Journal of World Englishes and Educational Practices
ISSN: 2707-7586
DOI: 10.32996/jweep
Journal Homepage: www.al-kindipublisher.com/index.php/jweep

JWEEP
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

The Use of Big Data in Education: Students' Attitudes and Perceptions at the School of Letters and Human Sciences, Dhar El Mehraz, Fes (Morocco)

Imane ZERYOUH[1], Imane JAI LAMIMI[2] and Maegan KEARNS[3]
[123]Faculty of Humanities and Social Sciences, Euro-Mediterranean University, Fez, Morocco
Corresponding Author: Imane ZERYOUH, **E-mail**: i.zeryouh@ueuromed.org

| ABSTRACT
This paper describes students' attitudes and perceptions of Big Data as a form of educational technology. These insights provide contexts and implications of Big data in improving student's learning environment and experience. The overarching data-fused systems allow for achieving better standards and enhancing the educational landscape. This study is descriptive, and the Validity of Big Data Applied to Education (VABIDAE) questionnaire has been employed to collect quantitative data. It is a three-level scale that assesses students' attitudes and perceptions of Big data as an ed-tech. Thus providing a comprehensive deduction of students' tendencies, concerns, and appreciation of the advantages and drawbacks of this technology. The hypothesis suggests that students' perception of big data as an educational technology is reflected on their attitude. To avoid ambiguity, perception and attitude represent two distinct variables. While perception represents students' views of positive and negative aspects of utility of Big Data in education, attitude represents their subjective point of view, such as feelings and beliefs. To test the hypothesis, This study answers three research questions. The first research question describes students' overall perception of Big data as an educational technology, while the second describes their attitude toward its implementation in their learning journey. The third and last research question investigates the correlation between the two variables. The findings of this study allow students, educators, researchers, policymakers, and shareholders in the field of education to better understand the advantages and challenges related to implementing Big data systems. In conclusion, considering how students view and approach the incorporation of this technology into education significantly contributes to improving learning outcomes and overall educational experiences.

| KEYWORDS

Affordances in Big Data, Big data, Educational technology, Higher education

| ARTICLE INFORMATION

ACCEPTED: 23 September 2025 **PUBLISHED:** 01 October 2025 **DOI:** 10.32996/jweep.2025.7.3.2

Introduction

Big data is data that is too big, too fast and exceeds the analytical properties or capacities of the conventional architectures of database systems, such as on-hand database management tools and traditional data processing applications (Dumbill, 2012). In every sector, data is growing exponentially, in huge volumes and with high velocity; for instance, until 2003, 5 billion gigabytes of data were created. In 2011, 5 billion gigabytes were created every two days. In 2013, the same amount was created every 10 minutes. These enormous volumes of data are necessary today, and they are acquired everywhere, in all aspects of human life, such as science, social networks, business, healthcare, government, media, education, national security, and transportation (Nugent, et al., 2013). Unlike the earlier generations of data that only required human efforts, these data necessitate analytical methodologies and computing resources to be processed and interpreted (Ward & Barker, 2013).

This study is motivated by the scarcity of research on using Big Data in education within the Moroccan context. It describes university students' attitudes and perceptions toward integrating Big Data technology into their learning journey. Its benefits are determined by three types of utility: Knowledge discovery, or novelty discovery, such as discovering something rare (one in a billion or trillion) object or event; association discovery, which is finding unusual or co-occurring associations through making connections between different people or events, thus, finding improbable co-occurring combinations of things; and lastly, class discovery, through which new classes of behaviours or objects are discovered (Dumbill, 2012).

To meet the study objectives, the VABIDAE test battery has been implemented to collect data from a sample of 174 undergraduate students at the School of Letters and Human Sciences, Dhar El Mehraz, Fès. The findings highlight positive attitudes and perceptions towards using Big Data, whereas the correlation between the two variables is negative. This study, being the first of its kind within the Moroccan context, aims towards being employed as a reference for upcoming research in the field.

Levels of Big Data:

Fischer et al. (2020) claim that there are three levels of Big Data: micro, mesolevel and macro. However, it is important to note that these levels are not completely separate or distinct, as they may overlap.

Microlevel Big Data, or interaction data, is retrieved directly and automatically from users. It is collected at the granularity of seconds to gather information about learners, their actions and the context in which they occurred (Ochoa & Worsley, 2016). In education, this data is produced by the exchanges between learners and educational platforms that have data collection platforms such as simulations, intelligent tutoring systems and (MOOCs). The affordances of Microlevel data lie in its utility to provide and detect cognitive strategies and self-regulated learning (SRL) behaviours and feedback on students' learning and engagement.

