Big Data Technology-Enabled Analytical Solution for Quality Assessment of Higher Education Systems

Samiya Khan[1], Mansaf Alam[4]
Department of Computer Science
Jamia Millia Islamia
New Delhi
India

Xiufeng Liu[2]
DTU Management
Technical University of Denmark
Lyngby
Denmark

Kashish Ara Shakil[3]
College of Computer and
Information Sciences, Princess Nora
Bint AbdulRahman University
Saudi Arabia

***Abstract*—Educational Intelligence is a broad area of big data analytical applications that make use of big data technologies for implementation of solutions for education and research. This paper demonstrates the designing, development and deployment of an educational intelligence application for real-world scenarios. Firstly, a quality assessment framework for higher education systems that evaluate institutions on the basis of performance of outgoing students was proposed. Secondly, big data enabled technological setup was used for its implementation. Literature was surveyed to evaluate existing quality frameworks. Most existing quality assessment systems take into account the dimensions related to inputs, processes and outputs, but they tend to ignore the perspective that assesses the institution on the basis of outcome of the educational process. This paper demonstrates the use of outcome perspective to compute quality metrics and create visual analytics. In order to implement and test the framework, R programming language and a cloud based big data technology that is Google, BigQuery were used.**

***Keywords*—Education big data; educational intelligence; educational technology; higher education; quality education**

I. Introduction

The concept of educational intelligence [1] was introduced as an umbrella term for all analytical solutions created for education and research sectors. A range of applications have been proposed in recent times, ranging from applications for improving the operational efficiency of educational and research institutes, to specific predictive analytical applications for foretelling student dropout rates and prescriptive analytical solutions to improve the quality of education. It is important to note that a majority of the proposed solutions have not been prototyped or implemented. Firstly, this paper proposes an outcome-based quality assessment framework for higher education systems. Secondly, it demonstrates the implementation of an educational intelligence solution with the help of base technologies used for big data storage and processing.

Higher education is the backbone of the education system of any country. Effective management and quality assessment of the higher education system is not just important, but it is also necessary. However, the concept of quality in higher education has found varied definitions and descriptions in literature. The most recent and widely accepted definition of quality describes it as conformance of standards and meeting the set objectives. Besides this, there are many other definitions that incorporate the perspectives of different stakeholders like students, organization and parents, in addition to others. However, none of the dimensions covers the outcome perspective of the education system. The proposed framework evaluates quality from this perspective.

The key process involved in higher education systems is the process of imparting education or the education process. The outcome of this process is determined by the performance of students after passing out of the course concerned, which is assessed using the information about the university or company that student joins after course completion. Quality score is computed using this information and analytics are generated on the basis of the cumulative study of these quality scores.

The analytical framework proposed in this paper can be used for evaluating the performance of an educational organization on the basis of the cumulative quality scores' analysis of the students who pass out in a year. Moreover, predictive analysis can also be generated to monitor progress and make interventions as and when required, to maintain quality of the educational organizations and system, at large.

There are several ways in which such a quality metric may be relevant. Most quality metrics assume that the responsibility of an educational institution for a student's performance is restricted to the time that the student concerned spends while he or she is enrolled in the institution. However, the responsibility of the student's performance on his or her alma mater extends beyond this timeframe. This is perhaps the reason why organizations take pride in their alumni networks and achievers who hail from their institutes. Therefore, a quality evaluation basis that assesses the organization's performance on the basis of student performance after he or she completes the course is considerable.

Inclusion of this quality dimension to the assessment framework will bring the attention of organizations to this aspect. Quality assurance in this regard shall boost postgraduate and research enrollments, promoting higher and advanced studies. Besides this, such an assessment will also drive educational institutions to groom students for prospective future degrees or jobs, bridging the gap that usually exists between these transitions. In entirety, this quality dimension will bring educational organizations a step closer to fulfilling their purpose, meeting institutional vision and conforming to standards in the outcomes context.

In countries like India, which boast of 34,211 thousand enrollments [1] for the academic year 2014-15 in higher education, the amount of data collected is immense. This data is high in volume, contains images and textual data for variety and is generated on a yearly basis. Moreover, the use of data mining and machine learning in educational analytics [2] makes big data technologies a relevant solution for educational analytics [3]. When projected in the global domain, education data is certainly a 'big' data set.

Owing to the volume, variety and velocity of data generated by the education system, education data can easily be termed as a class of big data called 'big education data'. Some of the information that is recorded as part of this system includes profile of teachers, students and operational data. Therefore, big data technologies can be used to develop educational intelligence solutions. In line with this, the case study done for implementing the proposed framework and providing a proof of concept makes use of Google BigQuery [4] and R programming language [5] to generate analytics.

The motivation behind this paper is to propose a simple technological framework that makes use of base technologies for development of educational intelligence applications. Therefore, this paper demonstrates the designing, development and deployment of an educational intelligence application. The contributions of this paper include the following: (1) proposes an outcome-based quality assessment framework for higher education systems (2) implements an educational intelligence application that use the proposed framework with the help of base technologies (3) deploys the educational intelligence application on cloud for use.

The rest of the paper has been organized in the following manner: Section II provides insights on related work for educational intelligence applications, their viability for higher education systems and the concept of quality and its definition in the higher education context. Section III provides details about the proposed framework. Section IV and Section V describe how big data analytics solutions can be used for handling big data generated by higher education systems and provides a case study of how this framework can be implemented using cloud-based big data technology, BigQuery, and R programming. Lastly, the paper concludes in Section VI and Section VII providing insights on scope for improvements and future work.

II. Related Work

The applicability of big data technologies [6], implemented using cloud-based infrastructures, to real-world, data-intensive scenarios has given rise to many sub-research areas. Big data has particularly found applications in healthcare [7], geospatial analytics [8] and business intelligence [9], in addition to many others. Education data is a form of big data and can make use of big data technologies for generating valuable analytics for an educational organization [10]. González-Martínez et al. [11] investigated the use of cloud computing in education, citing it as one of the key new-age technologies to have found applications in education.

IBM's big data model defines five big data characteristics [12], namely, volume, velocity, variety, veracity and value. Evidently, data collected at the student level is 'big' in 'volume', both on a per-year basis as well as over-the-years. Student profiles include textual and image data like scanned copy of the student's photograph and signature. The different types of data included in the collection makes up for the 'variety' characteristic of big data. Moreover, this data is generated on a yearly basis, accounting for 'velocity' of data.

Student data used for processing is manually entered. As a result, the probability of error and associated uncertainty are rather high, which makes veracity a significant characteristic in the education system context. Lastly, educational intelligence solutions [1] can be used for improving the operational efficiency of the system and support administrative processes with improved decision-making, adding value to the data and analytical solutions that it can produce. Educational data mining [13] has also been used for predicting student performance [14], which is one of the typical scenarios that can be converted into a big data case study.

For the purpose of feasibility evaluation of educational intelligence solutions for real world education systems, the Indian higher education system has been considered. The numbers of universities and institutions that can utilize educational intelligence solutions have been given in Table I, with a breakdown of the different types of organizations that it entails. This analysis can be projected to the global scenario to conclude that the education data set is indeed 'big'. Besides this, the total number of students pursuing different courses in the Indian higher education system has been detailed in Table II.

