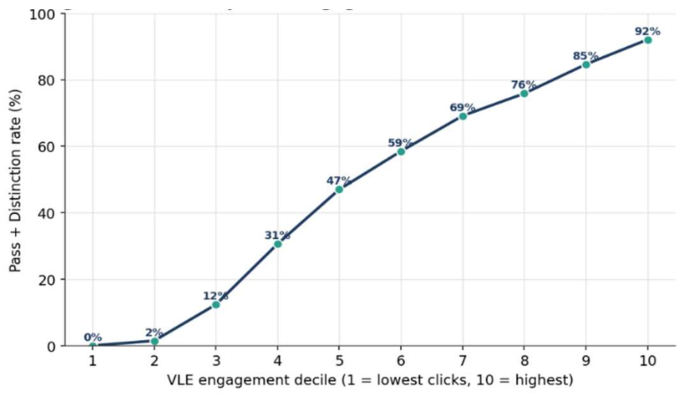

Figure 9. Graded observational association between VLE engagement decile and academic success rate. The pattern does not establish a causal dose–response effect.

The disaggregation by resource type (**Figure 10**) illustrates the locations of engagement activity. Content covered within the core course (oucontent), discussion forums (forumng), online tests, and the course home page make up an enormous proportion of the total number of recorded activities, out of 39,605,099 aggregated activities. This quartet of key points, including the teaching of course content, discussion, and testing, is the main area where the digital pedagogical activity takes place. These counts describe where recorded activity occurred but do not indicate whether individual interactions involved sustained attention, deep learning, or effective knowledge construction.

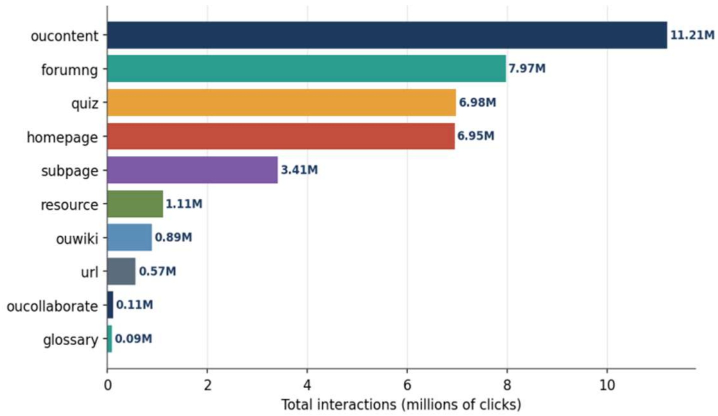

Figure 10. Total interactions by type of digital learning resource. Core content, forums, and quizzes dominate the engagement landscape.

3.5. Module-level synthesis

The learner-leveled engagement–outcome association mirrors at the course-designed level. **Table 5** shows the number of people enrolled, the median engagement time, success rate and withdrawal rate for each module in OULAD. The modules with the highest median engagement (FFF: 1683 clicks; AAA: 1160 clicks) achieve more above-average success rates, the ones that achieved the lowest median engagement (BBB: 320 clicks; GGG: 422 clicks) have more mixed performance. The pattern may reflect course design, learner composition, assessment structure, or other module-level differences and not just individual motivation, and that consequently the engagement profile, and thus the learning outcomes, of entire cohorts is shaped.

Table 5. Module-level enrolment, engagement, and outcome summary. Modules with higher median engagement generally achieve higher success and lower withdrawal rates.

Module	Enrolments	Median clicks	Success rate	Withdrawal rate
AAA	748	1160	71.0 %	16.8 %
BBB	7909	320	47.5 %	30.2 %
CCC	4434	525	37.8 %	44.5 %
DDD	6272	563	41.6 %	35.9 %
EEE	2934	1072	56.2 %	24.6 %
FFF	7762	1683	47.0 %	31.0 %
GGG	2534	422	59.7 %	11.5 %

3.6. Multivariable associations (RQ4)

The multivariable logistic regression was estimated using 31,482 complete cases. The model achieved McFadden $R^2 = 0.413$, ROC-AUC = 0.891, and classification accuracy = 81.0 %, indicating good discrimination within the analysed OULAD sample [26,27].