Mesolevel data includes computerized data of groups and organizations. In education, mesolevel data is systematically collected mainly through students' learning during the various activities and practices in the learning environment, such as assignments and online discussions. Mesolevel data provides opportunities to capture data on learners' development in both cognitive and social abilities, as well as affective states. Besides students' writings and assignments, mesolevel data can be collected through other sources such as website databases or intelligent tutoring systems. It is gathered in time durations that can range from minutes to hours. Mesolevel data provides insights into cognitive, social, affective, and behavioural processes (Fischer et al., 2020).

Macrolevel data is data collected at the institutional level. It includes, for instance, data about students' admission, demographics, grades, scheduled classes, courses and their description, campus service data, degree completion requirements and college major requirements. It is collected over the years and infrequently updated. Macrolevel data provides opportunities to improve data-driven administrative decision-making (Fischer et al., 2020). It improves students' experience and predicts success (Chaturapruek et al., 2018). Macrolevel data can also be applied as early-warning systems, information or course guidance systems (Jayaprakash et al., 2014), and administration-facing analytics (Méndez et al,. 2014).

Students Engagement and Performance

Studies, such as Ogan et al. (2015) and Fancsali et al., (2018), have examined (SRL) and metacognition within educational data mining. These constructions test learners' ability to control their learning processes (Roll & Winne, 2015). Modelling students' processes and behaviours in learning environments is a common step in educational data mining techniques for (SRL) analysis. This step identifies potential learning problems that system architects and designers may utilize to enhance user interfaces and experiences (Aleven et al., 2016; Roll & Winne, 2015). For instance, inferences about Non-cognitive dimensions related to engagement, motivation, and effects are allowed through micro-level data. Furthermore, academic emotions, also known as affective states, such as frustration, perplexity, boredom, and engaged concentration, are constructs that have also been the subject of research in the field (Botelho et al., 2017). The ability of educational data mining tools to recognize emotional states enables the use of emotional detectors to offer learners real-time feedback, scaffolding, and interventions.

Evaluating students' knowledge based on sets of accurate and inaccurate answers to problems, also known as knowledge inference or latent knowledge estimation, is an essential use of micro-level data. Three common techniques are deep knowledge tracing (DKT) (Pavlik et al., 2009), performance factors analysis (PFA), and Bayesian knowledge tracing (BKT) (Corbett & Anderson, 1995). These techniques employ several frameworks to determine a student's skill mastery level.

The oldest of these three methods is (BKT), and it uses a Hidden Markov Model to estimate four parameters for each distinct skill present in the data: the likelihood of a student's mastery of a skill before having the chance to practice it, the likelihood of a student's mastery of a skill after having the chance to practice it but before having the chance to do so again; and the likelihood of a student's mastery who has not yet mastered a skill but will do so in the future. This paradigm was enhanced to include item difficulty estimations, partially accurate answers, and various alternative states for certain knowledge components (González-Brenes et al., 2014; Ostrow et al., 2015; Falakmasir et al., 2015). (BKT) studies underlie adaptivity through various learning platforms, such as the Cognitive Tutor, and support fundamental research on affect detectors (Liu & Koedinger, 2017).

(PFA) uses logistic regression to estimate three parameters for each distinct skill within the data, and (BKT) uses a Hidden Markov Model to infer student knowledge (Pavlik et al., 2009). The three parameters are the degree to which correct answers are associated with better future performance, the degree to which incorrect answers are associated with better future performance and the general ease or difficulty of the skill being estimated. These variables represent the likelihood that a student has mastered a certain skill. Compared to (BKT), (PFA) parameters offer less insight into learners' early ability levels and tendency to commit careless errors. Nevertheless, (PFA) parameters show the relative difficulty of skills and the proportionate learning connected to correct and bad replies. Research is currently being done to extend (PFA) in order to, among other things, better understand mastery criteria (Käser et al., 2016), investigate the relative predictive value of recent performance compared to older performance (Galyardt & Goldin, 2015), and investigate individual differences in learning rate (Liu & Koedinger, 2015).

(DKT) has become a substitute for (BKT) and (PFA). It models skill knowledge and mastery using recurrent neural networks, resulting in a vector of the probability of mastery connected to each opportunity to practice a skill. (DKT) is generally more effective than the other ways of predicting student correctness while learning (Khajah et al., 2016; Yeung & Yeung, 2018), but it has not been widely applied in the real world due to interpretability and stability of estimates issues (Yeung & Yeung, 2018).

Personalized Learning

Prior studies examined thousands of participants in massive open online courses (MOOCs) and hundreds of students in classroom settings. Examples include modelling how different student groups who have different strategies emerge over time in their use of online course resources (Gasevic et al., 2017), identifying distinct patterns of engagement in (MOOCs), and identifying how different student groups work through a learning simulation as part of an experimental standardized test (Bergner et al., 2014; Guo & Reinecke, 2014; Kizilcec et al., 2013).

