

# Data Mining as a Tool for Quality Assurance in HLIs : Experience from Dissertation Evaluation Process at UDSM

Diana Philemon Mwiru<sup>1</sup>

<sup>1</sup> Department of marketing, University of Dar Es Salaam Business School, Dar es Salaam, Tanzania

Correspondence: Diana Philemon Mwiru, Department of marketing, University of Dar Es Salaam Business School, Dar es Salaam, Tanzania.

Received: November 27, 2017

Accepted: December 29, 2017

Online Published: January 2, 2018

doi:10.20849/abr.v2i3.275

URL: <https://doi.org/10.20849/abr.v2i3.275>

## Abstract

This paper seeks to better understand how data mining can be used to improve quality assurance system in HLIs in Tanzania. A time series analysis of 133 pairs of examiners' reports from the University of Dar es Salaam Business School (UDBS) was conducted. The analyses covered reports of three years from year 2014 to 2016. Analysis of means and correlation analysis were performed. In addition, content analysis of overall recommendations and comments provided by examiners was done. The results suggest that assessment reports contain useful information that could serve purpose other than awarding grades to students. The information contained could be used to inform future decisions geared to improving supervision, assessment and teaching in HLIs.

**Keywords:** data mining, quality assurance

## 1. Introduction

Quality assurance is at the heart of academic activities in Higher Learning Institutions (HLIs) in today's world. Although it is said to be a relatively new concept that has not matured enough (Yang, 2006), higher learning institutions worldwide strive to put in place quality assurance mechanisms. Increased labour market demand for qualified graduate who can contribute to improved performance of their organizations is among factors that necessitate higher learning institutions to institute quality assurance mechanisms. Increasingly, customers also demand for quality products (Gatfield et al. 1999, Fowler and Gilfillan, 2003), this makes quality assurance a necessity for organizations to remain competitive in the marketplace. Quality assurance in higher learning institutions simplify students mobility as it facilitates credit transfer and makes it possible for academic institutions to design and offer double degree or joint programmes.

There are mainly three pillars in the education process namely content of the program, delivery and assessment. To ensure total quality in education, quality assurance mechanisms need to be instituted in all the pillars. Quality assurance systems need to ensure that the content of the program will equip students with knowledge and skills that are needed to address problems in modern business organizations and hence assuring their relevance and employability in the business world. They also need to ensure that teaching are delivered in the most effective way so that the process impart the intended knowledge and skills. Finally, the system should ensure appropriate assessment procedures are in place which will portray students' performance level. Of the three pillars in the education process assessment produce massive information which could be used to inform various decisions in higher learning institutions. Assessment data if analyzed (data mining) can produce valuable information that can reveal relationships, trends, knowledge and quality gaps and enhance the ability of the HLIs to plan, assure and control quality. Assessment data can be used to inform programme review in terms of content and delivery, design of new programmes and may also help institutions to gauge the quality of the faculty and their relevance.

Among mechanisms instituted by HLIs in Tanzania to assure quality of research work at postgraduate level is the appointment of internal and external examiners to assess dissertations, projects and theses. Internal and external examiners produce reports which are mainly used to determine the ultimate dissertation grade of the student. The supervision for the masters' dissertation is done by the faculty who is normally appointed by the department. After completion of the research work the dissertation is marked by appointed independent internal and external examiners. As a matter of practice while for the dissertation to be accepted by the University both internal and external examiners have to pass it, the final grade awarded to the dissertation is the one by the external examiner.

While this practice has worked well and massive examination reports are being generated, these reports are hardly used for the purpose other than grading dissertations. Data mining seems to be detached from quality assurance process in the assessment of dissertations/theses. This particular study therefore, sought out to review and analyse the data contained in the reports and assess its relevance in decision making, particularly in improving quality of research conducted by students and assessment of dissertation and theses.

### *1.1 General Objective*

The overall objective of the study was to use the available data to assess information contained in the dissertations examinations reports and its relevance in decision making on quality assurance matters in HLIs.

### *1.2 Specific Objectives*

- To assess if there is correlation in grading dissertations between internal and external examiners
- To assess the extent to which overall recommendations given by the examiners match with the comments provided.
- To assess the degree of originality of dissertations produced by students

## **2. Literature Review**

Quality assurance entails all planned and systematic actions which are instituted to provide adequate confidence that the product or service will meet given requirements for quality. ISO 9000: 2015 defines **quality assurance** as that “part of *quality management* focused on providing confidence that *quality requirements* will be fulfilled.” Quality assurance provides confidence to the top management as well as customers, government agencies, regulators, certifiers and third parties. Juran and Gryna (1988) define quality as fitness for use, meaning that a quality product is the one that has features which meet customers’ needs hence providing the needed satisfaction to a customer. It entails a defect free product. Quality assurance is important to both service and manufacturing organizations as it contributes to ensure quality of outputs. Quality assurance can help an organization to occupy attractive position in the market. Quality performance resulting from proper quality assurance systems could help the organization attract more customers and remain competitive in the market. Quality assurance is not a onetime event, but rather a process that continuously seek to improve quality in the organization. It is indeed an organizational wide learning process.