TABLE. I. Number of Higher Education Institutions by Type 2014-15[1]

Level	Type	Total
University	Central University	43
	State Public University	316
	Deemed University	122
	State Private University	181
	Central Open University	1
	State Open University	13
	Institution of National Importance	75
	State Private Open University	1
	Institutions Under State Legislature Act	5
	Others	3
	Total	760
College		38498
Standalone Institutions	Diploma Level Technical	3845
	PGDM	431
	Diploma Level Nursing	3114
	Diploma Level Teacher Training	4730
	Institute Under Ministries	156
	Total	12276

[1] http://mhrd.gov.in/statist

TABLE. II. LEVEL-WISE ENROLMENT IN SCHOOL & HIGHER EDUCATION 2014-15[1]

Level	Total (in thousand)
Ph.D.	118
M. Phil.	33
Postgraduate	3853
Undergraduate	27172
PG Diploma	215
Diploma	2508
Certificate	170
Integrated	142
Higher Education Total	34211

The numbers of institutions that can utilize such solutions are high. This makes educational intelligence solutions commercially viable. Moreover, cloud-based big data technologies are deemed most appropriate for developing analytical solutions for education and research [15]. The integration of the analytics derived from such solutions can be integrated with insights to get value-based solutions [16].

The standard meaning of quality is a measurement of standard or excellence. In higher education, the concept of quality emerged in early 1980s and was derived from its commercial counterparts [17]. However, academic quality was considered an abstract term in those days. The traditional concept of quality [18] was inferred from the fact that world-class universities like Harvard and Oxford were considered benchmarks and no further dissection on the dimensions of quality was done.

Green [18] gave an extensive analysis of the literature that defines the term quality and categorized them into the following five approaches:

1) Conforming to standards in terms of the educational process and outcomes

2) Befitting the purpose: This is a contradictory definition as most scholars feel that if the institution meets standards, it fits its purpose, which may not always be the case.

3) Ability to meet set institutional goals and having a clear vision.

4) Meeting the needs of the customer or student: It is important to state here that the student can be considered a product, customer or both by the higher education institution. While the student is paying for getting education, which makes him or her the customer, it is the student's performance that will determine the quality of the institute, making him also the product.

5) The traditional concept defines quality as strive for excellence.

Recent definitions of quality have added dimensions like ethics and moral values [19] and accountability and accreditation [20] to the core system. Owlia and Aspinwall [21] gave a conceptual framework for measuring quality on the basis of dimensions that take student and administrative perspectives into account. It caters for dimensions from a service-oriented point of view. Another perspective that needs to be taken into account while defining quality is that of the stakeholders. In the higher education context, academics' and students are the main internal stakeholders, in addition to other stakeholders like the state, society, employers, parents and professional associations. Student's perspective on quality [22] is considered extremely important.

Related studies that cover higher education service quality dimensions include Parsuraman [23], Gronroos [24], Lehtinen and Lehtinen [25], Carney [26], Athiyaman [27], Lee et al. [28], Hadikoemoro [29], O'Neill and Palmer [30], Sahney et al. [31], Brooks [32] and Teeroovengadum et al. [33]. A comparison of the dimensions proposed by each of these works and the perspective they cover has been provided in Table III.

TABLE. III. COMPARISON OF QUALITY ASSESSMENT FRAMEWORKS FOR HIGHER EDUCATION SYSTEMS

Research Work	Brief Description	Dimensions Considered	Perspective Covered
Parsuraman [23] and Parsuraman [34]	Parsuraman [23] proposed 10 dimensions, which were later classified into five categories by Parsuraman [34]. Among the identified dimensions, empathy was considered least important while reliability was placed highest on the list.	The categories and corresponding dimensions for them are as follows - • Tangibles - This category includes dimensions like equipment, physical facilities and appearance of personnel. • Reliability - This category includes the ability of staff to provide the service accurately and consistently. • Responsiveness - This category includes the ability of staff to respond to demands of students and help them. • Assurance - This category includes dimensions like knowledge of employees and their ability to communicate trust and confidence. • Empathy - This category includes dimensions like the ability of staff to give individual attention and care for students.	Takes into account only the student perspective and considers student as a customer.
Gronroos [24]	Identified six criteria for classifying perceived service quality as good.	The identified criteria are as follows - • Skill and Professionalism - This criterion includes knowledge of staff and their ability to solve problems in a professional manner. • Flexibility and Access - This criterion includes dimensions like the ability of staff to adjust to the needs of students and their availability in times of need. • Behaviour and Attitude - This criterion includes dimensions like friendliness and genuine care for the needs of the student on part of the staff. • Reputation and Credibility - This criterion assesses the brand value of the institute by evaluating the perceived credibility and reputation of	Takes into account only the student perspective and considers student as a customer.

		the same. • Recovery - This criterion evaluates the ability of an institute to recover from an issue or problem.	
Lehtinen and Lehtinen [25]	Classified quality dimensions in three categories namely interactive quality, physical quality and corporate quality.	The dimensions included under each category are similar to those specified by Gronroos [13]. The categories are as follows - • Interactive quality - The quality dimensions, which are associated with the interaction between the institution and student are included under this category • Physical quality - The quality dimensions included under this category are measured regardless of the opinion of the student. • Corporate quality - The quality dimensions associated with the reputation and branding of the institute are included under this category. These dimensions take both the customer or student's perspective as well as the institution's perspective into account.	Takes into account the student perspective and considers student as a customer. However, this framework also considers dimensions that are independent of student perspective and can be seen as administrative perspective.
Carney [26]	Proposed a comprehensive set of attributes for studying the image of a college.	The variables identified relevant to the context include - • Student qualities • Student qualification • Quality of instruction • Faculty-student interaction • Academic reputation • Variety of courses run by the college • Career preparation • Class size • Student Activities • Athletic programs • Facilities and equipment • Community service • Physical appearance of the campus • Location • Friendly nature of staff • On-campus residence • Religious atmosphere • Caring atmosphere • Financial aid • Safety in campus	Takes into account the student and administrative perspective. Moreover, dimensions like student qualification and career preparation hint towards the impact of education on the outcome of the educational process, but there is no clear dimension catering to the outcome perspective.
Owlia and Aspinwall [21]	Gave a conceptual framework.	The six dimensions identified by this framework comprise of the following – • Tangibles which include infrastructure, ease of access and supporting infrastructure and facilities • Competence which includes student-staff ratio and quality of staff along with their ability to communicate effectively with students • Attitude which includes guidance and willingness to help • Content which includes curriculum, cross-disciplinarily of knowledge and relevance of courses for future jobs • Delivery which includes effective communication, student feedback and providing encouragement to students • Reliability which includes matching goals and handling complaints	Takes into account the student and administrative perspectives.
Athiyaman [27]	Proposed a framework containing eight characteristics for evaluating the services provided by a university.	The characteristics include - • Availability of staff for solving students' problems • Teaching skills of staff and their ability to teach the students well • Computing facilities • Class size • Library facilities • Recreational facilities • Student workload • Difficulty level of course content	Takes into account the student perspective.
Lee et al. [28]	This research work explained that there are two variables that can determine and predict the overall satisfaction level of the student with respect to the provided service.	The two variables include - • Overall impression of educational quality • Overall impression of school	Takes into account the student perspective.
Hadikoemoro	This research work identified	The identified dimensions are as follows:	Takes into account the