Joksimovic et al. (2015) examined how (MOOCs) students interacted with their courses on Facebook, Twitter, and blogs. The most popular subjects appeared early in the course. It has been discovered that the topics addressed were comparable across social media platforms. Mathematical problems that were largely created by instructors and provided to an intelligent tutoring system were subjected to an evaluation (Slater et al., 2016). Notably, Slater et al. looked at students working on arithmetic problems to find correlations between the issues' semantic aspects and student learning or engagement, which might help teachers choose which math problems to use in their lessons and how to create new math problems.

Learners' self-concept, judgment, and motivation were evaluated in studies that looked at affective components; these studies frequently looked at hundreds or thousands of students. For instance, Crossley et al. (2018) analyzed data from an online teaching environment and Natural language processing (NLP) techniques to determine how linguistic proficiency and identification in mathematics are related among learners (e.g., math value, interest, and self-concept). Data collected through open-ended questions about course expectations throughout (MOOC) enrollment processes and their associations with age and gender were studied by Crues et al. (2018); it revealed 26 reasons for course enrolment using latent Dirichlet allocation and correspondence analysis, all of which were related to learners' ages but not their gender. Similarly, Reich et al. (2015) employed structural topic modelling to identify semantic meaning patterns in unstructured text to comprehend students' enrollment motivations in an educational policy course.

Educational Data Mining

Educationl data mining (EDM) explores all sorts of data collected from educational environments. It also stands for the techniques to address and answer questions related to education (Lynn & Emanuel, 2021). Data mining in education and learning analytics are used interchangeably to refer to the same research domain. These two concepts correlate with many other research areas, such as machine learning, computer science, computer-based education, and educational statistics.

(EDM) applies to many different educational environments, including (traditional classes, classes through learning management systems (LMS), classes through Intelligent Tutoring Systems (ITS), and finally, (MOOCs). Educational Data could be anything that happens between instructors and students (interaction, quizzes, exercises, participation), administrative data (information about school and teachers) and demographic data about learners (age, gender, parents' marital situation), students' status towards learning (Internal/external motivation, emotional and psychological states). Education environments are regarded as fertile grounds for data from several sources. These sources vary in terms of the quantity of data they offer. The following data sources are organized from less to more data-providing sources, such as courses, students, activities, and events (see Table 1).

Data Mining/Processing

Available raw educational data cannot be used to offer insights about a phenomenon or answer some burning questions in education. Thus, it is of great necessity to convert the raw data and turn it into usable data with a different format. However, data mining stands for applying methods and techniques such as visualization, classification, and association analysis. These techniques enabled data harvesting to solve different education problems (Gushchina & Ochepovsky, 2019)

Table 1

Educational data mining methods

Method	Description	Application
Clustering	Finding similarities among a group of datasets hence providing observations.	Students with similar learning patterns are clustered together.
Prediction	Predicting a variable's behaviour according to an analysis done on other variables.	Predicting student performance and filling in knowledge gaps
Knowledge tracing	Creating a log for each student to trace skills' mastery.	Longitudinal studies for how student's knowledge changes with time
Causal mining	To identify causality among datasets or to find cause-effect	Finding direct causes of some students' low grades and dropping out of school.

Big Data and Educational Research

The reliance on Big data promises a wider understanding of educational processes on different application levels: Microlevel, mesolevel, and macro-level (Fischer et al., 2020). ***Microlevel Microlevel Big Data***

Microlevel Big data is concerned with the Data generated from the exchanges among learners, educators, and the data collection platform (MOOCs, LMS, simulations, and gamified learning) (Fancsali et al., 2018). At this level, Big Data provides insights about learners, their learning behaviour, actions within a given context, and data from the learning context itself. Data at this level is granted through students' accessibility to learning platforms (Mediators) that do not include many students. However, although the data is retrieved from samples of hundreds of students at most, millions, if not billions, of data points are generated due to the complexity of correlations found among variables. Therefore, microlevel Big Data is suitable when interventions for students' classroom behaviour are affordable. Furthermore, this data helps build a knowledge base capable of providing the correct feedback to students based on effective and cognitive datasets (DeFalco et al., 2018).

Mesolevel Big Data

Mesolevel Big Data uses students' generated academic texts, writings, and submitted assignments to learning management systems (LMS). The move toward digital texts has granted an abundance of user-generated Data added to the data lake. Mesolevel datasets are collected in a given time frame that ranges from a few minutes to a few hours. Furthermore, the data is analyzed using specialized software specialized in (NLP), which automates the whole process of analyzing data. Linguistics-infused Big Data tools can shed light on when students float lexical, grammatical, or morphological features in accordance with the previously machine-fed data (Fesler et al., 2019).