The complete set of coefficients and odds ratios is reported in **Table 6**, and the adjusted odds ratios are visualised in **Figure 11**. Standardised log-engagement had the largest adjusted odds ratio (OR = 27.074). Socioeconomic advantage was independently associated with success (IMD band OR = 1.196). Female learners had higher adjusted odds of success (OR = 2.483). Previous attempts (OR = 0.909), heavier credit load (OR = 0.752), and declared disability (OR = 0.745) were associated with lower adjusted odds of success.

Table 6. Multivariable logistic regression estimates for academic success. $n = 31{,}482$; McFadden $R^2 = 0.413$; ROC-AUC = 0.891; accuracy = 81.0 %. Reference categories are male, non-disabled, age 0–35, and A Level or equivalent. Continuous predictors were z-standardised. Statistical significance was evaluated using two-sided Wald z-tests. β = log-odds coefficient; SE = standard error; OR = odds ratio; CI = confidence interval.

Predictor	β	**SE**	**Wald z**	p	**OR**	**95 % CI for or**
Intercept	−1.124	0.033	−33.591	<0.001	0.325	0.304–0.347
VLE engagement, log and z-standardised	3.299	0.042	77.935	<0.001	27.074	24.918–29.415
IMD band, z-standardised	0.179	0.016	11.311	<0.001	1.196	1.160–1.234
Female	0.909	0.033	27.606	<0.001	2.483	2.328–2.648
Disabled	−0.294	0.053	−5.589	<0.001	0.745	0.672–0.826
Age 35–55	−0.148	0.035	−4.212	<0.001	0.862	0.805–0.924
Age 55+	−0.228	0.205	−1.110	0.267	0.796	0.532–1.191
HE qualification	−0.062	0.048	−1.286	0.198	0.940	0.856–1.033
Lower than A Level	−0.526	0.034	−15.463	<0.001	0.591	0.553–0.632
No formal qualification	−0.703	0.170	−4.130	<0.001	0.495	0.355–0.691
Postgraduate qualification	−0.146	0.199	−0.736	0.462	0.864	0.585–1.275
Previous attempts, z-standardised	−0.095	0.017	−5.630	<0.001	0.909	0.879–0.940
Studied credits, z-standardised	−0.285	0.017	−16.764	<0.001	0.752	0.728–0.778

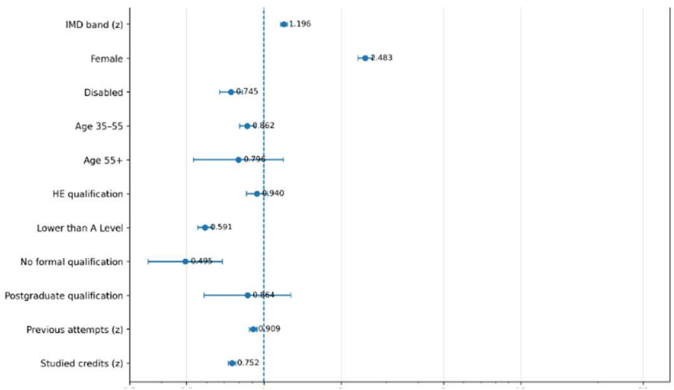

Figure 11. Adjusted odds ratios and 95 % confidence intervals from the multivariable logistic regression. Estimates represent conditional associations rather than causal effects. The vertical reference line indicates OR = 1.

3.6.1. Model diagnostics and robustness

The expanded model showed good discrimination, with ROC-AUC = 0.891 and accuracy = 0.810 at the 0.50 probability threshold. Sensitivity was 0.831, specificity was 0.792, precision was 0.776, and F1 was 0.803. The Brier score was 0.132. The confusion matrix contained 13,319 true negatives, 3,508 false positives, 2476 false negatives, and 12,179 true positives. Variance-inflation factors ranged from 1.017 to 1.462, providing no indication of problematic multicollinearity.

Student-grouped five-fold cross-validation, which prevented enrolments belonging to the same learner from appearing simultaneously in the training and test partitions, produced ROC-AUC = 0.891 and accuracy = 0.810. Results were also stable after winsorising clicks at the 99th percentile and after retaining all observations through median IMD imputation with a missingness indicator. When withdrawn enrolments were excluded, the adjusted association between engagement and success was smaller (OR = 7.036, 95 % CI [6.611, 7.487]) but remained statistically significant. This reduction indicates that the primary estimate partly reflects the mechanical reduction in clicks following withdrawal and should not be interpreted as a causal effect. The complete model diagnostics and sensitivity-analysis results are summarised in **Table 7**.