[29]	35 service-quality items based on interviews conducted for two groups and identified five dimensions.	• Academic Services - This dimension includes completeness of the infrastructure and facilities and accurate and dependable delivery of services. • Attentiveness and Readiness - This dimension assesses the ability of the university to help students and be promptly available fir any consultation • Fair and Impartial - This dimension includes the discipline maintenance on the campus and the ability of the institution authorities to implement a democratic campus regulation. • Tangible - This dimension includes equipment, facilities and the physical appearance of the campus. • General Attitude - The ability to handle students falls under this dimension of this framework.	student and administrative perspectives.
O'Neill and Palmer [30]	This research assessed the importance of dimensions for quality assessment based on student perceptions and it was found that process was the most important dimension. However, the analysis is exploratory and thus, a further analysis is required to establish the results.	The dimensions used by this research work are as follows: • Process - This dimension includes factors that assess the ability of staff to solve student's problems, their accessibility and availability and their ability to deal with students. Basically, the factors included in this dimension assess the quality of the process. • Empathy - The ability of employees to understand student needs and the amount of attention they are able to pay each student are assessed under this dimension. • Tangible - The facilities, equipment and physical appearance of the institution are assessed under this dimension.	Takes the students' perspective into account.
Sahney et al. [31]	The proposed framework emphasised on the need to identify design requirements of the system in addition to customer requirements for assessing quality. Moreover, it also focussed on measuring quality by comparing the perceptions of students with regard to the design requirements and expectations from the system.	The factors identified by the framework include: • Competence of the system - This factor refers to the infrastructure, faculty expertise and availability of skilled teaching and support staff. • Attitude of the staff - This factor refers to the attention they pay towards maintaining a healthy environment and remain accessible to students for consultation • Content - This factor refers to the course curriculum and future needs of the students from the course. Besides this flexibility of the course to adjust to interdisciplinary knowledge is also considered as part of this factor. • Delivery of content - This factor refers to the ability of staff to manage classroom and deliver content in a manner that makes the students feel at ease and comfortable. • Reliability - This factor includes dimensions like the clarity of course objectives, rules and regulations and guidelines of the institute.	Takes into account the student and administrative perspectives.
Brooks [32]	This research work emphasised on the inclusion of university activities for measurement of quality and gave a criterion for quality assessment.	The dimensions included in the proposed criteria are as follows: • Reputation • Faculty research productivity • Student educational experiences and outcomes - This dimension is further divided into program characteristics, program effectiveness and student outcome.	This framework takes into account the administrative perspective. However, there is a minor deviation and indication towards a third perspective called the outcome perspective. The student outcome dimension assesses the learning and career outcome of the student. This dimension covers the score and grades obtained by the student in the examination to assess the outcome of the process.
Teeroovengadum et al. [33]	Gave a factor/measurement model that was based on qualitative analysis of collected data, in addition to survey of existing literature. The hierarchical model proposed for measurement of service quality for higher educational institutions. This framework is also considered an extension of Owlia and Aspinwall [21].	The primary identified dimensions were as follows: • Administrative quality - This dimension has two sub-dimensions namely administrative processes and attitude and behaviours of administrative staff. • Physical environment quality - This dimension has three sub-dimensions namely general infrastructure, learning setting and support infrastructure. • Core educational quality - This dimension has four sub-dimensions namely pedagogy, competency, curriculum and attitude and behaviour. • Support facilities Quality • Transformative Quality	Takes into account only the student perspective and considers student as a customer.

III. Proposed Quality Framework

While most of the existing frameworks considered student's perspective and/or administrative perspective for defining quality dimensions, only Carney [26] and Brooks [32] hinted on the use of outcome perspective for measuring quality. Telford and Masson [35] emphasized on the importance of considering dimensions related to educational process as a key facet for measuring quality. With that said, computing the outcome of the educational process is one of the key measures for evaluating the education process.

In view of the definition of quality for higher education given in the previous section, it can be understood that a student is not just the customer of the system; he or she is also the product. Besides this, imparting education is a process and the performance of student in real-world scenario after the completion of the course, is the 'outcome' of the process. The proposed framework measures quality by computing quality scores for every student who passes out of a university in a given academic year. This score is calculated using the ranking of the organization that a student joins after passing out.

These scores are cumulatively analyzed to assess the average quality for the organization. Moreover, a year-wise analysis can provide trends and predictions in this regard. The content dimension of quality, which includes 'relevance of courses for future jobs' given by Owlia and Aspinwall [21], 'career preparation' by Carney [26], 'student outcome' by Brooks [32] and 'career prospect' objective mentioned by Teeroovengadum et al. [33] are quantitatively evaluated using the quality score. Consequently, quality monitoring can be performed using quantitative analytics.

A. Higher Education System Inputs and Outputs

Higher education system can be broken into three academic categories namely, undergraduate, postgraduate and research. The progress pathway of a student from one academic category to the next is shown in Fig. 1. A student admitted to undergraduate courses in a university, upon completion, may either choose to join a postgraduate course in the same or another university, take up a job or not pursue anything at all.

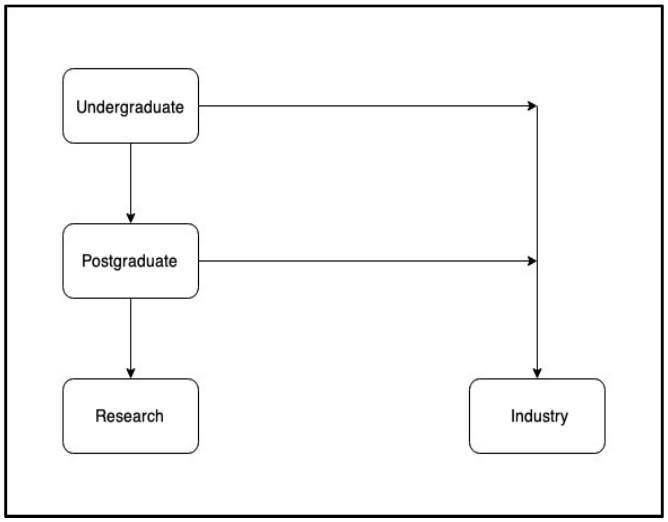


Fig. 1. Breakdown of Higher Education System.

Similarly, a postgraduate student may choose to take up a job, pursue research in the same or a different university or not pursue anything at all. Evidently, there are two transitional states. The first state of transition (Transition 1) is when a student completes an undergraduate course and takes up either a postgraduate course or joins industry. However, the state when a postgraduate student completes a course can be considered the second state of transition (Transition 2). This transition diagram forms the basis of the proposed quality assessment framework.

