# ENCOURAGING COURSE DESIGNER ENGAGEMENT WITH DATA ANALYSIS METHODS IN VIRTUAL LEARNING ENVIRONMENTS

**Stephen Smith[1], David Cobham[2], Kevin Jacques[3]**

*[1]Lincoln College (UK)*
*[2]University of Lincoln (UK)*
*[3]University of Lincoln (UK)*

## Abstract

Virtual Learning Environments (VLEs) provide a fundamental contribution to modern pedagogy in education. In addition to supporting student learning and programme management, they contain student usage data that has the potential to inform and improve this pedagogy. In earlier research, the authors explored how the development of data mining and log analysis systems for the Moodle virtual learning environment might improve course engagement by students [1]. They proposed that a student will complete missed tasks sooner if their utilisation of the VLE is automatically tracked and electronic prompts are sent when VLE activities are missed. A software tool named MooTwit was designed and used to test this proposal. In this paper the authors extend their research to explore how the development of data mining and log analysis systems for the Moodle virtual learning environment might encourage course designers' future engagement with data analysis methods for the evaluation of course resources.

The paper hypothesises that presenting course designers with a simple-to-use data mining and visualisation tool increases their future acceptance of data mining technology for informing course design with a longer-term intent; this should improve the quality of the online learning experience, ultimately improving student engagement. Exploring the hypothesis required the development of MooLog – a tool that extracts and presents summative information on VLE course utilisation. To ascertain if the acceptance of data mining for course evaluation could be improved, surveys were used before and after a demonstration of MooLog to a group of course designers. The pre-demonstration survey assessed existing planning and evaluation processes. The post-demonstration survey collected evaluations of the relevance of the information provided by MooLog and the likelihood of the software being used to evaluate course effectiveness. The results of the study established that many designers currently do not use data analysis as a method of informing course improvement and there was evidence to suggest the MooLog demonstration significantly increased acceptance of the potential of data mining. The findings show that educational data mining has the potential to improve the quality of VLE mediated education; it identifies opportunities to raise acceptance by course designers of data mining to improve the validity and quality of course evaluation.

Keywords: Virtual Learning Environments, Course Design, Higher Education, Social Networks.

## 1 INTRODUCTION

The use of virtual learning environments (VLEs) to augment the learning opportunities through online courses, supplementary activities and resources for learners has been identified as being utilised in a high proportion of surveyed educational institutions within the UK [2], [3]. The adoption of such technologies has led to a wider range of tools to engage learners with topics relating to their learning. This increased reliance on resources that are to be used by students independently has created new challenges for both the learner and the developer. For a learner to be successful in their studies they must engage with the VLE systems and designers must provide learning artifacts that make efficient use of their time learning from them. This study's aim is to improve the quality of a student's independent learning experience by encouraging the future engagement of course developers with data mining systems to identify online course resources and activities that are meeting learner needs.

This intent led to the research question "How might the development and demonstration a of data mining system for the Moodle virtual learning environment, encourage course designers' future engagement with data analysis methods to evaluate course resource effectiveness?".

There have been many studies [4]–[14] into the use of data from virtual learning environments that identify the level of utilisation of the activities and the resources within them. The common theme from these investigations is the success in generating the information from the data. The conclusions drawn from the studies consistently avoid establishing whether there was any positive impact as a consequence of providing the information to either the course designer/deliverer or the learners themselves. Conclusions typically make statements such as "expert teacher knowledge for learning analytics can sometimes be outperformed by knowledge derived by data-mining algorithms" [5] and "Educators with no expertise in data mining can also apply their hands in these fields." [6]. Although Kaur does evaluate the success of the accuracy of the information in identifying slow learners, there is no evidence of impact.

The investigation reported on in this paper involved course developers within Lincoln College who use the Moodle VLE system for purpose of academic study and delivery of Further Education and Higher Education qualifications, with the intention of improving student engagement with VLE resources by promoting the development of high-quality online learning environments and study habits within the College.

This research focuses on the impact of developing and demonstrating to course designers a simple-to-use data mining and visualisation tool MooLog, that graphically presents information about course resources to identify whether it would encourage designers to engage with course utilisation data analytics to improve course design.