Table 7. Model diagnostics and sensitivity analyses for the engagement–success association.

Analysis	***n***	**Engagement or (95 % CI)**	**ROC-AUC**	**Accuracy**	**Sensitivity**	**Specificity**	**McFadden R^2**
Main adjusted model	31,482	27.074 (24.918–29.415)	0.891	0.810	0.831	0.792	0.413
Five-fold student-grouped cross-validation	31,482	—	0.891	0.810	0.831	0.791	—
Cluster-robust SE by learner	31,482	27.074 (24.867–29.476)	0.891	0.810	0.831	0.792	0.413
Clicks winsorised at 99th percentile	31,482	27.123 (24.965–29.467)	0.891	0.810	0.831	0.792	0.413
Median IMD imputation plus missing indicator	32,593	26.944 (24.836–29.231)	0.891	0.811	0.836	0.788	0.414
Withdrawn enrolments excluded	21,562	7.036 (6.611–7.487)	0.843	0.811	0.927	0.565	0.310

4. Discussion

4.1. Digital engagement as an observational indicator relevant to sustainable learning

Across OULAD, recorded VLE engagement showed the largest adjusted association with academic success among the variables included in the model. This finding does not establish that increasing click counts would cause academic success. Engagement may reflect motivation, self-regulation, prior understanding, time availability, course design, access conditions, and impending withdrawal. Clickstream data are therefore better interpreted as behavioural risk indicators than as direct causal policy levers. They may help institutions identify patterns that warrant further support, but any early-warning or engagement-support system should be prospectively evaluated for discrimination, calibration, subgroup fairness, false-alert burden, and intervention effectiveness before operational deployment. The graded association shown in **Figure 9** is consistent with earlier learning-analytics research [27–29], but it should not be described as a causal dose–response relationship. These results complement and build on previous research in the field of learning analytics in higher education.

4.2. The equity caveat: Digital technology alone does not level the field

The 22-percentage-point IMD gradient and approximately 10-percentage-point disability gap document persistent outcome inequalities within the OULAD setting. They do not demonstrate that digital technology caused or perpetuated these gaps, and the present analysis did not estimate a statistical mediation model through engagement. Unobserved differences in connectivity, study time, prior preparation, employment, accessibility needs, and support resources may contribute to the observed patterns. Equity-oriented digital-learning strategies should therefore combine accessible design, connectivity support, human advising, and careful monitoring of differential model performance rather than assuming that greater platform activity alone will remove structural disadvantage. At the policy level, these equity-oriented measures are also consistent with broader calls to accelerate targeted investment and financing for SDG implementation [30].

4.3. Retention and the potential role of early-warning research

Withdrawal was the most common adverse outcome, affecting 31.2 % of enrolments. Previous online-learning research has demonstrated that dropout-risk systems can combine repeated prediction with targeted interventions, although such evidence does not validate the present retrospective model as a deployable early-warning system [31]. Recorded clickstream patterns may support the future development of withdrawal-risk tools; however, the present model used total interactions accumulated across the course and was not designed or prospectively validated as an early-warning system. A deployable early-warning model would need to use only information available before a clearly defined prediction date and would require evaluation of lead time, calibration, false positives, subgroup fairness, and the effectiveness of the subsequent intervention. The present findings therefore provide a rationale for prospective early-warning research rather than evidence that a specific early-warning intervention is already effective.

4.4. Institutional and design factors

Module-level differences in median engagement and success are descriptive. A sociotechnical perspective further emphasises that educational technologies operate through institutional, material, pedagogical, and social conditions rather than as context-independent solutions [32]. They may reflect course design, subject area, assessment structure, presentation duration, student composition, or other unmeasured factors. Consequently, the present comparisons do not isolate a causal effect of instructional design. Nevertheless, the variation indicates that course context should be explicitly considered in future multilevel or fixed-effects analyses of engagement and attainment.

4.5. Theoretical implications

At a broader theoretical level, NeuroIS offers complementary approaches for examining users' cognitive and affective responses to digital information systems; however, such processes were not measured in OULAD [33]. These results can be categorized into three different schools of thought. Self-efficacy theory better explains this: engagement is shaped by learners' confidence, resources, and self-regulatory capacity, which are constrained when socioeconomic deprivation limits access to the necessary material means [34]. Learning analytics theory further explains engagement as both a behavioural signal and an institutional feedback mechanism that links learner circumstances with timely support and course redesign [35]. Finally, the integration of AI-enabled adaptive learning tools into these frameworks represents a natural extension as such systems become more widely deployed in higher education [36].