From Fig. 1, three academic categories of students can be formulated, including Students Pursuing Higher Education (SPHE), Students Opting For Jobs (SOFJ) and Students With No Data Available (SWNDA). SPHE includes the students who choose to pursue a postgraduate degree after undergraduate degree or research after postgraduate degree. SOFJ includes the students who opt for on-campus placements or get an off-campus placement in the following academic year. Therefore, all students who are able to find an industry position within one year after graduation are considered under this academic category.

The last academic category, SWNDA, includes the students who do not fall into the other two academic categories. Since, this is a yearly analysis, any student who finds an off-campus placement after one year, chooses self-employment or does nothing at all for the first year after graduation, is considered under this academic category. It is important to mention that this academic category is added for comprehensiveness. However, for the sake of simplicity, the quality score for this academic category of students is taken as zero.

Understandably, self-employment is a special case scenario. However, even if the student takes up self-employment within the first year of leaving the university or institution, there is no parameter to judge the success of his or her venture in a given time period. Therefore, a more detailed framework is required for calculating the quality score for the self-employment.

B. Quality Score (QS) - Metric for Quality Assessment

This paper proposes a metric termed as 'quality score', which shall be calculated at student level on the basis of the transition outcome of the student from one state to another. The quality scores for students enrolled to a university are calculated year-wise on a scale of 1 to 10 and used for further analytics. Quality score calculations for SPHE and SOFJ academic categories of students will be provided in the following sections.

As mentioned previously, students, upon completion of their undergraduate or postgraduate degrees, are expected to pursue higher education or opt for campus placements. University or institute rankings are provided by Government organizations and private ranking agencies on a yearly basis.

In order to ensure and maintain authenticity of the base data used for analysis, the ranking provided by Government agencies is recommended for SPHE. Moreover, a student may move to a university in the same country or may opt to study abroad. In order to accommodate this case, the data for world university ranking must be taken for score calculation if the

student is taking admission abroad, otherwise country-wise ranking can be applied in the other case to accommodate for maximum universities.

In order to calculate QS for every student, the rank of the university in which the student is taking up a postgraduate degree or pursuing research and the maximum rank assigned by the ranking to any university are required as inputs. If the university in which the student is taking up postgraduate degree or research is not ranked in the list, the value zero is assigned to quality score for that student.

The value of QS for a SPHE student with a known university rank is calculated by performing linear scaling. The formula used for linear scaling[2] are given in equations (1), (2) and (3).

$$rate = \frac{scaled_{max} - scaled_{min}}{input_{max} - input_{min}} \quad (1)$$

$$offset = scaled_{min} - (input_{min} * rate) \quad (2)$$

$$ouput = (input * rate) + offset \quad (3)$$

The derivation of the formula for QS calculation in this scenario is given below. The value of variables used in equations (1), (2) and (3) are as follows:

$$input = rank$$

$$output = QS_{SPHE}$$

$$input_{min} = rank_{max}$$

$$input_{max} = 1$$

$$scaled_{min} = 1$$

$$scaled_{max} = 10$$

Substituting these values in equations (1), (2) and (3):

$$rate = \frac{(10-1)}{1-rank_{max}}$$

$$rate = \frac{9}{1-rank_{max}} \quad (4)$$

$$offset = 1 - (rank_{max} * rate) \quad (5)$$

$$QS_{SPHE} = (rank * rate) + offset \quad (6)$$

A description of the variables used in the described formula is given in Table IV.

Placement data like company ranking and package offered can be cumulatively used as base data for the students who opt for campus placed jobs. The company rankings can be taken from survey results of credible private agencies like Economic Times[3], which have created surveys with the objective to create top recruiters list. Data for package offered by companies to individual students is available with the university and can be directly used for analysis.

Given the fact that the best student performer in this academic category is the one who gets placed in a company with the highest ranking and gets the highest package. On the other hand, the worst performer is the one who gets placed in an unranked company at the lowest package. In order to compute the total quality score for SOFJ, QS calculated on the based of industry rankings is added to QS calculated on the basis of relative package score.

A description of the variables used in the described formula is given in Table V.

TABLE. IV. VARIABLES USED FOR SPHE QUALITY SCORE CALCULATION

Variable	Description
$input$	The input value that needs to be scaled
$output$	The scaled value
$input_{min}$	The minimum value of the input scale
$input_{max}$	The maximum value of the input scale
$scaled_{min}$	The minimum value of the output scale
$scaled_{max}$	The maximum value of the output scale
$rate$	The rate of scaling
$offset$	The offset that needs to be applied for scaling
$rank$	The university rank for the concerned SPHE student
$rank_{max}$	The maximum rank assigned to a university in the University Ranking data used
QS_{SPHE}	Quality Score for SPHE of the concerned student

TABLE. V. VARIABLES USED FOR SOFJ QUALITY SCORE CALCULATION

Variable	Description
$input$	The input value that needs to be scaled
$output$	The scaled value
$input_{min}$	The minimum value of the input scale
$input_{max}$	The maximum value of the input scale
$scaled_{min}$	The minimum value of the output scale
$scaled_{max}$	The maximum value of the output scale
$rate$	The rate of scaling
$offset$	The offset that needs to be applied for scaling
$industry\_rank$	The company rank for the concerned SOFJ student
$industry\_rank_{max}$	The maximum rank assigned to a company in the Company Ranking data used
QS_{IR}	Quality Score for SOFJ of the concerned student based on company ranking
$package_{min}$	The minimum package that has been offered to any SOFJ student
$package_{max}$	The maximum package that has been offered to any SOFJ student
$package$	Package offered to the concerned SOFJ student
QS_{PO}	Quality Score for SOFJ of the concerned student based on package offered
QS_{SOFJ}	Cumulative Quality Score for SOFJ of the concerned student

[2] https://www.courses.psu.edu/e_met/e_met430_jar14/pid/ioscal.html
[3]http://economictimes.indiatimes.com/