Macrolevel Big Data

Macrolevel big data is collected over months and years, including students' personal data, data about courses, records, and grade reports. This data reflects the overall status of the educational body and highlights where the call for action is needed. Macrolevel Big Data has many applications within the realm of education (Fischer et al., 2020).

Warning systems

A Warning System is an extension added to educational platforms to warn students at risk of underachieving in tests and exams. These systems act as predictive models which rely on datasets from tens of thousands of students as part of institutional Big Data to help instructors identify knowledge gaps and further focus on students who may need additional guidance and support (Chaturapruek et al., 2018). Additionally, Fischer et al. (2020) claim that pilot applications to warning systems successfully predicted failure at the course level, relying on data based on (LMS) sessions and academic interaction among students, platforms, and instructors.

Methodology

Research Problem

Big Data systems help educators assess students accurately, allowing better progress and achievement. The correlation between university students' perception and attitude toward this technology has not been adequately discussed through research. As it has always been a challenge to monitor the behavioural patterns of students effectively, Big Data systems enable interpretations and insights into students' daily operations during a course/class. As for any technology implemented in teaching and learning processes, gaining insights into how it is perceived from a student/learner-centered point of view is critical. This study, therefore, describes students' perception and attitude towards Big Data as an educational technology and investigates whether students' overall perception impacts their attitude towards it. This study provides visibility and the recognition of how Big Data is perceived from the point of view of a population of students, which is a critical step in enhancing their learning outcomes/achievements.

The Objective of the Study

Big Data is changing education and students' learning experience. It serves to meet the quality standards of the education landscape and build enhanced models for it. From a learner's learner-centred point of view, adopting Big Data as a learning technology helps educators better understand their students' needs and learning habits or behaviours; however, considering the students' views in parallel is of equivalent value. This leads to improving their learning outcome, performance and overall experience. The study aims to gain insights into students' perception and attitude toward Big Data with regard to it as a technological tool implemented in their learning process.

Research Design

Given that this study is descriptive, relying on a questionnaire as a data collection instrument is legitimate. The questionnaire instrument used to gather quantitative data on students' attitude and perceptions is the (VABIDAE) (Validity of Big Data Applied to Education). This 3-level scale is used in this study to collect the data used in this study.

The first two sections of the questionnaire were used to collect data regarding students' perceptions. The first section was used to collect data about positive perception, the second was for negative perception, and the third section was used to collect data about their attitude. A Likert scale is used (see appendix).

In this study, attitude stands for a more subjective point of view that relates to their personal beliefs, feelings and judgement. Conversely, perception is the variable that represents students' negative and positive views of Big Data as a technological tool, including its convenience and drawbacks; therefore, it depicts how these students perceive its utility concerning different criteria and educational contexts.

Data Collection

In order to answer the research questions, this study relied on quantitative data collected using the (VABIDAE) questionnaire. It was distributed to a convenient sample of 174 students from the English Department of Sidi Mohamed Ben Abdellah University; School of Letters and Human Sciences, Dhar El Mehraz- Fès. During the data collection process, respondents' informed consent has been granted; and, the anonymity and confidentiality of their contribution has been guaranteed.

To guarantee the validity of the study, the study opted for a questionnaire pilot study to gain insights and feedback on the topic. The aim behind the piloting is to scrutinize the relevance, clarity and overall understanding of the questionnaire and test the measurability of the studied variables. In this regard, sample surveys were randomly distributed among a small number of students within our sample, and they were asked to fill them out and grade the understandability, the design and the wording. The input gained from questionnaire piloting contributed to refining the design, which supported our questionnaire's validity and reliability.

Research Questions

1- What is students' overall perception towards the utility of Big Data as an educational technology?

2- What is students' attitude towards the implementation of Big Data in their learning experience?

3- Is there a relationship between students' attitudes and perception of Big Data as an educational technology?

Ethical Consideration

Before collecting the necessary data, permission to use the data collection tool was asked for and granted. Consequently, we contacted the researchers who designed the (VABIDAE) test battery to seek their consent and any potential stipulations for its use. Permission has been granted with the following directive from the designers: "Feel free to use (VABIDAE), and only I request you to cite us properly." In light of this, we have ensured that all references to (VABIDAE) in our work provide appropriate and accurate citations, honoring the wishes of its creators and upholding the ethical standards of academic research.

