DEAnER: An Effective Assessment Model

Rodany A. Merida, Melie Jim F. Sarmiento
University of the East - Manila

Abstract

The study was all about developing an effective assessment model for the exam analysis and evaluation. The study aims to give way to a software DEAnER that can provide the necessary reports for the end-users with the integration of new technology. In this study, Scantron machine is the sole provider of data in which, the authors think of a way wherein they can maximize these technologies to the fullest. The College of Computer Studies and Systems conducts departmental exams where all students taking the same subject undertake an equal exam. One of the technologies used is the Scantron machine which scans answer sheets of students, which makes the checking of the paper an easy and efficient task. On the other hand, the results like determining the percentage of passing students, the lists of those who failed and passed, standard deviation and variance and the topnotch are manually processed giving a hard time for faculty members to submit their reports on time. Due to its limitation, a Departmental Exam Analysis and Evaluation Report (DEAnER) was developed that will serve as an assessment model. This software intends to efficiently process the raw results of the departmental exam from the Scantron machine into useable information. The effectiveness of this assessment model shows in the reports being produced and the correct interpretation of raw data that resulted to precise evaluation reports in a departmental examination.

1. Introduction

With the status of Center of Development in Information Technology, the College of Computer Studies and Systems of the University of the East, Manila seeks to improve its processes and systems through the use of Information Technology. Necessary as means of ensuring the quality of education, each college of the University conducts Departmental Exams where all students taking the same subject undertake an equal exam. The University utilizes a Scantron machine which scans answer sheets of students by identifying lead or pencil marks. This practice deeply makes the checking of the answer sheets an easy and efficient task. However, the results are manually processed by faculty members to determine the needed reports to be submitted to the College like percentage of passing students and the topnotch or highest score list.

“Every faculty member wants to better understand his or her students’ strengths and weaknesses and measure how well they are grasping key concepts. But, without the right support technology, using assessment to do that can be tedious,” said Patrick Chadd, Manager of Academic Systems and Educational Technology at Rowan University’s School of Osteopathic Medicine.

Since the process of determining and analyzing the score results of the students is manually done, time and effort is needed before such reports and postings of the result can be achieved. Scantron machines used by the University outputs the result of the checking as .DAT files, or text files that follows a strict protocol and syntax in displaying the results. Faculty members examines each .DAT file, record and tally the results to generate a list of passing students, list of scores and other general reports. These general reports such as percentage of passing students and the statistics of scores are needed by the College Management for the overall performance of the students, the professors and ultimately the College.

In order to address the problems in the manual processing of Departmental Exam results using the Scantron, the researchers developed a system the will effectively used to process extracted data (.DAT files) thus, will be used to provide a learning assessment of the students leading to faculty and management assessment as well.

2. Background of the Study

Many universities nowadays are using Information Technology in administering examinations to the students. Every school’s information technology infrastructure is different. It is now possible for instructors to choose the option that best meets their needs in giving exams to their students. Among these options are the use of online examination and the exam analysis and evaluation system. In the exam analysis and evaluation system, it gives you instant access to valuable information like:

- List of students who passed and failed the exam
- Which question or topic was most difficult and which was easiest?
3. Objectives of the Study

The general objective of this study is to develop an effective assessment model to be named as Departmental Exam Analysis and Evaluation Report (DEAnER) software system for the College of Computer Studies and Systems of the University of the East Manila Campus.

The researcher has identified the following as the specific objectives of this study:

- To computerize the processing of Departmental Exam results of the University of the East College of Computer Studies and Systems.
- To design a prototype that will reduce the time needed by faculty members to analyze and process the Departmental Exam results, enabling them to focus on more important tasks.
- To design a prototype that is compatible with the hardware used by the College of Computer Studies and Systems.
- To provide a solution that will minimize if not eliminate the use of other computer software such as Microsoft Excel in processing the Departmental Exam results.
- To design a prototype that provides formatted reports of the Departmental Exam ready for printing.
- To test and evaluate the acceptability of each part of the system by gathering feedback from perspective users and as well as technical experts.

4. Scope and Limitation

The study is about developing a new computerized system that will serve as an assessment model that processes the results of the Departmental Exams. The researchers limit the scope of this study to Departmental Exams conducted by the College of Computer Studies and Systems of the University of the East, Manila Campus at Recto Avenue, Manila. These Departmental Exams are computer or related subjects, offered and taught by the college and its faculty, conducted at every end of each semester.