Quality score computed on the basis of industry ranking (QS_{IR}) makes use of the same concept as used by QS_{SPHE}. If the company in which the student is taking a campus placement is not ranked in the list, the value zero is assigned to for QS_{IR} of that student. In order to accommodate this case, the value of QS_{IR} is calculated on a scale of 1 to 5. The value of QS_{IR} can be calculated using the equations (1), (2) and (3) with the following parametric values:

$$input = industry\_rank$$

$$output = QS_{IR}$$

$$input_{min} = industry\_rank_{max}$$

$$input_{max} = 1$$

$$scaled_{min} = 1$$

$$scaled_{max} = 5$$

Substituting these values in equations (1), (2) and (3),

$$rate = \frac{(5-1)}{1-industry\_rank_{max}}$$

$$rate = \frac{4}{1-industry\_rank_{max}} \quad (7)$$

$$offset = 1 - (industry\_rank_{max} * rate) \quad (8)$$

$$QS_{IR} = (industry\_rank * rate) + offset \quad (9)$$

Scaling of package offered to students opting for campus-placed jobs on a scale of 1 to 5 also requires linear scaling. Therefore, equations (1), (2) and (3) are used with the following parameters:

$$input = package$$

$$output = QS_{PO}$$

$$input_{min} = package_{min}$$

$$input_{max} = package_{max}$$

$$scaled_{min} = 0$$

$$scaled_{max} = 5$$

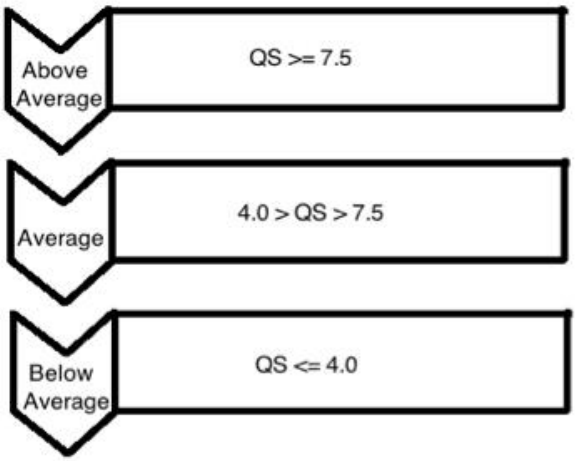


Fig. 2. QS-based Quality Framework.

Substituting these values in equations (1), (2) and (3),

$$rate = \frac{(5-0)}{package_{max}-package_{min}}$$

$$rate = \frac{5}{package_{max}-package_{min}} \quad (10)$$

$$offset = -(package_{min} * rate) \quad (11)$$

$$QS_{PO} = (package * rate) + offset \quad (12)$$

Quality score for SOFJ (QS_{SPHE}) is determined using the following equation:

$$QS_{SOFJ} = QS_{IR} + QS_{PO} \quad (13)$$

C. Quality Score (QS) – Based Analytics

Once quality score is determined for the students passing out of an institute or university at the end of a specific academic year, these values can be used to generate varied types of analytics and graphical interpretations. In order to use QS for decision-making and organizational efficiency management, students need to be divided into three QS categories Fig. 2 illustrates the framework that can be used for student categorization.

On the basis of this categorization, average QS for each category can be determined. Moreover, graphical illustrations like pie charts indicating the share of 'Above Average', 'Average' and 'Below Average' to indicate organization performance for a specific year can be created. Besides this, analytics that use data generated over the years can be used to create line charts for indicating the performance patterns of the organization in all the three categories. These analytics can be used to assess the performance of an institute or organization for an academic year as well as over the years. Cumulatively, these visualizations can be used for performance assessment and comparison.

IV. Methodology and Implementation

The proposed framework utilizes student data profiles to compute quality score per student and use the computed values for advanced analytics. Profiles of students passing out each year are scanned for computing quality scores and quality score data per year is stored for generating time-based analytics.

In order to implement and test the proposed framework, Google BigQuery[4] and R[5] are used as base technologies. A web-based application was developed using shiny package[6] available for R. The backend programming for computation of quality score and creation of visualizations like pie charts and line charts was performed using R programming language.

Dummy data set has been used for testing and validating the framework. The data set used for computation was stored in Google BigQuery, which is a cloud-based big data-warehousing technology. The schema for the three tables used for the implementation of this framework has been shown in Table VI, Table VII and Table VIII.

[4] https://cloud.google.com/bigquery/
[5] https://www.r-project.org/about.html
[6] https://cran.r-project.org/web/packages/shiny/index.html

There are a plethora of big data technologies and tools [36] available in the form of commercial as well as open-source solutions. The choice of these technologies was made because of the cost-effective nature of Google BigQuery and the simplicity of R language for developing analytical visualizations. Moreover, the availability of BigQuery API allows easy integration of Google BigQuery with R programming environment. This is also one of the reasons for this selection.

TABLE. VI. TABLE SCHEMA FOR STUDENT DATA TABLE

Field Name	Data Type	Description
course	STRING	Course to which the student is enrolled
eyear	INTEGER	Year of enrolment
code	STRING	Code of course to which the student is enrolled
Id	INTEGER	Student ID
Gender	STRING	Gender
Region	STRING	Region to which the student belongs
He	STRING	Highest education
Imd	STRING	IMD Band
Age	STRING	Age Bracket to which the student belongs
prev_attempt	STRING	Number of previous attempts taken
Credit	STRING	Credits studied
Disability	STRING	Whether suffering from a disability
final_result	STRING	Final result (Pass or Fail or Withdrawn)
Univ	STRING	University to which student has taken admission after course completion
Comp	STRING	Company that the student has joined after course completion
Package	FLOAT	Package offered
univ_f	STRING	Whether joined a university after course completion (Y/N)
comp_f	STRING	Whether joined a company after course completion (Y/N)
q_score	FLOAT	Quality score (assigned to zero for initialization)

The developed tool is named "Quality Management Tool for Higher Education Systems". The user interface of the web application is shown in Fig. 3. It is important to mention that the application uses server-side authentication. Therefore, in order to make the application work, the user will have to provide the JSON file with authentication token, which can be downloaded from the user's Google BigQuery Account. The application was firstly deployed on the local server and tested. Consequently, it was deployed on the Shiny server and tested. This application is available on Shiny Server, at https://qmhes.shinyapps.io/qmhes/.

TABLE. VII. TABLE SCHEMA FOR UNIVERSITY RANK TABLE

Field Name	**Data Type**	**Description**
univ_code	STRING	University code
univ_name	STRING	Name of the university
univ_city	STRING	City of location
univ_state	STRING	State of location
univ_score	FLOAT	Score
univ_rank	INTEGER	Rank
uryear	INTEGER	Year in which rank was generated

TABLE. VIII. TABLE SCHEMA FOR COMPANY RANK TABLE

Field Name	**Data Type**	**Description**
comp_name	STRING	Name of the company
comp_sector	STRING	Sector of operation
comp_subsector	STRING	Sub-sector of operation
comp_area	STRING	Area/Continent of operation
comp_country	STRING	Country of operation
comp_para1	FLOAT	Financial parameter
comp_para2	FLOAT	Financial parameter
comp_para3	FLOAT	Financial parameter
comp_para4	FLOAT	Financial parameter
comp_rank	INTEGER	Rank
Cryear	INTEGER	Year for which rank was generated