MooLog is a bespoke software artefact aimed at ease-of-use for the Moodle course developer, eliminating the need for generic tools and their configuration. MooLog automates the analysis of the access logs relating to a specific course and communicates to the designer a breakdown of the course utilisation in a format that is easy to understand.

Previous research by [5] used KEEL (Knowledge Extraction based on Evolutionary Learning) "…a software tool to assess evolutionary algorithms for Data Mining…" to extract and analyse data for checking student performance and [6] used the existing data mining tool WEKA (Waikato Environment for Knowledge Analysis) machine learning algorithms on the data. A weakness in their methodology was that the intended user had to recreate the whole processing system, configuring, and setting up WEKA and KEEL to achieve a result.

The expectation is that a better, more user-friendly tool will improve designers' engagement with data mining and its resulting information to improve the quality of their learning resources. Given that previous research [7] has suggested that VLEs "…can enhance students' HE goals achievement…", evaluation of the quality of resources is a desirable activity with the positive outcomes from this improvement in resources having a direct benefit for learners in the quality of the learning experience provided by the institutions' VLE and, provide course designers with greater insight into the utilisation of the learning activities they create, enabling them to focus on the development of course elements with high impact on learning.

## 2 METHODOLOGY

The intent of the study was to identify if data analysis methods and techniques were demonstrated as a sufficiently simplified process for the participants within the investigation; would this increase their acceptance of data mining as a technique for evaluation of the effectiveness of VLE course content based on the level and frequency of student engagement? The simplification of the process was achieved through the development of a software application MooLog that automated data analysis and presentation of the resulting information in an accessible format.

### 2.1 Participant selection

Initially an investigation was made into the users of Moodle in the college to identify the teaching staff that actively engaged with the VLE to ensure the survey was being sent to staff who were regularly modifying courses on the system. Frequent system users were identified by filtering Moodle's access log for staff who logged in and modified material on Moodle at least once over a seven-day period, this process was repeated for four weeks resulting in 58 candidates who consistently engaged with the system. A sample size of 51 was identified as being a representative sample of the 58 frequent users of Moodle for teaching within the college, giving 5% margin of error with a confidence level of 95% [15].

## 2.2 MooLog data analysis tool

Fig. 1 shows the output from the MooLog data analysis tool which processes data related to student accesses to the materials and activities contained within a Moodle based course. Upon importing a data file the course designer can optionally link a grades file to the learners who have accessed a course and select and filter by student and grades.