Validity of the Assessment of Big Data Applied to Education (VABIDAE)

Given prior research, the study conducted by Matas et al. (2020) analyzed the psychometric characteristics of the (VABIDAE) questionnaire, which is designed to assess how students feel and think about Big Data. A non-probability sample of 337 students from both Peru and Spain have participated in this study, and the findings have provided different factors, manifested in a robust psychometric profile that encompasses emotional and cognitive dimensions related to perceptions of Big Data in education. Furthermore, the study has revealed that the questionnaire is indeed relevant and effective in identifying the participants' perspectives, by shedding light on their attitudes and emotions towards this technology. In addition, it emphasizes its utility in exploring and building upon studies that shed light on the implications of Big Data for Inclusive and sustainable education. Notably, the (VABIDAE) questionnaire demonstrated high reliability with a Cronbach’s alpha coefficient of 0.897, which is an indicator of strong reliability.

Limitations

While the use of Big Data analysis offers unprecedented insights into the dynamics of learner interaction and early identification of students at risk of underachieving, sample size limitations are important to acknowledge. The findings based on 174 respondents, suggest a negative correlation between attitudes and perception. To establish broader applicability, it would be valuable to replicate this study across multiple courses, diverse institution types, and perhaps different geographical settings. Additionally, the research at hand focuses on students' attitudes and perceptions toward the use of Big Data as an educational tool. Future studies could touch upon teachers' views on adopting Big Data technologies to routine tasks such as assignment submission trends, engagement with online resources, or even real-time monitoring of student attention via learning platform analytics.

The sample size was an unavoidable limitation due to the available research conditions. It proved to be challenging to have access to a large sample within the given educational context. Furthermore, the premise of working on Big Data directly intertwines with having access to a large population for robust findings and generalizations at the level of the institution studied, but since accessibility was limited; the current study opted for a smaller population to highlight and explore the current views students hold on Big Data use as an educational technology. Furthermore, the implications of a smaller sample size raise questions and concerns of generalizability, highlighting the need to interpret findings cautiously and consider whether they are confidently extrapolated to broader student populations or diverse educational contexts. Henceforth, the aim behind the current research is not to generalize the findings but only to explore this new field of studies in the Moroccan context while the findings are only representative of the studied sample.

Results

Descriptive Results

Research Question One What is students’ overall perception of using Big Data as an educational technology? To answer this research question, data were collected through the first two sections of the (VABIDAE) questionnaire that suggests the different positive (first section) and negative aspects (second section) of using Big Data in education.

Table 2

Perception descriptive statistics

Descriptive Statistics			
	N	Mean	Std. Deviation
Positive Perception	174	3.7969	.56212
Negative Perception	174	3.2557	.58185
Perception	174	3.5263	.40062
Valid N (listwise)	174		

As displayed in Table 1, the means of positive and negative perceptions are calculated separately. The mean of positive perception is (M=3.79), while the mean of negative perception is (M=3.25). The difference between both means entails that students only show a slightly higher positive perception towards the positive aspects of using Big Data. Both means having (SD<0.6) means that the dataset is clustered for both groups. Overall, both variables' mean showcases that students simultaneously have negative and positive perceptions.

Table 3

Calculating students' overall perception

			Statistic	Std. Error
Perception	Mean		3.5263	.03037
	95% Confidence Interval for Mean	Lower Bound	3.4664	
		Upper Bound	3.5863	
	5% Trimmed Mean		3.5341	
	Median		3.5182	
	Variance		.160	
	Std. Deviation		.40062	
	Minimum		2.37	
	Maximum		4.45	
	Range		2.08	
	Interquartile Range		.55	
	Skewness		-.253	.184
	Kurtosis		.184	.366

The Mean (M=3.52) of positive and negative perceptions is generated by computing the variable (Perception) to get students' overall perception of Big Data; Thus, it answers the first research question. (SD=0.4) reflects low data dispersion; thus, it indicates that the value of both variables is balanced in their contribution to the mean of the computed variable (perception). Based on the descriptive statistics, the result is higher than average.

Table 1

Correlation between students' positive and negative perceptions of the use of Big Data in education

Correlations			
		Positive Perception	Negative Perception
Positive Perception	Pearson Correlation	1	-.019
	Sig. (2-tailed)		.802
	N	174	174
Negative Perception	Pearson Correlation	-.019	1
	Sig. (2-tailed)	.802	
	N	174	174

A statistical correlation measure has been conducted to measure the statistical relationship between the two variables. The correlation (r=-0.019) indicates a weak negative correlation (close to r=0), suggesting almost no linear association between the variables.

Research Question Two

What are students' attitudes on implementing Big Data in their learning experience? The third section of the VBIDAE questionnaire has been employed to answer this question. The data is collected through a Likert scale to measure how students relate to statements representing how they feel about Big Data as educational technology.