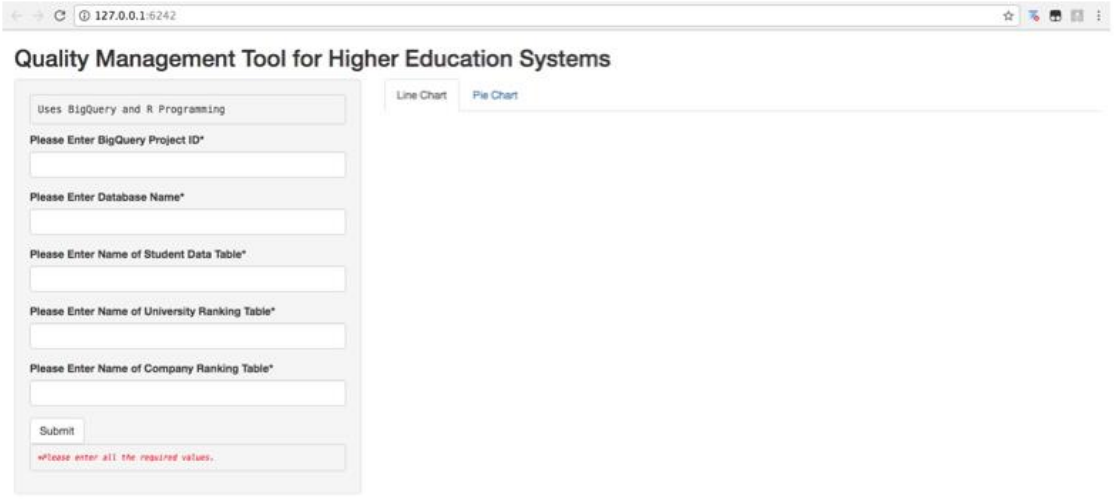


Fig. 3. User Interface of 'Quality Management Tool for Higher Education Systems'.

V. EVALUATION

The dataset used for testing the implementation of this framework contains dummy data for students, university rankings and company rankings for the years 2013 and 2014. Quality score for every student is calculated on the basis of whether the student joins a research degree or gets a job in a company. The rank of the university determines the quality score of students joining a university.

On the other hand, company ranking and package offered determine the quality score of students joining a company. The package offered to a student is the annual compensation that a company is willing to offer the student upon joining. The computation of quality score for the data set used has been illustrated in Fig. 4 and Fig. 5.

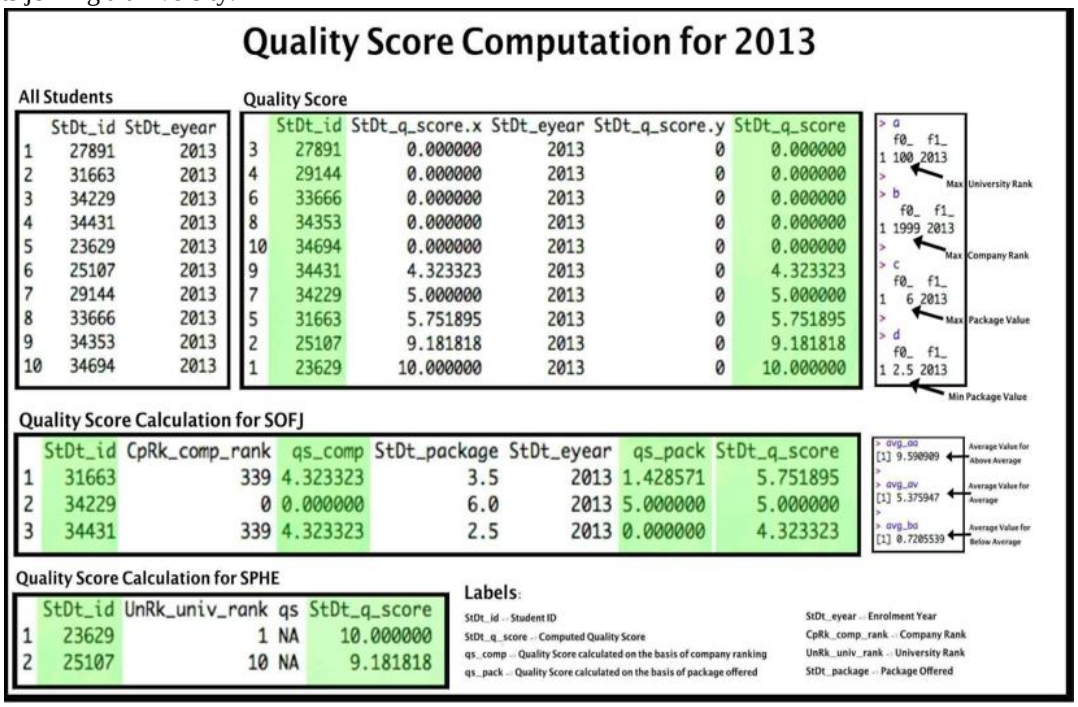


Fig. 4. Computation of Quality Scores for Students enrolled in the Year 2013.

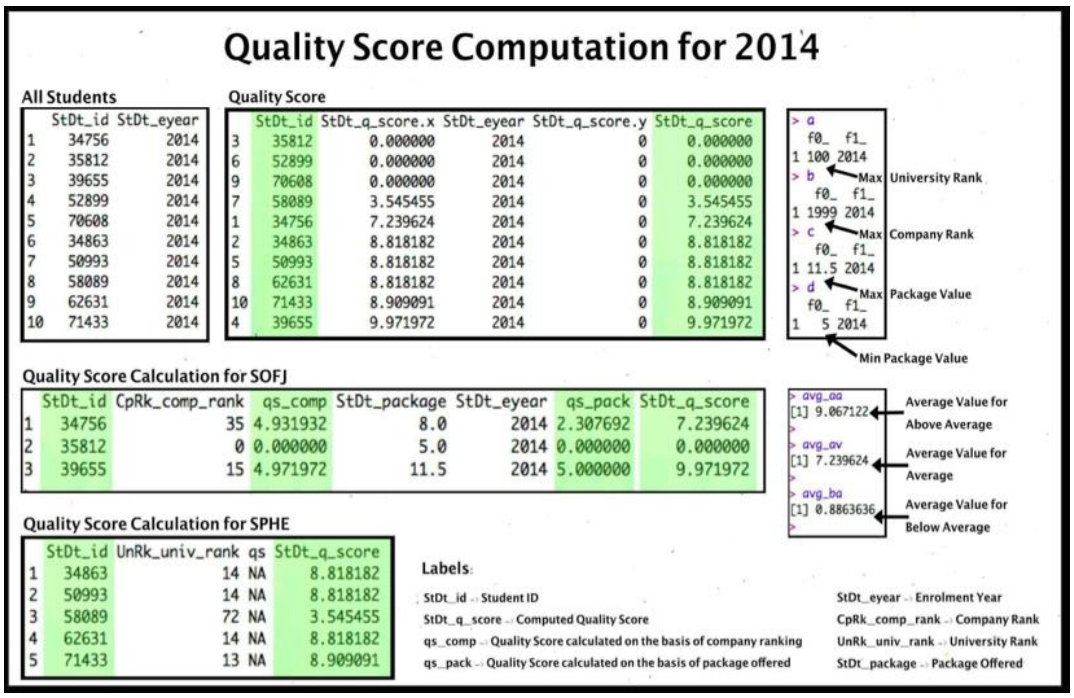
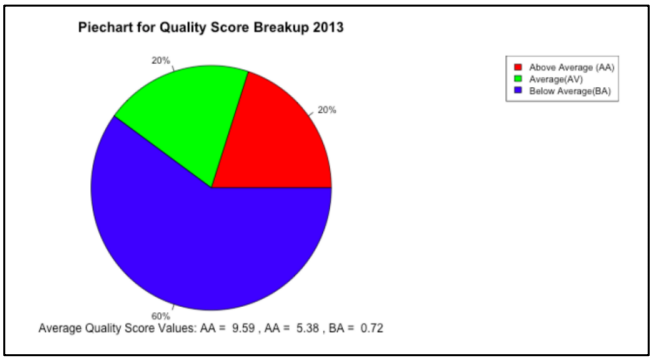
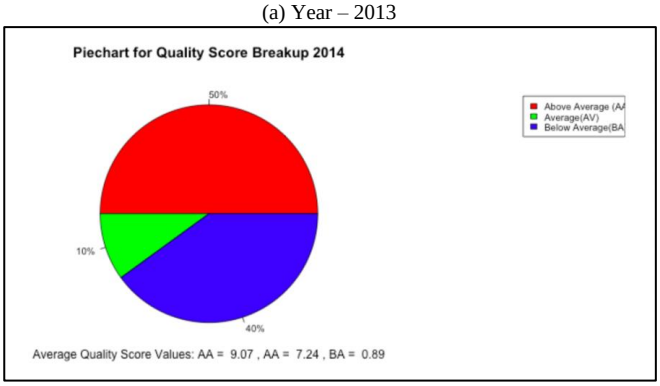