The study covers the analysis of the previous situation and the processing of the Departmental Exam results as relayed by faculty members from September of 2009 to May of 2010. This includes interviewing key faculty members and analyzing the flow of data in processing the Departmental Exam results. It also involves the discussion about the developed system which includes the system features, data requirements, the modules and their respective functions and as well as the software and hardware that are needed in order to run the developed system.

In order to solve the problems in the Departmental Exam results processing, to minimize the workload needed and to improve the integrity of results and reports, the researchers aims to develop the DEAnER system which facilitates the computerization of the processing of .DAT files from the Scantron machine that checks student answer sheets. It includes computerization of the following reports: Top Ten Students for every Departmental Exam subject, Percentage of Passing Students, and Percentage of Failing Students, which are all output of the developed system.

This assessment model is a LAN-Based computer software system which consists of multiple
integrated modules or sub-systems each performing a different operation and can only be accessed by specific users. It does not require the acquisition of new computers; instead it is installed to existing computer resources used by the CCSS faculty members. However the system does need Scantron machine results or .DAT files as input or raw data for processing, though the system is not linked directly to the Scantron machine to receive the .DAT files.

5. Review of Related Literature

Assessment of and for students’ learning is the process of gathering and analyzing information as evidence about what students know, can do and understand. It is part of an ongoing cycle that includes planning, documenting and evaluating students’ learning. (Adapted from The Early Years Learning Framework for Australia, p.17)

“Access to actionable outcomes data and insights is important to faculty members. But, the last thing they need is another cumbersome piece of technology that eats up time.” said Daniel Muzquiz, Chief Executive Officer of ExamSoft. “Everyone is talking about big data today, but big data doesn’t mean anything if it is not relevant, timely, and a simple click away from faculty fingertips.

According to the Centre for the Study of Higher Education, for most students, assessment requirements literally define the curriculum. This based on the study entitled Assessing Learning in Australian University wherein they tackle the Core principles of effective assessment.

The ideas and strategies in the Assessing Student Learning resources support three interrelated objectives for quality in student assessment in higher education.

<table>
<thead>
<tr>
<th>Three objectives for higher education assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. assessment that guides and encourages effective approaches to learning;</td>
</tr>
<tr>
<td>2. assessment that validly and reliably measures expected learning outcomes, in particular the higher-order learning that characterises higher education; and</td>
</tr>
<tr>
<td>3. assessment and grading that defines and protects academic standards.</td>
</tr>
</tbody>
</table>

Using the Teachers Guide to Assessment the paper defines Assessment as the process of gathering and interpreting evidence to make judgments about student learning. It is the crucial link between learning outcomes, content and teaching and learning activities. Assessment is used by learners and their teachers to decide where the learners are at in their learning, where they need to go, and how best to get there. The purpose of assessment is to improve learning, inform teaching, help students achieve the highest standards they can and provide meaningful reports on students’ achievement.

Item analysis can be a powerful model to analyze the responses of different people answering a particular set of questions. For these questions to be analyzed, the question should be a valid measure of instructional objectives. “Further, the items must be diagnostic, that is, knowledge of which incorrect options students select must be a clue to the nature of the misunderstanding, and thus prescriptive of appropriate remediation.” Furthermore, according to the Michigan State University Academic Technology Services uses the Scoring Office as their software to analyze the student’s answers and responses to a specific item.

Item Analysis Reports

“As the answer sheets are scored, records are written which contain each student’s score and his or her response to each item on the test. These records are then processed and an item analysis report file is generated. An instructor may obtain test score distributions and a list of students’ scores, in alphabetic order, in student number order, in percentile rank order, and/or in order of percentage of total points. Instructors are sent their item analysis reports as e-mail attachments. The item analysis report is contained in the file IRPT####.RPT, where the four digits indicate the instructor’s GRADER III account. A sample of an item response pattern is shown below.”

![Figure 1. Sample item analysis for Item 10 of 125](image)
Item Analysis describes the statistical analyses which allow measurement of the effectiveness of individual test items. An understanding of the factors which govern effectiveness (and a means of measuring them) can enable us to create more effective test questions and also regulate and standardize existing tests.

There are three main types of Item Analysis: Item Response Theory, Rasch Measurement and Classical Test Theory. Although Classical Test Theory and Rasch Measurement will be discussed, this document will concentrate primarily on Item Response Theory.

5.1. The Models

Classical Test Theory (traditionally the main method used in the United Kingdom) utilizes two main statistics - Facility and Discrimination.

- **Facility** is essentially a measure of the difficulty of an item, arrived at by dividing the mean mark obtained by a sample of candidates and the maximum mark available. As a whole, a test should aim to have an overall facility of around 0.5, however it is acceptable for individual items to have higher or lower facility (ranging from 0.2 to 0.8).