### *2.1 Quality Assurance in HLIs*

Quality assurance in higher learning institutions entails all planned and systematic actions required to provide adequate confidence in education quality as a result of receiving the content that enhances employability, content being delivered in student centered manner that enhances students’ skills, knowledge and competences and assessed fairly (Alzoabi et al. 2008). In institutions of higher learning quality assurance is a monitoring process with four steps (see fig.1) that inform and provide feedback to the three pillars of the education process.

Assessing educational needs is the first step of the quality assurance process in HLIs. The educational needs are provided by potential customers (students) and employers. Students choose education programmes on the basis of the career path they wish to take. While employers suggest the type of skills and knowledge relevant to the nature of their business activities. This step informs the design of the curricula and programmes with specified outcomes and standards which is the first pillar in academic process that relates to the content of the programme. The designed curriculum is then delivered to students. Programme implementation is monitored to ensure quality. Students are given chance to evaluate the programme during their studies. At the end of their studies graduates together with other stakeholders including employers are given chance to provide feedback through tracer studies. The feedback is expected to be used to modify programmes, teaching, learning and assessment approaches.

### *2.2 Study Contribution*

This particular study aimed at contributing knowledge to quality assurance system, specifically, by looking at how data mining could be used to improve the quality assurance process in HLIs. The study analysed the data from internal and external examiners reports to uncover trends, relationships, knowledge and any quality shortcomings which can be used in future decisions to improve the assessment process and quality of dissertations.



Figure 1. Quality assurance process

Despite the varied data mining definitions by different authors, they all agree that data mining is the process of sorting and analyzing large data to discover useful patterns and information in large data repositories. This information is found as patterns on the data in different structures (Selimaj, 2011). Data mining can be defined as the process of extracting data, analyzing it from many dimensions or perspectives, then producing a summary of the information in a useful form that identifies relationships within the data. There are two types of data mining: descriptive, which gives information about existing data; and predictive, which makes forecasts based on the data. Hand (1998) defines data mining as a process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners. This definition echoes maintenance of individual databases which is highly opposed in today's world which advocates for making data accessible to many. Weis and Indurkha (1998) define data mining as a search for valuable information in large volumes of data. This definition does not point out the usefulness of the information to individuals and organizations. Haskett (2000) looks at data mining as a set of techniques that are used in an automated approach to explore and uncover complex relationships in large datasets. Data mining is seen as a process of finding interesting pattern in the data that are not explicitly part of the data (Witten et al, 2005). Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers and develop more effective marketing strategies as well as increase sales and decrease costs. Data mining depends on effective data collection and warehousing as well as computer processing. In this study understanding the importance of data sharing we modify Hand (1998) definition and define data mining as a process of analysing secondary large databases aimed at establishing unsuspected valuable relationships which can inform individual and organizational decisions.

Data mining is a process that involves a number of steps. First is the identification of data sources. This could range from electronic to manual data bases. The second step is data gathering which involves sampling of the data and transforming it. The third step entails modelling that is, creating a model to be used to test the data and evaluate it and the final stage entails deploying the model, where by the information generated from the models is used to inform decisions. The user is expected to take actions on the basis of the results produced by the model. Using data mining, HLIs in Tanzania are expected to take actions based on the information generated from the data mining models. A number of techniques can be used in data mining which include decision tree, rule induction, Genetic Algorithms and Artificial Neural network (Thearling, 2002).

### 2.3 Data Mining and Quality Assurance

The use of ICT in data collection and storage presents unique opportunity to organizations, such as, Universities which collect significant amount of data to get insights from the data generated. In the light of increased demand for quality academic programs and increased informed decisions, data mining becomes an important aspect in quality assurance. Data mining can improve quality of research work and overall learning process in HLIs by generating knowledge from massive data contained in the assessment reports. Data mining could also be used in resource allocation including staff. It could inform not only the appointment of supervisors and examiners of the

dissertations but also the grading system and fairly awarding grades to dissertations. So the central research question that the study sought to answer is *can data mining be used as a tool to improve quality of dissertations and theses assessment in higher learning institutions in Tanzania?*

This study uses decision tree( the classification and regression Trees, CART) and rule induction technique (Thearling,2002) to uncover useful information contained in the assessment reports and argues on its relevance in improving quality of students research work in HLIs.