Fig. 5. Computation of Quality Scores for Students enrolled in the Year 2014.



(a) Year – 2013



(b) Year - 2014

Fig. 6. PieChart Generated for Different Years

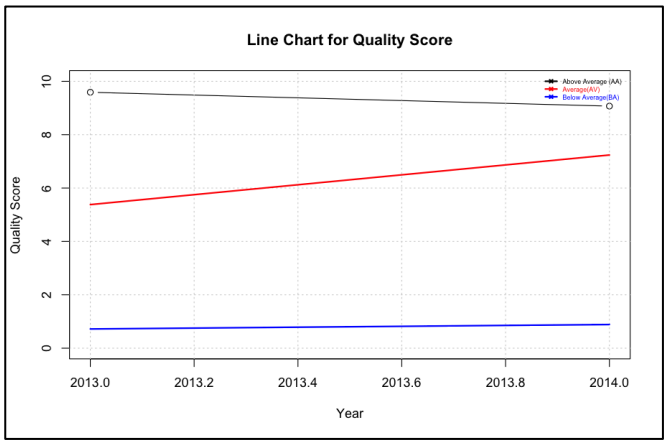


Fig. 7. Line Chart Generated.

On the basis of the quality scores generated for every student passing out of the university during one academic year, average quality scores are calculated. These average values are used to generate line charts for demonstrating trends of quality score values for an educational organization. On the other hand, the numbers of students falling under specific QS categories are used to generate piecharts for specific years. The analytics–pie charts and line chart – created using quality score data computed for the dataset used are shown in Fig. 6(a), Fig. 6(b) and Fig. 7. The inputs required to generate the line chart include project's BigQuery Project ID, name of the database (inside the BigQuery project) and names for tables containing data related to students, university rankings and company rankings. In order to generate year-specific pie charts, an addition field of year also needs to be filled.

VI. Discussion

Data about students, who get jobs through off-campus placement immediately after the degree or after taking a gap year, is usually not available. Moreover, data collection related to students who opt for self-employment is also limited. However, off-campus placements within one year of course completion are only considered because the framework generates yearly analysis. This limits the capabilities of the proposed analytical framework by restricting the amount and diversity of data available for analysis. A more robust data collection at the organization level or alumni association levels can be a significant step towards a more efficient outcome-based analysis in the educational context. Moreover, self-employment is not accounted for in the given framework.

This paper is an exploratory study, which encompasses several limitations. The QS is based on the institutions and companies' rankings, so the proposed framework assesses the quality of Higher Education Institutions (HEIs) based on rankings. The base dataset of rankings used for QS calculation is and must be taken from credible sources like Government agencies and reputed private organizations. This can be a controversial way to "measure" organizations' quality. The computation of the QS assumes that the type of companies they enroll in after graduation measures the 'quality' of the graduates. So, best students will have a job in higher rank companies. But this may not always be true. A graduate may find a job in an excellent company on the basis of references and not because he or she is a high quality student. Looking at such limitations of using the proposed indexes for quality assessment, the use of these indexes along with other quality metrics is recommended for a more comprehensive evaluation.

Therefore, this study intends to contribute to the discussion on the possibilities of enlarging QA schemes to make them abler to include outcome perspective along with both the customer and product perspectives. Usually the frameworks include dimensions related to inputs, processes and outputs, but tend to overlook the outcomes, that is what happens after the students graduate and leave institutions. The innovativeness of the proposed concept lies in adding a new dimension to QA schemes and implementing the same using big data technologies. This study demonstrates how big data technologies can be used to develop educational intelligence applications.

VII. Conclusion

The next generation of analytical solutions for education and research are drifting towards educational intelligence. Big data technologies are put to use for storage and analysis of big education data. An educational intelligence application for quality assessment of higher education systems has been designed, implemented and deployed in this paper.

Existing quality frameworks proposed for higher education cover many dimensions of quality including content and delivery. However, most of these frameworks consider education as a service and evaluate quality from a service-oriented perspective. Other systems consider students a product and evaluate quality from the product perspective. However,

student is also the 'outcome' of the educational process and this dimension is unexplored.

The proposed framework uses this unexplored dimension of quality. In view of the fact that the students' performances after graduating are the outcomes produced by higher education institutions, an outcome-based analysis of this transition data can give useful insights into the quality of education provided by the university. Moreover, analytics of this nature can be helpful in decision-making and administrative planning for the educational institutions, contributing to quality improvement.

The framework is implemented in the form of a web-based application, using big data technology, Google BigQuery, and R programming. It is tested using a synthetic data set to demonstrate how quality scores are calculated. With the help of the generated quality scores, analytical visualizations like piecharts and line charts are generated.

VIII. Future Work

It is important to mention that the data available for analysis is limited to students who take up a postgraduate course/research degree or campus placement after the completion of their respective courses.

Future research work shall include designing a comprehensive data collection framework and exploration of other variables that may affect and govern outcomes analysis other than transitions from one degree to the next level. Academic categories like self-employed students can also be explored to extend the framework and make it more comprehensive. Therefore, computation parameters for quality score can be improved in future work.

Besides this, the use of other cloud-based technologies for building a big data architecture [37] for comprehensive quality assessment of higher education systems is also planned for future to improve the performance and usability of the educational intelligence application.

IX. Availability of Data and Materials

Quality Management Tool for Higher Education Systems (QMHES) is available as an open source project in GitHub (https://github.com/samiyakhan13/qmhes-1). Synthetic dataset was used to test the application. An electronic version of data will be made available and shared with interested researchers under an agreement for data access (contact: samiyashaukat@yahoo.com).

References

[1] S. Khan, K. A. Shakil, and M. Alam. "Educational intelligence: applying cloud-based big data analytics to the Indian education sector" in 2016 2nd international conference on contemporary computing and informatics (IC3I), IEEE, 2016, pp. 29-34.

