Computer Assisted Audit Tools and Techniques in Real World: CAATT's Applications and Approaches in Context

Isabel Pedrosa [1], Carlos J. Costa [2]

[1] Instituto Superior de Contabilidade e Administração de Coimbra, ISCAC-IPC
Quinta Agrícola, Bencanta, 3040-361, Coimbra, Portugal
ipedrosa@iscac.pt

[2] Information Science and Technology Department, ADETTI-IUL, ISCTE-IUL
Av.ª das Forças Armadas, 1649-026 Lisboa, Portugal
carlos.costa@iscte.pt

Abstract: **Nowadays, Computer Aided Audit Tools (and Techniques') support almost all audit processes concerning data extraction and analysis. These tools were firstly aimed to support financial auditing processes. However, their scope is beyond this, therefore, we present case studies and good practices in an academic context. Although in large auditing companies Audit Tools to do data extraction and analysis are very common and applied in several contexts, we realized that is not easy to find practical examples to teach this subject. Computer Science students' and teachers usually use application manuals', which are adequate to explore the tools itself, but they are not enough to do real data analysis: a "hands-on" approach is essential. Despite that, our students didn't belong to any Computer Science graduation. CAATTs are good examples of software aimed to help auditors to analyze data, to explore and study distinct data sets in auditing processes. CAATTs represent an innovative approach to do data analysis, mainly because they allow the manipulation of very large data records with no significant added costs. Frequently, users have another difficulty: they are not able to adequate their data sets or to prepare their data in order to create useful input data, to do their analysis. In this paper, we present a proposal of extraction and data analysis, based on data sets obtained from a LMS platform (Learning Management System), a real data set that can be used to support students learning about CAATTs topic. Fundamental procedures to reach specific answers using a CAATT as a main tool to do data extraction and analysis are also referred. The context for this case study was an Information Technologies applied to Auditing course in a Master Degree graduation..**

Keywords: **CAATTs, Auditing, Information Systems, Fraud detection, E-Learning, Education, Knowledge discovery**

I. Introduction

Nowadays, audit processes are supported by several Computer Aided Audit Tools and Techniques, CAATTs. CAATTs can be defined as "any use of technology to assist in the completion of an audit. This (broad) definition would include automated working papers and traditional word processing applications" [2] or as "the use of certain software that can be used by the auditor to perform audits and to achieve the goals of auditing" [3]. CAATTs can be classified "in four board categories: Data analysis software; Network security evaluation software/utilities; Operating Systems and Data Base Management System security evaluation software/utilities; Software and code testing tools", according [3].

In this paper we are focused on Data Analysis software tools which, among other main uses, can be helpful to detect fraud evidences': "evidence are gathered by interviews, document reviewing, CAAT use" [6].

Several studies before state that mainly big companies have tools to do data analysis in data sets with large amount of data records [2], [7], [10]. Regardless of this usage and the fact that this tools are common in organizational context, students are not very used to work with it in a learning context [16]: software full packages and licenses are expensive and is not easy to use, create and prepared data sets. Sometimes, to take advantage of the empowerment of these tools, there is the need to do a complex data "makeover" process to allow the use of a data analysis tool, so the most time consuming operation can be data understanding and data preparation, most similar to SPSS, a tool that Master students were already familiar with. Other reason is software licenses: an education version could be enough to explore some data sets, however some vendors don't have it or the process to achieve it is complicated and too much bureaucratic. However, the Return on that Investment worth it: mainly, those efforts had good results among students.

II. ISCAC Contect about CAATTs

ISCAC, Instituto Superior de Contabilidade e Administração de Coimbra, a Polytechnic Portuguese Accounting, Administration and Management School, has a Master

Graduation Degree on Audit for Public and Organizational Enterprises that includes a subject focused on Information Technologies Applied to Auditing (ICT on Auditing). The first year was 2008/2009 and fifty-nine students attended the first edition; thirty-six, during the second and twenty during the third. Mostly, these students had their first contact with CAATTs in this subject. The assessment was based on students' labor experiences, which included education and training, city council, audit organizations and accounting offices. Students made use of that experience concerning the applications they chose to do data analysis using CAATTs software. However, to introduce and to practice the CAATTs' contents, we realized we need data sets to test the concepts, all the students at the same time, conducing to real learning and good practices that they can use, later, in their specific data sets. A demo educational version was available for students and teachers: classrooms and students' laptops were included. That educational version was limited to databases with less than 10.000 records. The full CAATT version "is adequate when data amount does not allow assuring audit process without a technological aid: paper amount will be impossible to manage" [3]. The course was taught using a Learning Management System, MOODLE, Modular Object-Oriented Dynamic Learning Environment, supported by several forums, documents, students' contributions and submissions. As in [11] and [12], "the power of Internet in learning management, trough an E-Learning platform, MOODLE, was explored as a tool to motivate students and to improve their final skills". In order to promote CAATTs' usage, this project was also inspired in [13], where is stated that Audit Command Language, ACL, an auditing software classified as Data Analysis was used "to help students become more "technologically adept"".