4.6. Limitations, endogeneity, and contemporary relevance

Several limitations delimit the interpretation of the findings. First, OULAD is observational; consequently, the reported coefficients represent conditional associations rather than causal effects. Recorded engagement may be endogenous because motivation, self-regulation, prior ability, available study time, internet and device access, employment responsibilities, and course selection may affect both VLE activity and academic outcomes. Reverse causality is also plausible because learners who understand the material or expect to succeed may engage more, whereas withdrawal mechanically prevents later interactions from being recorded. Consistent with this concern, excluding withdrawn enrolments reduced the engagement odds ratio from 27.074 to 7.036, although the association remained positive. OULAD does not contain a clearly defensible instrumental variable that would affect academic outcomes only through engagement. Fixed-effects or longitudinal panel models could address some stable learner- or course-level heterogeneity but would not remove time-varying unobserved confounding. No causal mechanism is therefore claimed.

Second, total VLE clicks capture the volume of recorded online activity rather than the quality, purpose, attention, cognitive effort, or learning strategy associated with that activity. Identical click totals may reflect substantially different learning behaviours, resource types have different pedagogical meanings, and offline study is not captured. The click-count variable should therefore be interpreted as a behavioural trace rather than a complete measure of meaningful engagement.

Third, OULAD represents one UK distance-education provider during 2013–2014. The data predate pandemic-scale digital transformation, widespread videoconferencing, mobile-first learning, generative artificial intelligence, and current adaptive-learning ecosystems. OULAD remains useful as a historical benchmark for large-scale distance education, but its estimated associations should not be assumed to represent contemporary campus-based, blended, international, or AI-supported higher education. Replication using post-2020 data from multiple institutions and countries is necessary.

Fourth, some learners appear in more than one enrolment record, and module-level differences may reflect subject matter, learner composition, assessment structure, presentation timing, or course design. Student-grouped cross-validation and cluster-robust standard errors were used to evaluate dependence among repeated learner records, but the transferability of the model to new institutions remains untested.

4.7. Future research directions

Several directions warrant further investigation. First, quasi-experimental or experimental designs — randomised deployment of early-warning systems, for example — would enable causal claims about engagement-support interventions. Second, the qualitative dimension of engagement (depth, purposefulness, and metacognitive strategy) merits investigation alongside the quantitative click-count proxy employed here. Third, multi-institutional and cross-national replication studies would test the generalisability of the engagement–outcome–equity chain beyond the single-institution OULAD context. Fourth, the integration of AI-enabled adaptive learning and personalisation tools into learning-analytics frameworks represents a natural extension as such systems become more widely deployed in higher education [36–38]. Future studies should also construct time-bounded predictors using only information available before the prediction date, validate models prospectively, assess calibration and subgroup fairness, and evaluate whether interventions triggered by learning-analytics alerts improve outcomes through randomised or credible quasi-experimental designs.

5. Conclusion

This study examined associations among recorded VLE engagement, learner characteristics, and academic outcomes in 32,593 enrolments from the Open University Learning Analytics Dataset. Engagement differed substantially across final-outcome groups, and log-transformed engagement was positively correlated with academic success. In the complete-case multivariable model, a one-standard-deviation increase in log-engagement was associated with higher odds of success after adjustment for socioeconomic deprivation, gender, disability, age, prior education, previous attempts, and studied credits. The model achieved ROC-AUC = 0.891 and accuracy = 81.0 %. However, the engagement odds ratio decreased from 27.074 to 7.036 when withdrawn enrolments were excluded, showing that withdrawal-related truncation contributed to the magnitude of the primary estimate.

The analysis also identified substantial socioeconomic and disability-related outcome gaps. These disparities demonstrate that access to digitally delivered higher education does not automatically produce equivalent outcomes across learner groups. Institutions considering learning-analytics support should therefore combine accessible course design, connectivity and study support, human advising, and formal assessment of algorithmic fairness. Recorded engagement may provide a useful risk indicator, but the present analysis does not establish that increasing click activity will itself improve academic outcomes.