- **Discrimination** measures how performance on one item correlates to performance in the test as a whole. There should always be some correlation between item and test performance, however it is expected that discrimination will fall in a range between 0.2 and 1.0.

Item Response Theory (IRT) assumes that there is a correlation between the score gained by a candidate for one item/test (measurable) and their overall ability on the latent trait which underlies test performance (which we want to discover).

IRT can be used to create a unique plot for each item (the Item Characteristic Curve - ICC). The ICC is a plot of Probability that the Item will be answered correctly against Ability. The shape of the ICC reflects the influence of the three factors:

- Increasing the **difficulty** of an item causes the curve to shift right - as candidates need to be more able to have the same chance of passing.

- Increasing the **discrimination** of an item causes the gradient of the curve to increase. Candidates below a given ability are less likely to answer correctly, whilst candidates above a given ability are more likely to answer correctly.

- Increasing the **chance** raises the baseline of the curve.

Using IRT models allows Items to be characterized and ranked by their difficulty and this can be exploited when generating Item Banks of equivalent questions. It is important to remember though, that in IRT2 and IRT3, question difficulty rankings may vary over the ability range.

5.2. Rasch Measurement

Rasch measurement is very similar to IRT1 - in that it considers only one parameter (difficulty) and the ICC is calculated in the same way. When it comes to utilizing these theories to categorize items however, there is a significant difference. If you have a set of data, and analyze it with IRT1, then you arrive at an ICC that fits the data observed. If you use Rasch measurement, extreme data (e.g. questions which are consistently well or poorly answered) is discarded and the model is fitted to the remaining data. (Assessment Issues: Item Analysis)

According to David Curtis, in the analysis of data, which arise from the administration of multiple choice tests or survey instruments and which are assumed to conform to a measurement model such as Rasch, it is normal practice to check item fit statistics in order to ensure that the items used in the instrument cohere to form a unidimensional trait measure. However, checking whether individuals also fit the measurement model appears to be less common. It is shown that poor person-fit compromises item parameter estimates and so it is argued that person-fit should be checked routinely in the calibration of instruments and in scoring individuals. Unfortunately, the meanings that can be ascribed to person-fit statistics for attitude instruments is not clear. A proposal for seeking the required clarity is developed. (Item Response Theory, Rasch, person-fit statistics, attitude)

To conduct a Rasch analysis using Winsteps, according to Jams Sick, the user must first create a control file that specifies the model parameters, data structure, and output format using a special Winsteps control language. This control file is saved as a text file and then run from the Winsteps program. The data to be analyzed can either be appended to the end of the control file, or stored in an Excel file that is addressed in the control file. Because the control file syntax can be intimidating to non-programmers, a graphical control file set-up module has now been added to Winsteps. This allows users to set up their analyses in a more familiar graphical fashion, by selecting radial buttons and filling in labeled text fields.

5.3. Basic Item Analysis for Multiple-Choice Tests

According to Kehoe (1995), “the basic idea that we can capitalize on is that the statistical behavior of
‘bad’ items is fundamentally different from that of ‘good’ items. Of course, the items have to be administered to students in order to obtain the needed statistics. This fact underscores our point of view that tests can be improved by maintaining and developing a pool of ‘good’ items from which future tests will be drawn in part or in whole. This is particularly true for instructors who teach the same course more than once.”

5.4. Ranking

A ranking is a relationship between a set of items such that, for any two items, the first is either ‘ranked higher than’, ‘ranked lower than’ or ‘ranked equal to’ the second. In mathematics, this is known as a weak order or total preorder of objects. It is not necessarily a total order because two different objects can have the same ranking. The rankings themselves are totally ordered. For example, materials are totally preordered by hardness, while degrees of hardness are totally ordered.

By reducing detailed measures to a sequence of ordinal numbers, rankings make it possible to evaluate complex information according to certain criteria. Thus, for example, an Internet search engine may rank the pages it finds according to an estimation of their relevance, making it possible for the user quickly to select the pages they are likely to want to see.

Analysis of data obtained by ranking commonly requires non-parametric statistics.

5.5. DAT files (.dat)

DAT files are data files, usually comma delimited, that contain data in ASCII format. These files are able to be open in a variety of programs including Microsoft Word and Microsoft Notepad. (DAT Files, 2003).