### 3. Methodology

A time series research design was used to assess if data mining could be used to improve quality of supervision and assessment of dissertations and theses in HLIs in Tanzania. The study used Internal Examiners (IEs) and External Examiners (EEs) reports from the University of Dar es Salaam Business school. The University of Dar es Salaam is the oldest state University in Tanzania. To assure quality of its products the University of Dar es Salaam has instituted several mechanisms some of them implemented through the Quality Assurance Bureau and others are implemented at unit level. Dissertation and thesis assessment is done at Unit level guided by the University wide established procedures. The study used secondary data covering examination reports produced from 2014 to 2016 retrieved from the school data warehouse. The data was analysed with the help of Statistical Package for Social Sciences (SPSS). A total number of 133 examination report sets (266 individual reports) were analysed. Analysis of means and correlations were performed to establish if there are associations in assessment between internal and external examiners and the degree of originality of the dissertations. Content analysis was also performed to see if the overall examiners recommendations matched with the specific comments given in subsections of the dissertations.

### 4. Findings

#### 4.1 Internal and External Examiners Dissertation Assessment

The first study objective was to assess if there is correlation between internal and external examiners assessment of the dissertations. To achieve this objective, Pearson correlation analysis was performed and the results are presented in table 1. The findings indicate very weak correlation between IEs and EEs grades awarded to dissertations. Furthermore, analysis of means shows on average high scores being awarded by EEs with high standard deviations compared to grades awarded by IEs. The results can partly be explained by the fact that internal examiners for theses/dissertations are appointed on the basis of their research areas and specialty in the courses they teach at the University. It is therefore, very possible for the internal examiner to be more current and knowledgeable about the subject matter than the external examiner. For example, faculty in the department of Finance, further specialize in different areas such as Financial institutions, corporate finance etc. Hence giving them chance to be more precise when assessing dissertations.

Table 1. Correlations between internal and external examiners grades

		IE GRADE	EE GRADE
IE GRADE	Pearson Correlation	1	.345**
	Sig. (2-tailed)		.000
	N	133	133
EE GRADE	Pearson Correlation	.345**	1
	Sig. (2-tailed)	.000	
	N	133	133

\*\* . Correlation is significant at the 0.01 level (2-tailed).

#### 4.1.1 Scores on Dissertations Main Subsections

The study also assessed the pattern in scores for the main sections of the dissertations namely; Introduction (INTR), Literature review (LIT), Research methodology (METH), Findings (FIND) and conclusions. The findings are presented in table2. Generally, the findings indicate higher mean scores in the main subsections for EEs than those of IEs. From the findings, the lowest scores are observed in the Introduction chapter followed by the data analysis

and findings sections. This implies that, students face challenges in developing context of the study which include identification of the research problem. These findings somehow related to the observed average mean score in the originality of the research work. Since students do not come up with new topics, they are challenged to show the research gap. The findings also show low scores in the data analysis and findings sections. This implies that student fail to use appropriate tools for analysing data or do not have strong data analysis skills. The highest scores are observed in the research methodology section. Arguably, one would expect scores in the findings section to relate to those in research methodology. However, the current situation may imply failure by students to apply statistical tools of analysis learnt in the research methodology and quantitative methods courses. Content analysis revealed failure by students to choose the appropriate analytical tools to analyse data. For example, it has been observed in the reports that, about 40 percent of the sampled students' reports used frequencies and percentages to capture the asserted causal relationship between variables. Causal relationship cannot be captured using frequencies (Kothari, 2006)

Table 2. Examiners scores in main sections of the dissertations

GENDER		EE INTR/05	EE LIT/10METH/20	EE FIND/35	IE INTR/05	IE LIT/10	IE METH/20	IE FIND/35	
F	Mean	3.096	6.663	13.356	21.346	2.991	6.123	12.179	19.817
	N	52	52	52	52	53	53	53	52
	Std. Deviation	.5861	1.0324	2.3249	2.8961	.5234	1.0091	2.5853	3.8756
M	Mean	3.125	6.656	13.706	22.531	2.885	6.052	12.096	20.185
	N	80	80	80	80	80	80	80	80
	Std. Deviation	.6137	1.1654	2.4964	4.5412	.6673	1.0214	2.3012	3.8505
Total	Mean	3.114	6.659	13.568	22.064	2.927	6.080	12.129	20.040
	N	132	132	132	132	133	133	133	132
	Std. Deviation	.6009	1.1109	2.4274	4.0050	.6141	1.0133	2.4091	3.8498

#### 4.1.2 Dissertations Assessment Trend over the Three Years

The study also tried to find out if there is any trend in terms of assessment of the dissertations over the years. The overall findings in table 3 and 4 do not reflect any trend in the assessment. Likewise, scores in the main subsections do not portray any specific trend. This is probably attributed to by the University of Dar es Salaam practice to change examiners over time which eliminate the halo effects in the assessment process.