The findings should be interpreted within the limitations of an observational dataset from one UK distance-education institution during 2013–2014. Unmeasured motivation, self-regulation, prior ability, available study time, course selection, and reverse causality may influence the estimated associations. Total clicks also measure recorded activity volume rather than engagement quality. Future research should use post-2020 multi-institutional data, time-bounded predictors, qualitative engagement measures, and experimental or credible quasi-experimental evaluations of learning-analytics interventions. OULAD remains a valuable historical benchmark, but contemporary and causal claims require additional evidence.

Author contributions: Conceptualisation, ARK and SM; methodology, ARK; software, ARK; validation, ARK, SM and BH; formal analysis, ARK; investigation, ARK; resources, SM; data curation, ARK; writing — original draft preparation, ARK; writing — review and editing, SM and BH; visualisation, ARK; supervision, BH; project administration, BH. All authors have read and agreed to the published version of the manuscript.

Funding: None.

Ethical approval: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The dataset used in this study is the Open University Learning Analytics Dataset (OULAD), openly available at: https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad. The data were downloaded in January 2024, cleaned, and pre-processed as described in Section 2.4. No modifications were made to the underlying anonymised records. The pre-processed analytical dataset is available from the corresponding author upon reasonable request.

Acknowledgements: The authors gratefully acknowledge the Open University and the OULAD development team (Kuzilek, Hlosta, and Zdrahal) for making the dataset openly available under a Creative Commons licence. The authors also thank the editorial team and anonymous reviewers for their constructive comments.

Conflict of interest: The authors declare no conflict of interest.

References

1. UNESCO. Education in a Post-COVID World: Nine Ideas for Public Action. International Commission on the Futures of Education. UNESCO: Paris, France; 2020. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000373717 (accessed on 10 February 2025).
2. Tlili A, Shehata B, Adarkwah MA, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments. 2023; 10: 15. doi:10.1186/s40561-023-00237-x
3. Mayer RE. Multimedia Learning, 3rd ed. Cambridge University Press: Cambridge, UK; 2020. doi:10.1017/9781316941355
4. Crompton H, Burke D. Artificial intelligence in K–12 education. SN Social Sciences. 2022; 2(7): 113. doi: 10.1007/s43545-022-00425-5
5. Dutta S, Lanvin B. The Network Readiness Index 2023. Portulans Institute: Washington, DC, USA; 2023. Available online: https://networkreadinessindex.org/nri-2023/ (accessed on 15 January 2025).
6. UNESCO GEM Report. Technology in Education: A Tool on Whose Terms? UNESCO: Paris, France; 2023. Available online: https://www.unesco.org/gem-report/en/2023 (accessed on 10 March 2025).
7. World Bank. The Human Capital Index 2020 Update: Human Capital in the Time of COVID-19. World Bank: Washington, DC, USA; 2020. doi:10.1596/34432