DAT files are created by a specific application; typically accessed only by the application that created the file; may contain data in text or binary format; text-based DAT files can be viewed in a text editor. (DAT File Extension - Open .DAT Files)

5.6. Report

Tina Stevens, Head of English in Light Hall School, shares her experience in using the Enhanced Results Analysis (ERA). It is a free online model that provides an instant breakdown of exam results at the click of a mouse. With clear and simple snapshots, Enhanced Results Analysis provides information on how your school, subject, class and individual students performed.

According to Ms. Stevens, ERA makes her job easier and they were able to teach more effectively. Aside from that ERA helps them to:

- **Review** – on results days, you can immediately see and analyze your students’ results. You can also instantly compare the average mark your students received to the previous year, similar centers or all AQA centers.
- **Provide feedback** – after reviewing the results, you can use the mark scheme and examiners’ report to explain to students where and why they lost marks. You can also identify students who would benefit from a re-sit.
- **Address areas for improvement** – with the weaker areas highlighted you can use this insight to deliver personalized revision plans or amend your lesson plans.
- **Target set and plan for next year** – with year-on-year comparison of results you can easily identify trends and areas for improvement.

According to Dr. Raymond M. Zurawski, Associate Professor and Coordinator of Psychology at St. Norbert College, in his article “Making the Most of Exams: Procedures for Item Analysis”, one of the most important tasks confronting faculty members is the evaluation of student performance. This task requires considerable skill, because it presents so many choices. Decisions must be made concerning the method, format, timing, and duration of the evaluative procedures. Once designed, the evaluative procedure must be administered and then scored, interpreted, and graded. Then, feedback must be presented to the students. Accomplishing these tasks demands a broad range of cognitive, technical, and interpersonal resources on the part of faculty. But an even more critical task remains, one that perhaps too few faculty undertake with sufficient skill and tenacity: investigating the quality of the evaluative procedure.

The author of a leading textbook on the subject of psychological and educational assessment, Lewis Aiken (1997), contends that a “postmortem” evaluation, just like in medicine is necessary in classroom testing. This postmortem procedure for exams is called item analysis, which is a group of procedures for assessing the quality of exam items. Furthermore, the purpose of this item analysis is to improve the quality of an exam by identifying items that are candidates for retention, revision, or removal. More importantly, it can also clarify what concepts the examinees have and have not mastered. There are two broad categories involved in item analysis: qualitative and quantitative item analysis.
5.7. Qualitative Item Analysis

Qualitative item analysis procedures include careful proofreading of the exam prior to its administration for typographical errors, for grammatical cues that might inadvertently tip off examinees to the correct answer, and for the appropriateness of the reading level of the material. These procedures can also include small group discussions of the quality of the exam and its items with examinees who have already taken the test or even experts in the field. There are also faculty that use a “think-aloud test administration” (cf. Cohen, Swerdlik, & Smith, 1992) in which examinees are asked to express verbally what they are thinking as they respond to each of the items on an exam. This procedure can assist the instructor in determining whether certain students misinterpreted particular items, and it can help in determining why students may have misinterpreted a particular item.

5.8. Quantitative Item Analysis

In quantitative item analysis, three numerical indicators are often derived during an item analysis; Item Difficulty Index (p), Item Discrimination Index (D), Item Distractor Analysis.

5. Conceptual Framework

Figure 2. Conceptual Framework of the Proposed System

The Departmental Exam Analysis and Evaluation Report system processes scores from Departmental Exam. Knowledge on how to process the input and turn them into useable information and in correct format is vital to the success of the system. The researchers used Visual Basic, Net, the widely-used general-purpose object oriented language that is especially suited for LAN- based systems and is easily connected to SQL databases. Microsoft SQL Server 2005 Express Edition will be the Database Engine of the proposed system, thus, requires the proponents to know how to use SQL statement such as SELECT, INSERT, UPDATE, DELETE. To be able to come up with a interactive web application

Gathered information were analyzed and evaluated by the researchers. The researchers studied the errors and debugs during the testing and evaluation. During the implementation is the installation and publishing of the proposed system.

6. Methodology

The developed assessment model named as Departmental Exam Analysis and Evaluation Report system for CCSS is composed of three major components mainly the users of the system which are faculty members, Scantron output files or .DAT files, a Database and Report Output of the System itself.

This model is comparatively smaller and dependent to Scantron output files type of system which has a single main module, the faculty module. The faculty module has different access restrictions depending on status of the faculty member who is accessing it.