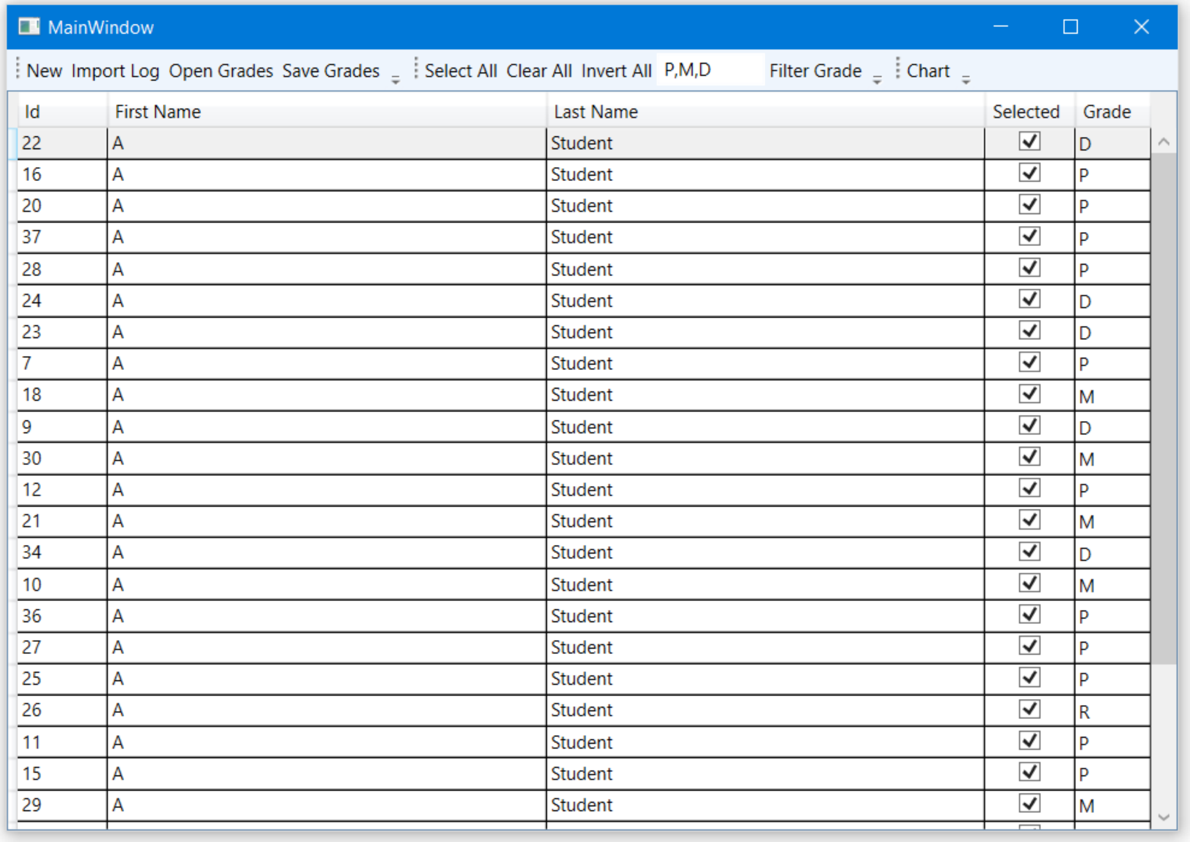


*Figure 1. MooLog - Student data linked to grades screen.*

Fig. 2 shows how selecting the chart option to create a dynamic chart of access information based on the filtered data.

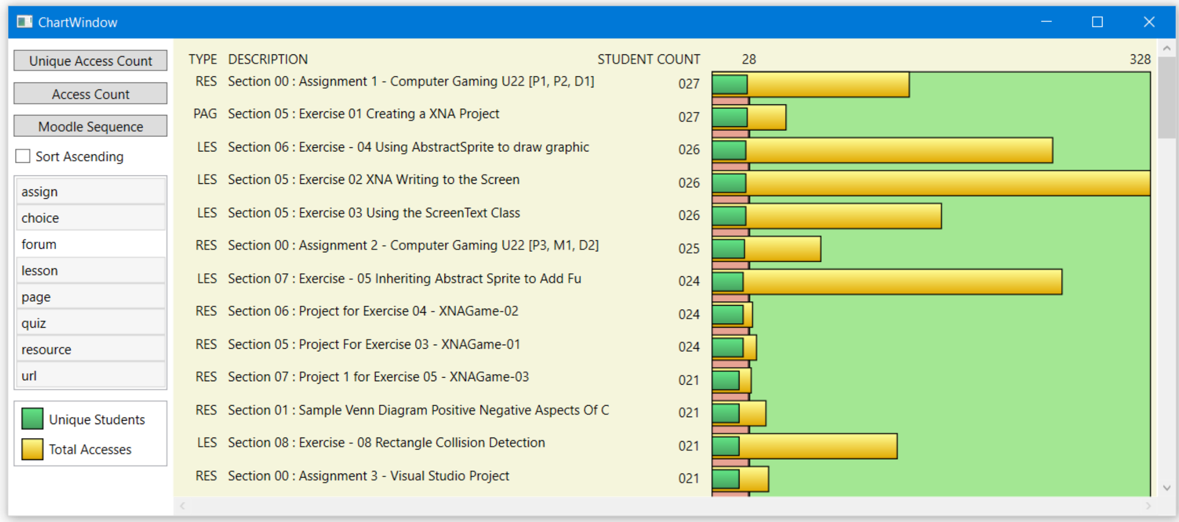


*Figure 2. MooLog - Summary data analysis with filtering.*

The resulting chart can be sorted by the number of unique accesses, total accesses or by the sequence the activities appear in the course. The designer can also dynamically filter the results to different types of activities. The software demonstration consisted of a short video of the software in operation with captions to identify the features being used. This was presented without any narrative or voiceover to try to reduce any bias towards the usefulness of the software being introduced.

## 2.3 Data collection

The participants were surveyed twice using online surveys one containing a video demonstration, as it "can be easily used when we know the basic population and have online access to its members" [16], Google forms and YouTube were used as the survey and demonstration platform as they facilitated effective distribution of information and collection of quantitative data. The survey links were distributed to each candidate via email and to ensure a sufficient response rate tutors were contacted the week after the initial email to encourage completion. The surveys were created with a focus on the two key areas of the study.