In many situations, original data sets are "prepared" just before being imported on CAATTs. A spreadsheet can be used on data "preparation". However, using the CAATT, new data fields can be added and calculated, taking other native or non native fields – this is also an important moment on the process: it was considered a core operation to learn how to program using functions specific to this CAATT. The CAATT functions' syntax was not similar to any other language they learned before. Indeed, only few students that had worked with DBase IV mentioned that the syntax was easy to learn and to use and similar to Dbase. However, in order to simplify this step, we've done the analogy with similar situations using a spreadsheet, a tool that all the students knew and felt comfortable with. Although formulas syntaxes' are different, the arguments and the main aim were similar.

III. Literature Review

CAATTs can be defined as "any use of technology to assist in the completion of an audit. This (broad) definition would include automated working papers and traditional word processing applications" [2] or as "the use of certain software that can be used by the auditor to perform audits and to achieve the goals of auditing" [3]. CAATs' use is, nowadays, accepted in data analysis mainly because it is "useful when large amounts of data are involved or complex relationships

of related data need to be reviewed programmatically to glean appropriate evidence from the aggregated data" [4]. They also "increase the efficiency of the conclusions about data analysis" [5]. Despite all the arguments look like so recent, we can conclude in [5], that CAATTs are not a new concept or approach: this study "An Audit Approach to Computers: A new practice manual" was published in 1978 and the paradigm is still changing as in [1]: GAS, Generalized Audit Software, running on Web 2.0, using a collaborative approach and several different profiles to improve Audit Software usage.

CAATTs can be classified "in four board categories: Data analysis software; Network security evaluation software/utilities; OS and DBMS security evaluation software/utilities; Software and code testing tools", according [3]. We can include IDEA, Interactive Data Extraction and Analysis, and ACL, Audit Command Language, in the first category and, among other main uses, can be tools used as fraud detection: "evidence are gathered by interviews, document reviewing and CAAT use" [6].

According [7], auditors accepted the CAATs, but more recent studies [8] and [9], suggest that CAATTs acceptance is very low, it is variable according to the companies and it is dependent on the company dimension. In [9], the Technology Acceptance Model, TAM, was used to study Information Technologies' acceptance among auditors and "the results from an experiment with experienced auditors suggest that firms have the ability to influence the implementation of new technology by using longer-term budget and evaluation periods and by communicating the approval of remote superiors regarding the software. In the absence of such firm interventions, the individual characteristics of the auditor (risk-aversion and perceptions of budgetary pressure) determine implementation decisions".

In [17], Unified Theory on Users Acceptance and Use on Technology, UTAUT, studied the "examination of contextual factors and individual characteristics affecting technology implementation decisions in auditing". We realized that the contact that auditors (present or in a near future) can have with the CAATTs during a Master Degree, could be relevant to mitigate the individual characteristics, difficulties and resistances referred before. Consequently, we decided to define specific supporting materials to teach Data Analyses Tools and to help teachers and students in this learning approach. Other factors in CAATTs acceptance can be relevant, whoever, they will be future work.

IV. Proposing a Process for data Analysis and Exploration using a CAATT

To do an efficient Data Analysis, we can use the same methodology defined as the main techniques on Data Mining and add new steps:

- • Extraction
- Exploration,
- Transformation
- Model building and validation
- Deployment.
- Reuse.

Extraction is the process to obtain the data set, usually from data center or a data server. On exploration and transformation, the process "usually starts with data preparation which may involve cleaning data, data

transformations, selecting subsets of records and - in case of data sets with large numbers of variables ("fields") - performing some preliminary feature selection operations to bring the number of variables to a manageable range" [18]. Indeed, in this sentence the statement "to explore data" is used with the meaning of "data understanding as an elementary step to be succeeded". Databases, attributes and data types are not easy to figure out and previous background knowledge on the area can be a core preparation on accomplishing the right result [19]. This paper focus on exploration, transformation and on CAATTs' usage to achieve efficient results.

CAATTs to do Data Analysis may also be referred as a tool to deployment and to reuse: after we answered to core questions on the data, we can create macros to set some parameters and to repeat all the process later with similar but new data.

V. Two Cases

Before we decided to create this Cases/Tutorials, we followed other approaches, namely, in the first year of the course, several IDEA data analyses applications were run during classes and assessments: IDEA for Education in "MOODLE data analysis in a Polytechnic School" and "Students grades and absences control in a Secondary School"; IDEA for audit organizations: "Inventory control"; and IDEA for Accounting offices: "Studding patterns using SAFT-PT file". As the MOODLE data can become the most usable data (because of the number of schools in Portugal that is currently using this e-Learning platform, we choose to improve that case study and to create a tutorial using data extraction from MOODLE Server at ISCAC, data transformation using IDEA and data analyses. The main aim was to study how could be possible to improve MOODLE usage at ISCAC taking advantage of students' background as MOODLE users.