8. Bates AW. Teaching in a digital age: Guidelines for designing teaching and learning for a digital world, 3rd ed. Tony Bates Associates Ltd: Vancouver, BC, Canada; 2022. Available online: https://teachonline.ca/teaching-in-a-digital-age (accessed on 10 March 2025).
9. Ifenthaler D, Yau JY-K. Utilising learning analytics to support study success in higher education: A systematic review. Educational Technology Research and Development. 2020; 68(4): 1961–1990. doi: 10.1007/s11423-020-09788-z
10. Hilliger I, Ortiz-Rojas M, Pesantez-Cabrera P, et al. Identifying needs for learning analytics adoption in Latin American universities: A mixed-methods approach. The Internet and Higher Education. 2022; 53: 100842. doi: 10.1016/j.iheduc.2020.100726
11. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations: New York, NY, USA; 2015. Available online: https://sdgs.un.org/2030agenda (accessed on 1 January 2025).
12. UNESCO. Education 2030: Incheon Declaration and Framework for Action for the Implementation of Sustainable Development Goal 4. UNESCO: Paris, France, 2015. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000245656 (accessed on 1 January 2025).
13. Leal Filho W, Lange Salvia A, Beynaghi A, et al. Digital transformation and sustainable development in higher education in a post-pandemic world. International Journal of Sustainable Development & World Ecology. 2024; 31(1): 108–123. doi: 10.1080/13504509.2023.2237933
14. Starkweather J, Moske AK. Multinomial logistic regression. Benchmarks RSS; 2023. Available online: https://it.unt.edu/benchmarks (accessed on 5 April 2025).
15. Williamson B, Eynon R, Potter J. Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. Learning, Media and Technology. 2020; 45: 107–114. doi: 10.1080/17439884.2020.1761641
16. Huang R, Liu D, Tlili A, et al. Handbook on facilitating flexible learning during educational disruption. Smart Learning Institute of Beijing Normal University/UNESCO: Beijing, China; 2020. Available online: https://iite.unesco.org/publications/handbook-on-facilitating-flexible-learning/ (accessed on 12 February 2025).
17. Kuzilek J, Hlosta M, Zdrahal Z. Open university learning analytics dataset. Scientific Data. 2017; 4: 170171. doi: 10.1038/sdata.2017.171
18. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Wiley: Hoboken, NJ, USA; 2013. doi: 10.1002/9781118548387
19. Committee on Publication Ethics (COPE). Core Practices. COPE: London, UK, 2022. Available online: https://publicationethics.org/core-practices (accessed on 20 March 2025).
20. Global Education Monitoring Report Team.Education for People and Planet: Creating Sustainable Futures for All. UNESCO: Paris, France; 2016. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000245752 (accessed on 15 January 2025)
21. Herodotou C, Rienties B, Hlosta M, et al. The online transition of activities in distance learning universities and the role of learning analytics: Lessons learnt from the Open University UK. The Internet and Higher Education. 2021; 48: 100780. doi: 10.1016/j.iheduc.2020.100780Becker GS.
22. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 3rd ed. University of Chicago Press: Chicago, IL, USA; 1993. doi: 10.7208/chicago/9780226041223.001.0001
23. Rienties B, Boroowa A, Cross S, et al. Reviewing three case-studies of learning analytics interventions at the Open University UK. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge; 2016. pp. 534–535. doi: 10.1145/2883851.2883886
24. OECD. Education at a Glance 2023: OECD Indicators. OECD Publishing: Paris, France; 2023.
25. Tight M. Systematic reviews and meta-analyses of higher education research. European Journal of Higher Education. 2019; 9(2): 133–152. doi:10.1080/21568235.2018.1541752
26. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis, 8th ed. Cengage Learning: Stamford, CT, USA; 2019.
27. Zhao Y, Watterston J. The changes we need: Education post COVID-19. Journal of Educational Change. 2021; 22: 3–12. doi: 10.1007/s10833-021-09417-3
28. Ahshan R. A framework of implementing strategies for active student engagement in remote/online teaching and learning during the COVID-19 pandemic. Education Sciences. 2021; 11(9): 483. doi: 10.3390/educsci11090483

29. Jawad K, Shah MA, Tahir M. Students' academic performance and engagement prediction in a virtual learning environment using random forest with data balancing. Sustainability. 2022; 14(22): 14795. doi: 10.3390/su142214795
30. Sachs JD, Lafortune G, Fuller G, Drumm E. Implementing the SDG Stimulus. Sustainable Development Report 2023. Dublin University Press: Dublin, Ireland; 2023. doi: 10.25546/102924
31. Bañeres D, Rodríguez-González ME, Guerrero-Roldán AE, Cortadas P. An early warning system to identify and intervene online dropout learners. International Journal of Educational Technology in Higher Education. 2023; 20: 3. doi: 10.1186/s41239-022-00371-5
32. Selwyn N. Education and Technology: Key Issues and Debates. London, UK: Bloomsbury Academic; 2022.
33. Davis FD, Riedl R, Vom Brocke J, et al. Information systems and neuroscience: Gmunden Retreat on NeuroIS 2022. Springer; 2023. doi: 10.1007/978-3-031-13064-9
34. Schunk DH, DiBenedetto MK. Self-efficacy and human motivation. Advances in Motivation Science. 2021; 8: 153–179. doi: 10.1016/bs.adms.2020.10.001
35. Khalil M, Prinsloo P, Slade S. The use and application of learning theory in learning analytics: A scoping review. Journal of Computing in Higher Education. 2023; 35: 573–594. doi:10.1007/s12528-022-09340-3
36. Lameras P, Arnab S. Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. Information. 2022; 13(1): 14. doi: 10.3390/info13010014
37. Kabudi T, Pappas I, Olsen DH. AI-enabled adaptive learning systems: A systematic mapping of the literature. Computers and Education: Artificial Intelligence. 2021; 2: 100017. doi: 10.1016/j.caeai.2021.100017
38. Crompton H, Burke D. Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education. 2023; 20:22. doi:10.1186/s41239-023-00392-8