### *2.3.1 Pre-demonstration*

A pre-demonstration survey with questions to establish existing working practices of the users.

1. How they use Moodle for learning
2. How they organize and structure resources on Moodle
3. If or how they evaluate their course design
4. Their ability to gather data about student activity on Moodle
5. The percentage of their Moodle course materials and activities that are vital to complete a course

### *2.3.2 Post-demonstration*

A second survey containing a software demonstration video followed by questions to identify if the software features were of use in the context of evaluating and supporting redevelopment of courses. The questions asked the participants to rate the usefulness (not useful; slightly useful; useful; important or essential) of:

1. Knowing how many of a student group accessed each resource
2. Knowing how many total accesses there were to a Moodle resource (popularity)
3. Being able to identify resources that had low or zero usage
4. Being able to identify resources that were used based on a cohorts' achievement level
5. Being able to find an individual students' usage

Additionally, the participants were also asked "if the software was made available to you to help you decide what resources and activities worked well with your students, would you use it?" with the capability to respond with Yes; Yes, with training and support; Maybe; No, I would not or Other.

The extracted data from the pre-demonstration survey provided a foundation for understanding of the extent that Moodle is used for learning and what techniques are used to evaluate the quality of the learning activities.

Data from the post-demonstration survey, was collected to provide insights into how valuable the extracted information was to participants in the context of understanding the effectiveness of the course materials and activities; the survey also identified the participants willingness to utilize these techniques in the future.

# 3 RESULTS

## 3.1 Pre-demonstration

The pre-demonstration survey established how the participants made use of the VLE usage, identifying the extent that the features of the VLE were used and to identify any existing techniques used to evaluate the course activities and materials.

### *3.1.1 How participants use Moodle for learning?*

To ascertain what types learning activity took place using Moodle developers were asked to identify their most frequently used elements in the VLE.

*Table 1. Course designers preferred VLE usages.*

| | |
|---|---|
| 48% | as a system for student self-study |
| 92% | as a store of reference material |
| 25% | for assessment |
| 36% | for communication and collaboration |

### *3.1.2 How participants organise and structure resources on Moodle?*

Developers were asked how they organised resources on Moodle to help identify if there was any specific structure in their organisation of materials. The 88% of the developers structured their materials in relation to topics or planned delivery sequence, the remaining 12% did not relate the materials to a student's study sequence. This indicated that most developers considered how the materials were to be accessed by the users, identifying there was some level of planning and structure to VLE courses within the college.

### *3.1.3 If or how participants evaluate their course design?*

To establish what systems designers already applied to when evaluating a course, they were asked to identify their existing mechanisms. Most developers used some form of evaluation with 53% using student feedback and 10% using data; the remaining 37% did not evaluate or guessed.

The results suggest that although designers previously indicated that they structured their VLE courses, a significant proportion of the designers did not evaluate the courses they created or used informal feedback to influence improvements. Very few designers stated they used data analysis, possibly indicating that this method is either unfamiliar, perceived as difficult or they lack the skills to use the method.

### *3.1.4 Participant's ability to gather data about student activity on Moodle.*

To establish the course developer's basic skill level in extracting data from the standard Moodle user interface, they were questioned about how able they were at retrieving information about a student's activity on the system. The survey results indicated that 32% were below average or without ability at retrieving data, 28% had an average level of ability to retrieve data and 40% were above average or highly skilled in retrieving data.

The previous chart indicated that only 10% of designers used data as a method of evaluation and it was proposed that factors that might explain this would include a lack of ability to retrieve the data. The results presented above indicate that 60% of the designers rated themselves as having from no ability to an average ability to retrieve data. Potentially this may be linked to the low usage of data analysis for evaluation and supports the argument for the use of simple tools to aid tutors in retrieving and analysing data.

### *3.1.5 The percentage of Moodle course materials and activities that are vital to complete courses created by participants.*

To establish the level of importance of the online resources provided to the students by the course designers, the developers were asked what proportion of a Moodle courses content was vital to the students being successful on their course. 56% of developers considered at least 40% of their resources as vital for success, 40% considered below 40% of resources as necessary for success and 4% did not answer the question. These results indicated that online resources are in many cases vital to a student completing successfully; at least half of the course designers provided a large proportion of essential resources online for their courses. With a high proportion of resources being vital, there is a need to evaluate them to identify if they are effective and being used by students, this supports the studies aim to encourage evaluation through data analysis.