Table 5

Calculating students' overall attitude of using Big Data in education

Descriptives				
			Statistic	Std. Error
Attitude	Mean		2.9347	.04575
	95% Confidence Interval for Mean	Lower Bound	2.8444	
		Upper Bound	3.0250	
	5% Trimmed Mean		2.9409	
	Median		3.0000	
	Variance		.364	
	Std. Deviation		.60344	
	Minimum		1.20	
	Maximum		4.50	
	Range		3.30	
	Interquartile Range		.72	
	Skewness		-.083	.184
	Kurtosis		.064	.366

The mean of attitude (M=2.93) suggests that respondents' overall view is relatively moderate. It is measured through their reaction to statements representing their attitude towards the theme of Big Data in general as an educational technology. Furthermore, as the confidence interval is narrow (see Table 4), it indicates a high level of precision, thus, the estimation of a true mean.

As the descriptive statistics demonstrate, The minimum value of the dataset is 1.2, and the maximum is 4.5. The variance is 0.36, and (SD=0.6) indicates a moderate level of diversity in the students' responses through a 1 to 5 Linkert scale. Overall, it can be concluded that the analysis at hand answers the research question by demonstrating an average level of perception, with a moderate variation in students' overall attitude.

Research Question Three

This article's third and final question aims to uncover the correlational nature between research questions one and two. In that way, the question inquires about the relationship between students' perception (question one) and attitude (question two).

The Pearson Correlation coefficient (r=-0.117) is conducted between the two variables, Perception and Attitude (see Table 5 below). As a negative correlation is denoted, one variable's rise indicates the other's decrease. Therefore, if the variable of 'Perception' changes, the variable of 'Attitude' changes in the other direction. However, the magnitude of the correlation (r=-0.117) is weak, and its significance value (p=0.124) indicates that the correlation is not statistically significant and that the observed correlation coefficient could have occurred by chance and could not represent the true relationship between the variables.

Table 6

Correlation between students' perception and attitude of Big Data use in education

Correlations			
		Perception	Attitude
Perception	Pearson Correlation	1	-.117
	Sig. (2-tailed)		.124
	N	174	174
Attitude	Pearson Correlation	-.117	1
	Sig. (2-tailed)	.124	
	N	174	174

To conclude, the results of the correlation study indicate a weak negative correlation between the variables "Perception" and "Attitude." The correlation is not statistically significant. Therefore, additional analysis or investigation may be required to determine the significance of the correlation between these variables.

Discussion

Research Question One

The objective of the first question in this study is to examine and describe students' overall perception regarding the use of Big Data as an educational technology. Data was collected through the VIBDAE questionnaire (uses the Likert scale). The first two sections represent positive and negative aspects of big data in education, thus deducing respondents' overall perceptions. As a first step in the data analysis process, the means of positive ($M=3.79$) and negative ($M=3.25$) were calculated separately. The difference between the two means is small, but it suggests that students exhibited a slightly higher level of positive perception towards the positive aspects of Big Data inclusion as a technology integrated into education. ($SD<0.6$) for both groups of data sets indicates that there is only a slight variability in the responses and that the data set is clustered. The variable 'Perception' is computed to represent the overall perception of students, including positive and negative perceptions. Its Mean ($M=3.52$) and ($SD=0.4$) suggest both variables' balanced contribution in computing 'Perception'. In addition to that, based on the statistical findings, it is observed that the level of students' perception is relatively positive as it is higher than average.

Furthermore, an analysis of the correlation coefficient has been conducted to understand the statistical relationship between positive and negative perceptions. The result ($r=-0.019$) showcases a negative correlation that is close to zero. This indicates that the linear association between the positive and the negative aspects of students' perceptions towards Big Data is almost null.

Research Question Two

To answer the second research question, the analysis explored students' attitude, as learners, towards Big Data as an educational technology. Hence, the employment of the third section of the VABIDAE questionnaire. This section consists of statements that capture students' positive and negative feelings and thoughts towards the theme and, thus, their attitude toward it. Based on the descriptive statistics generated, the Mean ($M=2.93$) indicates a relatively moderate overall attitude that is higher than average. Furthermore, ($SD=0.6$) and the variance of 0.36 suggests that students' responses (on a Likert scale of 1 to 5) are relatively diverse. The findings provide insight into the nature and level of students' overall attitude toward Big Data as a theme by introducing statements representing different positive and negative feelings about Big Data implementation as an educational technology in their learning experiences.

Research Question Three

This study's third and final question explores the relationship between students' perception and attitude (questions one and two) concerning Big Data as an educational technology. To answer this question, the analysis entails a Pearson correlation coefficient ($r=-0.117$) demonstrating the relationships between the variables 'Perception' and 'Attitude'. The correlation is observed to be negative; however, the magnitude suggests that it is weak. In addition, based on the significance value ($p=0.124$), it is deduced that the observed negative correlation is not statistically significant and could be due to mere chance. In other words, it may not represent the true relationship between the correlated variables.