Table 3. Trend in scores (in percent) over the past three years

Year	Introduction		Literature review		Methodology		Findings		Recommendations and conclusions	
	IE	EE	IE	EE	IE	EE	IE	EE	IE	EE
2014	58	60	61	67	59.8	67.5	56	63.7	55	62
2015	59.4	62	58.9	70	59.4	71	59.4	65	56	61.5
2016	56	62	61	64	62	66.5	57	60	53	60

Table 4. Trend of the Mean scores over the past three years

EXAMINATION YEAR		IE GRADE	EE GRADE
2014	Mean	58.321	64.522
	N	67	67
	Std. Deviation	8.4179	10.2663
2015	Mean	59.763	65.684
	N	19	19
	Std. Deviation	8.1503	7.2748
2016	Mean	59.117	62.872
	N	47	47
	Std. Deviation	7.2944	7.3407
Total	Mean	58.808	64.105
	N	133	133
	Std. Deviation	7.9572	8.9265

#### 4.2 Correlation between Comments Provided and Overall Recommendations

Content analysis was done to evaluate the comments and overall recommendations given to dissertations. There are six recommendations for the examiner to make. First, the dissertation can pass as it is without any revisions, second, the dissertation could pass with minor changes such as typos or formatting of tables. Third, the dissertation may pass with substantial corrections. Fourth the dissertation is not accepted but could be resubmitted after making major changes. The fifth recommendation is rejection however, the dissertation could be resubmitted for a lower award. Last is the outright rejection. The content analysis revealed that in some cases recommendations given by examiners did not match with the comments given in other subsections of the dissertations. The content analysis reveal that comments on poorly stated research problem, use of wrong analytical tools which would require the candidate to re do the analysis and that could probably affect the conclusions and recommendations are considered to be minor corrections. According to assessment guidelines minor changes refer to editorial corrections, slight re-organization of sections and minor modifications of tables, paragraphs or sentences. While additional data analysis (which is the appropriate recommendation for student using wrong analytical tool) calls for non-acceptance of the dissertation.

#### 4.3 Degree of Originality of Dissertations

The study sought to establish degree of originality of research work conducted by students in higher learning institutions. The 15 points scale is used to measure the degree of originality in which 15 points mean the research is original and generates new knowledge while 1 means the topic and context are both not new. To establish degree of originality analysis of means was conducted and the results are presented in table5.

Table 5. Originality of students' research work

GENDER		EE ORIG/15	IE ORIG/15
F	Mean	8.971	8.942
	N	52	52
	Std. Deviation	1.2657	1.7140
M	Mean	9.281	8.606
	N	80	80
	Std. Deviation	1.9272	1.7113
Total	Mean	9.159	8.739
	N	132	132
	Std. Deviation	1.6990	1.7137

The results indicate that on average in the scale of 15 points, the mean score for students' dissertations in higher learning institutions 8.7 (for female students) and 9.1 (for male students) in terms of originality. However, while the findings suggest that students do not produce new discoveries in research, their work contributes to the body of knowledge and offer useful insights to policy makers and practitioners. The results indicate a slightly higher originality mean score for male students than female students. They also show high mean score by external examiners compared to the internal examiners. Overall, these findings imply that students do not venture into new research areas but rather replicate studies conducted elsewhere in the world or in different industry within the country.

## 5. Conclusions and Recommendations

The findings of this study contribute to quality assurance process in higher learning institutions in Tanzania. They offer useful insights on how massive data can be analysed to uncover hidden trends, relationships and phenomenon which can inform decisions on how to improve quality in higher learning institutions. It tries to stimulate research in data mining, this is important given that higher learning institutions generate a lot of data which is underutilized. Overall, the study concludes that the massive data generated in the dissertation/thesis assessment can be used to inform decisions and improve quality assurance process in HLIs. These findings are in line with a study by Athanasiadis et.al (2010) who observed data mining to be useful in quality assurance in an environmental monitoring networks. Specifically, on the basis of the findings the study concludes that there is a very weak correlation between external examiners and internal examiners' assessment grades. Overall, external examiners grades are higher than internal examiners grades. The second conclusion is that there is no correlation between comments given and the overall recommendations and the third conclusion is that students do not produce original research findings.