## 3.2 Post-demonstration

The participants were surveyed to identify which of the functional parts of the software would aid them in evaluating the usefulness of elements in a course based on student utilization and to identify if the use of a software tool would be accepted by the participants if it was made available to them.

### *3.2.1 Knowing how many of your student group accessed each resource.*

Results identified that 88% of the respondents responded positively and 12% negatively, to the usefulness of knowing the total number of individual students that accessed a resource. The high positive response level gives some confidence in the usefulness of the tool in general with the usage data being provided to course designers being seen as an effective metric.

### *3.2.2 Knowing how many total accesses there were to a Moodle resource (popularity).*

In response to the usefulness of identifying resource popularity 74% of the respondents responded positively and 26% negatively, to usefulness of knowing the total number of individual times a resource was accessed. The popularity of a resource was identified as another useful metric based around usage data; this has a bearing on the level of acceptance of data mining for course evaluation by the designers; showing that data when provided to designers has a very low rejection rate.

### *3.2.3 Being able to identify resources that had low or zero usage.*

The usefulness of knowing which resources had a low or zero access count was seen by 82% of the respondents as useful and 18% indicated that had little usefulness. A high level of positive response to the ability to be able to identify low usage resources was expected in this question given that to evaluate the success of a resource, a key indicator of a poor resource would be that it is not being used. It was expected that the positive responses would be greater than the responses from the previous two questions, this was not the case, raising questions about the negative respondents; there is some indication that there is a core of respondents that have rejected the application facilities in general.

### *3.2.4 Being able to identify resources that were used based on a student's level of achievement.*

When questioned on the linking of resource utilisation to achievement 84% of the respondents responded positively to usefulness of the linking of information and 16% responded negatively to usefulness of linking a student's level of achievement with the resources they accessed. This question focused on providing linked information (grades to usage) compared to the previous questions that focused on single data items. A greater number of respondents saw this information as important compared to the basic utilisation information from the first two questions. This is an encouraging sign given that this type of information would require a high level of ability and would take a significant amount of time to produce without software and gives some confidence that the survey is identifying positive feedback from the designers.

### *3.2.5 Being able to find an individual students' usage.*

When asked about usefulness of being able to identify an individual's usage of the course elements 86% of the respondents responded positively and 14% negatively, to usefulness of knowing individual students access to resources. This question focused on identifying the level of importance of being able to mine data about individual students, this question was deliberately placed within the survey to encourage the participants to consider data mining as more than just a course evaluation tool. The increase in responses in the important to essential range may indicate that many of the designers are still more concerned about tracking what students are doing than evaluating the activities they provide to them.

### 3.2.6 Would the tool be used if provided in normal activity?

The participants were questioned about the usefulness of the software when applied their working processes. The specific question asked was: If this software was made available to you to help you decide what resources and activities worked well with your students, would you use it?

Of the 50 potential respondents 50 responded with 5 providing null responses, from the valid responses 65% responded positively, 29% were uncertain and 6% would not use the software. The average positive responses of 83% from the previous questions relating to the usefulness of the data combined with the 65% overall positive response to this question shows there is a clear shift in acceptance of the data mining tool and the data it provides. There does seem to be a core element of responses in the range of 4% - 8% in each question where data mining is rejected; this is low rejection rate and does suggest that overall, there would be a good uptake of data mining in future.

### 3.2.7 Comparison of levels of acceptance between designers who previously did/did not evaluate their courses.

Prior to the demonstration of the software designers were asked if they currently evaluated their courses, this resulted in a 37%(No)/63%(Yes) split within the group. Subsequent to the demonstration all participants were asked if they would adopt the software as part of their normal activities, the chart below identifies the level of adoption of the software by the two groups. The level of acceptance in the two groups showed comparable levels of acceptance of the software as part of their normal design and evaluation activities.