Conclusion

The findings of this study describe and shed light on students' perceptions and attitudes regarding Big Data and its use as educational technology. Based on the data collected and the analysis conducted, it is indicated that students showcase a slightly higher level of positive perception compared to negative perception. In addition to that, variability in the data set is weak, and the correlation between the variables in this research is statistically insignificant. Furthermore, it is necessary to point out that it is of paramount importance to inspect other research and literature to understand and explore the implications of Big Data in education. Stemming from the research findings of this paper, it is concluded that students' perception of Big Data does not have an impactful relationship with their personal attitude towards it, and vice versa. However, it is also important to state that the interpretations and deductions of this study are specific to the studied data set; hence, they are potentially not pertinent to generalize over all students and/or educational contexts.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205-223.
[2] Bergner, Y., Shu, Z., & von Davier, A. (2014, July). Visualization and confirmatory clustering of sequence data from a simulation-based assessment task. In *Educational Data Mining 2014*.
[3] Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings 18* (pp. 40-51). Springer International Publishing.
[4] Chaturapruek, S., Dee, T. S., Johari, R., Kizilcec, R. F., & Stevens, M. L. (2018, June). How a data-driven course planning tool affects college students' GPA: evidence from two field experiments. In *Proceedings of the fifth annual ACM conference on learning at scale* (pp. 1-10).
[5] Chaturapruek, S., Dee, T. S., Johari, R., Kizilcec, R. F., & Stevens, M. L. (2018). How a data-driven course planning tool affects college students' GPA: Evidence from two field experiments. Proceedings of the 5th Annual ACM Conference on Learning at Scale, L at S 2018. https://doi.org/10.1145/3231644.3231668
[6] Crossley, S., Ocumpaugh, J., Labrum, M., Bradfield, F., Dascalu, M., & Baker, R. S. (2018). Modeling Math Identity and Math Success through Sentiment Analysis and Linguistic Features. *International Educational Data Mining Society*.
[7] Crues, R., Bosch, N., Anderson, C. J., Perry, M., Bhat, S., & Shaik, N. (2018). Who They Are and What They Want: Understanding the Reasons for MOOC Enrollment. *International Educational Data Mining Society*.
[8] DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2018). Detecting and Addressing Frustration in a Serious Game for Military Training. International Journal of Artificial Intelligence in Education, 28(2), 152–193. https://doi.org/10.1007/S40593-017-0152-1/FIGURES/9
[9] Dumbill, E. (2012). *Planning for big data*. Sonoma County, CA, USA: Sebastopol.
[10] Falakmasir, M., Yudelson, M., Ritter, S., & Koedinger, K. (2015). Spectral Bayesian Knowledge Tracing. *International Educational Data Mining Society*.
[11] Fancsali, S., Yudelson, M. V, Berman, S. R., & Ritter, S. (2018). Intelligent Instructional Hand Offs. EDM.
[12] Fesler, L., Dee, T., Baker, R., & Evans, B. (2019). Text as Data Methods for Education Research. Https://Doi.Org/10.1080/19345747.2019.1634168, 12(4), 707–727. https://doi.org/10.1080/19345747.2019.1634168
[13] Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., ... & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130-160.
[14] Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. Review of Research in Education, 44(1), 130–160. https://doi.org/10.3102/0091732X20903304
[15] Galyardt, A., & Goldin, I. (2015). Move Your Lamp Post: Recent Data Reflects Learner Knowledge Better than Older Data. *Journal of Educational Data Mining*, *7*(2), 83-108.
[16] Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113-128.
[17] González-Brenes, J., Huang, Y., & Brusilovsky, P. (2014). General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *The 7th international conference on educational data mining* (pp. 84-91). University of Pittsburgh.
[18] Guo, P. J., & Reinecke, K. (2014, March). Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 21-30).
[19] Gushchina, O., & Ochepovsky, A. (2019). Data mining for the e-learning risk management. Turkish Online Journal of Distance Education, 20(3), 181–196. https://doi.org/10.17718/TOJDE.601941
[20] Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6-47.

[21] Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015, March). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 156-165).

[22] Käser, T., Klingler, S., & Gross, M. (2016, April). When to stop? Towards universal instructional policies. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 289-298).

[23] Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing?. *arXiv preprint arXiv:1604.02416*.

[24] Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179).

[25] Klašnja-Milićević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education*, *25*(6), 1066-1078.

[26] Liu, R., & Koedinger, K. R. (2015). Variations in Learning Rate: Student Classification Based on Systematic Residual Error Patterns across Practice Opportunities. *International educational data mining society*.