[2] E. Salwana, S. Hamid, and N. Mohd Yasin. "Student Academic Streaming Using Clustering Technique." Malaysian Journal of Computer Science, Vol. 30, No. 4, 2017, pp. 286-299.

[3] B. Williamson. "The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'". International Journal of Educational Technology in Higher Education. Vol. 15, No. 1, 2018, pp. 12.

[4] K. Sato. "An inside look at Google BigQuery". White paper, 2012. Available from: https://cloud. Google. com/files/BigQueryTechnicalWP. pdf.

[5] S. B. Siewert. "Big data in the cloud. Data velocity, volume, variety, veracity", 2013, pp. 4-21.

[6] A. Valdez, G. Cortes, S. Castaneda, L. Vazquez, A. Zarate, Y. Salas and G. H. Atondo, "Big Data Strategy" International Journal of Advanced Computer Science and Applications (IJACSA), 10(4), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100434

[7] H. Ehtesham, R. Safdari, and S. Tahmasebian. "Big Data in Health: New Challenges and New Solutions in Data Management (A Lifecycle Review)". Indian Journal of Science and Technology. Vol. 10, No. 13, 2017.

[8] S. Khan, K. A. Shakil, and M. Alam. "Cloud-based big data analytics—a survey of current research and future directions". In Big Data Analytics. Springer, Singapore, 2018, pp. 595-604.

[9] G. D. Puri, and D. Haritha. "Survey big data analytics, applications and privacy concerns". Indian Journal of Science and Technology. Vol. 9, 2016, pp. 1-8.

[10] B. Daniel. "Big Data and analytics in higher education: Opportunities and challenges". British journal of educational technology. Vol. 46, No. 5, 2015, pp. 904-20.

[11] J. A. González-Martínez, M. L. Bote-Lorenzo, E. Gómez-Sánchez, and R. Cano-Parra. "Cloud computing and education: A state-of-the-art survey". Computers & Education. Vol. 1, No. 80, 2015, pp. 132-51.

[12] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. A. Netto, and R. Buyya. "Big Data computing and clouds: Trends and future directions". Journal of Parallel and Distributed Computing, Vol. 1, No. 79, 2015, pp. 3-15.

[13] L. C. Liñán, and A. A. Pérez. "Educational Data Mining and Learning Analytics: differences, similarities, and time evolution". International Journal of Educational Technology in Higher Education, Vol. 12, No. 3, 2015, pp. 98-112.

[14] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider. "Analyzing undergraduate students' performance using educational data mining". Computers & Education. Vol. 1, No. 113, 2017, pp. 177-94.

[15] S. Khan, X. Liu, K. A. Shakil, and M. Alam. "A survey on scholarly data: From big data perspective". Information Processing & Management. Vol. 53, No. 4, 2017, pp. 923-44.

[16] M. K. Saggi, S. Jain. "A survey towards an integration of big data analytics to big insights for value-creation". Information Processing & Management. Vol. 54, No. 5, 2018, pp. 758-90.

[17] J. Newton. "Views from below: academics coping with quality". Quality in higher education. Vol. 8, No. 1, 2002, pp. 39-61.

[18] D. Green. "What Is Quality in Higher Education?" Taylor & Francis, 1900 Frost Road, Bristol, PA 19007-1598; 1994.

[19] A. Prisacariu, and M. Shah. "Defining the quality of higher education around ethics and moral values". Quality in Higher education. Vol. 22, No. 2, 2016, pp. 152-66.

[20] A. Y. Hou. "Quality Assurance in Asian Higher Education: Challenges and Prospects" in The Palgrave Handbook of Asia Pacific Higher Education. Palgrave Macmillan, New York. 2016, pp. 381-392.

[21] M. S. Owlia, and E. M. Aspinwall. "A framework for the dimensions of quality in higher education". Quality Assurance in Education. Vol. 4, No. 2, 1996, pp. 12-20.

[22] N. Elassy. "The concepts of quality, quality assurance and quality enhancement". Quality Assurance in Education. Vol. 23, No. 3, 2015, pp. 250-61.

[23] A. Parasuraman, V. A. Zeithaml, and L. L. Berry. "A conceptual model of service quality and its implications for future research". Journal of marketing. Vol. 49, No. 4, 1985, pp. 41-50.

[24] C. Gronroos. "Service quality: The six criteria of good perceived service". Review of business. Vol. 9, No. 3, 1988, pp. 10.

[25] U. Lehtinen, and J. R. Lehtinen. "Two approaches to service quality dimensions". Service Industries Journal. Vol. 11, No. 3, 1991, pp. 287-303.

[26] R. Carney. “Building an Image” in Proceedings Symposium for the Marketing of Higher Education, New Orleans, Lousiana: American Marketing Association. 1994.

[27] A. Athiyaman. “Linking student satisfaction and service quality perceptions: the case of university education”. European journal of marketing, Vol. 31, No. 7, 1997, pp. 528-40.

[28] H. Lee, Y. Lee, and D. Yoo. “The determinants of perceived service quality and its relationship with satisfaction. Journal of services marketing”. Vol. 14, No. 3, 2000, pp. 217-31.

[29] S. Hadikoemoro. “A comparison of public and private university students' expectations and perceptions of service quality in Jakarta, Indonesia”.

[30] M. A. O’Neill, and A. Palmer. “Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education”. Quality assurance in education. Vol. 12, No. 1, 2004, pp. 39-52.

[31] S. Sahney, D. K. Banwet, and S. Karunes. “Customer requirement constructs: the premise for TQM in education: a comparative study of select engineering and management institutions in the Indian context”. International Journal of Productivity and Performance Management. Vol. 53, No. 6, 2004, pp. 499-520.

[32] R. Brooks. “Measuring university quality”. The Review of Higher Education, 2005. Vol. 29, No. 1, 2005, pp. 1-21.

[33] V. Teeroovengadum, T. J. Kamalanabhan, and A. K. Seebaluck. “Measuring service quality in higher education: Development of a hierarchical model (HESQUAL)”. Quality Assurance in Education. Vol. 24, No. 2, 2016, pp. 244-58.

[34] A. Parasuraman A, L. L. Berry, and V. A. Zeithaml. “Understanding customer expectations of service”. Sloan management review. Vol. 32, No. 3, 1991, pp. 39-48.

[35] R. Telford, and R. Masson. “The congruence of quality values in higher education”. Quality assurance in education. Vol. 13, No. 2, 2005, pp. 107-19.

[36] S. Khan, K. A. Shakil, and M. Alam. “Big Data Computing Using Cloud-Based Technologies: Challenges and Future Perspectives” in Networks of the Future Chapman and Hall/CRC, 2017, pp. 393-414.

[37] F. Azzedin and M. Ghaleb, “Towards an Architecture for Handling Big Data in Oil and Gas Industries: Service-Oriented Approach” International Journal of Advanced Computer Science and Applications (IJACSA), 10(2), 2019. http://dx.doi.org/10.14569/IJACSA. 2019.010026.