On the basis of the findings the study recommends to HLIs to make use of massive data generated by the dissertation/thesis assessment to improve the assessment and ultimate quality assurance system. Analysis of assessment data can provide insights that can be used to improve course delivery, specifically, research methodology course. Given the variation in IE and EE grades, HLIs should consider using both internal examiners and external examiners grades in determining students' final grade in the research work. Internal examiners are independent appointed faculty. They are appointed on the basis of their area of specialization in teaching and research. Therefore, they are expected to be more current and informed on the topics they are examining. On the other hand, external examiners could be appointed on the basis of functional specialization but may not necessarily be current on the specific research area. Using both reports to determine grades could iron out elements of subjectivity in dissertation examination and contribute to fair assessment of the dissertations.

Based on the observed weakness in data analysis, there is a need for the research methodology course to be more practical especially on analytical tools. Students seem to fail to use the analytical tools to analyse data. Choice of wrong and simple analytical tools to capture complex relationships such as causal relationships. Inferential analysis is done using frequencies and percentages. Students should be given proper guidance on how to choose appropriate analytical tools depending on the title and nature of the study. Furthermore, students should be encouraged to venture into new research areas to generate new knowledge. Examiners should be objective in giving recommendations. Minor corrections should only be given to minor changes as stipulated in the assessment tool. This is important if HLIs need to ensure quality of the dissertations/theses. This study, utilized data from the university of Dar es salaam Business School, future analysis could be conducted using data from more HLIs to uncover other useful trends which can inform design, implementation and management of postgraduate programmes in Tanzania.

## References

- Alzoabi, Z, Diko, F, & Alnoukari, M. (2008). *Enhancing Education Quality Assurance using Information Systems QAAS system International Symposium on Information Technology*. ITSIM 08, Malaysia.
- Athanasiadis, I, Emilio, R., & Daniel, B (2010). *Data Mining methods for quality assurance in an Environmental Monitoring network, Artificial Neural Network*. ICANN.
- Davier, A. (2012). *The Use of Quality control and data mining techniques for monitoring scales scores*. Research Report, ETS RR-12-20.
- Fowler, A., & Gilfillan, M. (2003, Decemebr). A framework for Stakeholder Integration in Higher Education information systems projects. *Technology, Analysis and Strategic Management*, 15(4).

- Gatfield, T, Barker, M., & Graham, P. (1999). measuring Student Quality Variables and the Implication for the management Practices in higher Education Institutions: An Australian International Students perspective. *Journal of higher Education Policy and Management*, 21(2).
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts And Techniques*(Morgan-Kaufman series Of Data Management System). San Diego: Academic Press.
- Hand, D. J. (1998). Data Mining: Statistics And More? *The American Statistician*, 2, 112-118.
- Haskett Mitch. (2000). An Introduction To Data Mining.Part 1.Understanding The Critical Data Relationship In The Corporate Data. *Warehouse.Enterprise System Journal*, 15, 32-34.
- ISO 9000:2015: Quality management systems—Fundamentals and vocabulary.
- Juran, J. M., & Gryna F.M. (1988). *Juran's quality control handbook*, Mcgraw-Hill.
- Selimaj, L. (2011). *Data Mining and Automatic Quality Assurance*. Bachelor Project, Copenhagen University college of Engineering.
- Thearling, K. (2002) Data Mining Techniques. Retrieved From <http://www.3shore.net/zkut/text/dmwhite/dmwhite.htm>
- Weiss. S. H., & Indurkha. N. (1998). *Predictive Data Mining*. A practical Guide:Morgan Kaufman Publishers San Francisco CA.
- Witten, I. H. et al. (2005). Data Mining. *Practical Machine Learning Tools And Techniques* (2<sup>nd</sup> ed,Morgan-Kaufman Series Of Data Management Systems) San Francisco:Elsevier.
- Yang, C. (2006). problems in Quality Assurance Under open Source Development Mode. Retrieved from [www.inf.fu-berlin.de/inst/ag-se/teaching/S-OpenSource2006/YangC06\\_Quality\\_Assurance\\_Ausarbeitung.pdf](http://www.inf.fu-berlin.de/inst/ag-se/teaching/S-OpenSource2006/YangC06_Quality_Assurance_Ausarbeitung.pdf)

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).