Mainly as Scantron result processing software, the functions of the Faculty Module are, Reading of .DAT files, Analyzing and Sorting of .DAT files, Checking of Student Scores, and Generation of Reports. The other component is the Database houses saved .DAT files and processed reports. The developed system shows only a single entity, the user or faculty member entity. This is the only type of user allowed to access the system since it deals with
academic sensitive information such as student name, number and scores and the actual answer key to the Departmental exam which has heavy consequences if an unauthorized person accessed it from the system. Specifically the users are faculty members of CCSS that administers subjects that needed Departmental Exams. First users need to gain access to the system. After doing so, users need to supply. DAT files of Student answer results and .DAT files of Exam answer keys. After those two requirements are entered into the systems which processes and saves the information, users generate reports.

The figure above provides a detailed explanation on the flow of data and main processes of the system. Users or faculty members needed to supply. DAT files. These files are namely Student Answer Keys and Exam Answer Keys. The Student Answer Keys .DAT file contains the student numbers who took the exam, the code of the exam and subject as well as the sequenced answer of the students for a single section. The Exam Answer Keys .DAT file contains the subject code and the correct answers in sequence. After the two .DAT file requirements are provided, the system sorts out the data contained in the files, sort and save them into 4 databanks or databases, namely: Student, Key, Subject and Scores databanks.

As soon as the information has been sorted and saved, the answers of students can now be compared to the Correct Answer Key and the result is saved into the Score databank. Once the scores are available reports can be generated. These reports are: Top 10 Students for each subject, Lowest 10 Students for each subject, All Student Scores per section, and Top 10 Mistakes for each subject and All Mistakes per Question.

7. Testing and Evaluation Procedure

After developing the system, testing and evaluation was conducted by the researchers to test the acceptability of the developed system. Normally, testing comes in two phases, the Preliminary Testing and Evaluation and the Final Testing and Evaluation. Preliminary Testing and Evaluation was done after completing the developed system. During this phase, the developers / researchers tested the system for bugs or errors. Because automated unit testing was not done during the development of the system, the developers / researchers made their own tests which is focused more on checking for errors or bugs in the integration of all the components.

After errors, bugs or flaws found after integrating all the components were corrected or fixed, Final Testing and Evaluation was done to test the acceptability of the developed system. During this phase, a survey was conducted to evaluate the acceptability of the system.

Direct Users and Technical Experts were the main respondents. Results from direct users include, professors and College of Computer Studies and Systems. The technical experts on the other hand consist of technical personnel who are involved in Application programming, Systems administration, system engineering and supervision. In groups, they were asked to use the system and gauge the level of acceptability with the use of the evaluation instrument.

Based on the survey instrument that was answered by each respondent, for each indicator of every criterion, a numerical equivalent scale is applied 1 means Not Acceptable; 2 means Slightly Acceptable; 3 is for Moderately Acceptable; 4 is assigned for Acceptable; and 5 is for Highly Acceptable. The accomplished Survey forms were collected. For each form, the average of each criterion was computed and the forms was classified according to type of respondent then tabulated. Then finally, the overall average score for each type of respondents were tabulated.

8. Summary of Findings, Conclusions and Recommendations

8.1. Summary of Findings

With all the procedures done in this research such as interviewing the professors involved in this system, programming, testing and evaluating, the
Departmental Exam Analysis and Evaluation Report has been developed and implemented.

This assessment model achieved an overall rating of 4.04, which makes the system acceptable.

In addition to the average being given by the evaluation of technical experts, the end-users also had their share. The system gained an average rating of 4.10, making the developed system also acceptable for the other end-users. It was observed that end-users had their lowest average being given to the functionality criterion in the evaluation form, having a 3.95 average rating.

8.2. Conclusion

Evaluation results being extracted in the answered forms were collected and interpreted in such a manner that it will dictate the acceptability of the system. Having an overall acceptability rating of 4.04, which is equivalent to acceptable, the system proves its effectiveness, accuracy and user friendliness that yields to correct interpretation of raw data that results to precise evaluation reports in a departmental exam of the College of Computer Studies and Systems.

The model provided for an effective assessment of departmental exam answers most of the problems encountered by the direct end-users. This make assessing student learning becomes easy.

8.3. Recommendations

Such is fact that software is continuously evolving and improving. In addition, most of the companies and researchers need holistic software that is complete enough with the provision of extensible modules in the system. For future studies, the researchers have recommended the following improvements, modifications, enhancements and alterations that will give a more refined version or replica of the system:

- A web-based version of the system so that faculty is not limited on using a single terminal computer and can accept other file type aside from .DAT file.
- A Departmental Exam Analysis and Evaluation Report software system made up of open source languages and programming frameworks.
- A system that will include a repository for historical reports such as the not retrieved evaluation reports.
- A system that integrates encryption techniques such as MD5 and AES encryption algorithms to secure confidential data.
- New learning assessment output should be added.

15. References