## 4 CONCLUSIONS

In terms of future work, it is hoped that the results will contribute to larger areas of study including the production of additional mechanisms for understanding how learners learn effectively when working independently online and how they can be encouraged to engage successfully with online learning systems that can be structured to their needs. It is also hoped that the software developments undertaken within the study will provide software that could be used in the wider educational community and a framework for future VLE data analysis developments.

## ACKNOWLEDGEMENTS

The content of this paper is based on research carried out by Stephen Smith for the award of MSc by Research under the supervision of Dr David Cobham and Dr Kevin Jacques. The original thesis can be found at reference [17] below. All authors contributed to the writing of the paper and all authors have approved the final version.

## REFERENCES

[1] S. Smith, D. Cobham, K. Jacques, “The Use of Data Mining and Automated Social Networking Tools in Virtual Learning Environments to Improve Student Engagement in Higher Education” in *International Journal of Information and Education Technology*, vol 12, no 4, April 2022.

[2] J. Walker, R., Voce, J. and Ahmed, “Survey of technology enhanced learning for higher education in the UK,” p. 10, 2014.

[3] Ofsted and The Office for Standards in Education, “Virtual Learning Environments : an evaluation of their development in a sample of educational settings,” *Alexandra House*, no. January, pp. 1–28, 2009.

[4] D. Y. T. Liu, D. Richards, P. Dawson, J.-C. Froissard, and A. Atif, “Knowledge Acquisition for Learning Analytics: Comparing Teacher-Derived, Algorithm-Derived, and Hybrid Models in the Moodle Engagement Analytics Plugin,” in *Pacific Rim Knowledge Acquisition Workshop*, pp. 183–197, 2016

[5] C. Romero, S. Ventura, and E. García, “Data mining in course management systems: Moodle case study and tutorial,” *Comput. Educ.*, vol. 51, no. 1, pp. 368–384, Aug. 2008.

[6] P. Kaur, M. Singh, and G. S. Josan, “Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector,” *Procedia Compu. Sci.*, vol. 57, pp. 500–508, 2015.

[7] E. Lacka, T.C. Wong, M.Y. Haddoud, “Can digital technologies improve students’ efficiency? Exploring the role of Virtual Learning Environment and Social Media use in Higher Education” in *Computers & Education*, vol 163, p.104099, 2021.

[8] D. Yen, -Ting Liu, J.-C. Froissard, D. Richards, and A. Atif, “An enhanced learning analytics plugin for Moodle: student engagement and personalised intervention,” *Glob. Connect. Digit. enabled. Proc. ascilite 2015*, pp. 180–189, 2015.

[9] S. K. Mohamad and Z. Tasir, “Educational Data Mining: A Review,” *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 320–324, Nov. 2013.

[10] C. Romero, C. Castro, and S. Ventura, “A Moodle Block for Selecting, Visualizing and Mining Students’ Usage Data,” 2013.

[11] C. Larmuseau, M. Evens, J. Elen, W. Van Den Noortgate, P. Desmet, F. Depaepe, “The relationship between acceptance, actual use of a virtual learning environment and performance: An ecological approach”, in *Journal of Computers in Education*, vol 5 no 1, pp.95-111, 2018.

[12] S.U. Hassan, H. Waheed, N.R. Aljohani, M. Ali, S. Ventura, F. Herrera, “*Virtual learning environment to predict withdrawal by leveraging deep learning*,” in International Journal of Intelligent Systems, vol 34 no 8, pp. 1935-1952, 2019.

[13] C.A. Boulton, C. Kent, H.T. Williams, “*Virtual learning environment engagement and learning outcomes at a ‘bricks-and-mortar’ university*”, in Computers & Education, 126, pp.129-142, 2018.

[14] E. Ryan, C. Poole, “*Impact of virtual learning environment on students’ satisfaction, engagement, recall, and retention*,” Journal of Medical Imaging and Radiation Sciences, vol 50 no 3, pp.408-415, 2019.

[15] CheckMarket, “Sample size calculator,” 2015. [Online]. Available: https://www.checkmarket.com/sample-size-calculator/. [Accessed: 07-Oct-2015].

[16] G. Szolnoki and D. Hoffmann, “Online, face-to-face and telephone surveys - Comparing different sampling methods in wine consumer research,” *Wine Econ. Policy*, vol. 2, no. 2, pp. 57–66, 2013.

[17] S. Smith, Thesis for the award of MSc by Research, Available on request from University of Lincoln library.