[27] Lynn, N. D., & Emanuel, A. W. R. (2021). Using Data Mining Techniques to Predict Students' Performance. a Review. IOP Conference Series: Materials Science and Engineering, 1096(1), 012083. https://doi.org/10.1088/1757-899x/1096/1/012083

[28] Matas Terrón, A., Leiva Olivencia, J. J., & Núñez Rojas, N. (2020). El Big Data desde la perspectiva de una muestra de estudiantes de universidades del norte de Perú. *Propósitos y representaciones*, *8*(1).

[29] Méndez, G., Ochoa, X., & Chiluiza, K. (2014, March). Techniques for data-driven curriculum analysis. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 148-157).

[30] Nugent, A., Halper, F., Hurwitz, J. S., & Kaufman, M. (2013). *Big data for dummies*. John Wiley & Sons.

[31] Ochoa, X., & Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213-219.

[32] Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education*, *25*, 229-248.

[33] Ostrow, K., Donnelly, C., Adjei, S., & Heffernan, N. (2015, March). Improving student modeling through partial credit and problem difficulty. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 11-20).

[34] Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis--A New Alternative to Knowledge Tracing. *Online Submission*.

[35] Reich, J., Tingley, D. H., Leder-Luis, J., Roberts, M., & Stewart, B. M. (2014). Computer-assisted reading and discovery for student generated text in massive open online courses.

[36] Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7-12.

[37] Slater, S., Baker, R., Ocumpaugh, J., Inventado, P., Scupelli, P., & Heffernan, N. (2016). Semantic Features of Math Problems: Relationships to Student Learning and Engagement. *International Educational Data Mining Society*.

[38] Slater, S., Baker, R., Ocumpaugh, J., Inventado, P., Scupelli, P., & Heffernan, N. (2016). Semantic Features of Math Problems: Relationships to Student Learning and Engagement. *International Educational Data Mining Society*.

[39] Ward, Jonathan Stuart, and Adam Barker. "Undefined by data: a survey of big data definitions." *arXiv preprint arXiv:1309.5821* (2013).

[40] Yeung, C. K., & Yeung, D. Y. (2018, June). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1-10).

Appendix

Questionnaire on Big Data in Education: Assessment Scale of Big Data Applied to Education (VABIDAE)

This questionnaire aims to collect information about your opinion regarding Big Data. For this reason, we ask that you be as sincere as possible.

Big Data is data that is too big, too rapid, and beyond the analytical capabilities of traditional data base structures, including readily available data base administration tools and conventional data processing software. These vast amounts of data are required today, and they are being collected everywhere, in the fields of science, social networks, commerce, healthcare, government, media, education, and national security. From educational data mining strategies to SRL (self Regulated Learning) analysis, modeling the procedures and actions that students take in educational settings is a common stage. Studies examined hundreds of students in classroom settings and thousands of participants in (MOOCs), which has demonstrated how various student groups adopt various tactics that develop over time when utilizing the course resources. recognizing distinct MOOC engagement patterns and observing how various student groups interact with a learning simulation; leading to better performance, a personalized learning experience, and increased engagement.

Before starting, keep in mind the following conditions:

- Your **voluntary** participation is **requested**.

- No **private** or **sensitive** information is collected. In any case, the data will be used for academic and scientific purposes. Likewise, any information will be managed in accordance with current regulations.
- The questionnaire is only four short questions that will not take you more than 5 minutes.

Please indicate if you voluntarily agree to answer this questionnaire according to the previous conditions: -YES ☐ -NO ☐

[   ] Age:

Gender: MALE ☐ FEMALE ☐ Other ☐

Indicate, according to your criteria, what Big Data in Education could be used for:

	Not at all	I think not	I don't know	I think so	Totally Agree
Better serve the needs of students					
Improve academic results					
Personalize Education					
Improve employability					
avoid plagiarism					
Improve the organization of schools					
Improve teacher selection					
Produce educational resources adapted to students					
Facilitate decision-making at the political level					
Promote educational quality in general					
Helps prevent school failure					

	Not at all	I think not	I don't know	I think so	Totally Agree
Loss of student privacy.					
Loss of teacher privacy.					
Loss of school socialization.					
Computer attacks.					
Loss of teacher functions.					
Increase in the power of centre managers.					
Increasing the power of politicians (related to education).					
Produce educational resources adapted to students.					
System handling.					
Control of the education system by governments.					
Company control of the education system.					

Indicate, in your opinion, to what extent the following negative aspects could occur with Big Data:

Please, indicate to what extent the following statements are consistent with your attitude:

	Nothing at all	I think not	I don't know	I think so	Totally agree
The theme amuses me					
Gives me hope					
Makes me feel proud					
Makes me feel annoyed					
Makes me anxious					
I feel uncomfortable					
I feel negligent					
I find it pleasant					
I feel impotent					
Makes me bored					

What is your overall perception on Big Data in education or in general? All inputs are highly appreciated