A. CASE 1: using CAAT Data Bases as a first approach to use the tool

1) The very beginners: using IDEA Data Bases to a first approach to CAATs

IDEA has an example Project, composed by eleven databases IDEA Metadata files, IMD extension, with data sets for a first approach to IDEA to beginners. Using that data set, our main aims were:
- present, gradually, IDEA tool with examples to guide the students for real important data verification and validation, sums, patterns and error/"missings" detection;
- use native data sets (the ones distributed with the software), easy to use, to learn basic functionalities on IDEA;
- to promote class/students interaction and knowledge discovery based on students questions. The focus was on new information and not on the tool.

2) Elementary aims about CAAT tool

To achieve successfully our plan we choose *Sample-Bank Transactions*: it has only four attributes, it was simple, all the students had background on this subject, and, as a consequence on that, the focus was not on understanding complex data but on the right questions and processes we wanted the students to learn.

Sample-Bank Transactions has the attributes described in Table 1, 1166 records and some of those records are in Figure 1.

Table 1. Bank Transactions Native attributes

Name	Description	Comments
TransID	Transaction number	This must be unique (primary key)
Type	Cheque or Deposit	It must be filled in
Date	Day for the transaction	It must be filled in with a data value
Amount	Transaction amount	Depending on transaction Type, the amount must be positive for Deposits (Credits) and negative for Cheques (Debits)

Figure 1. Bank Transactions: first records

Using *Sample-Bank Transactions*, students were focused on the answers to the following questions:
- Are there any null Deposits or null Cheques? If yes, in which dates did it happened? Are there other important transactions during the same reference days?
- Are there any duplicated transactions?
- Are there negative Deposits and/or positive Cheques? What are the dates? Are there other transactions during those days?
- Which are the transactions with the highest values? Can we find a justification about it?

Most of the previous questions were stated by students: after, the focused on learning how to answer the questions with IDEA.

3) IDEA functionalities

To answer last questions, we realized that a "hands on" approach could the most effective. We just introduced briefly the IDEA tool, explored the main aspects in the application main window, as in Figure 2, and we started to answer the questions stated just before.

They learned how to:
- Use **Field Statistics** to find the null values and store it on a new database named *null transactions*: they found 4 Deposits in 17-02-2008, 07-05-2008, 26-07-2008 and

01-10-2008. Using the functionality **Join Databases** to cross *null transactions* with *Sample-Bank Transactions* they discovered all the transaction done concerning the referred days.

- Use **Duplicate Key Detection** to detect the duplicated transactions: only one, the TransID 1505, referring to two Cheques -11.359,36 and -9.758,84.
- Use **Summarization** to group records by Type, create two new databases one for *Deposits* (769 records) and the other to *Cheques* (397 records). **Criteria** and **Equation Editor window** were introduced to explain how to filter the negative or positive amounts in each database. No irregular records were found.
- Use **Top Records Extraction** to check the top 5 extractions: a total of 7.826.192,90, mainly on 2 transactions 3.899.925,00 (TransID 125) and 3.536.451,99 (TransID 420).

After this exercise, planned for 2 hours, the students felt much more confident with this tool and the aims that we established just before. We proceeded, analyzing the databases **Sample-Detailed Sale** and **Sample-Sales Representatives**. To explore the formulas and to use IDEA as a tool to discover information in a specific and real context, we decided to use MOODLE data.
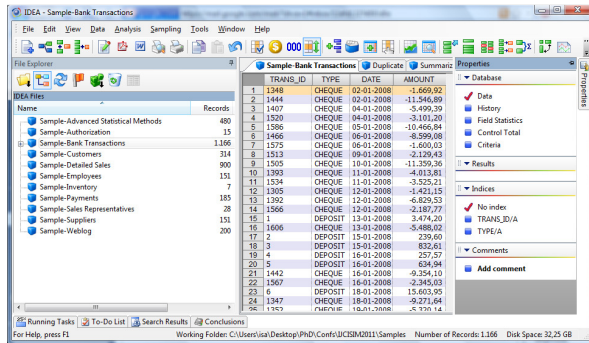

Figure 2. Main IDEA Window Application

B. Case 2: LMS data analysis in a Polytechnic Institute

ISCAC is a Portuguese Accounting, Administration and Management School, focus on Accounting, Administration, Management, Informatics and Law (Solicitors). MOODLE is being used as an e-learning platform since September 2006. Considering the period from the beginning until August 2007, two scientific working papers were published about the usage of MOODLE among students and teachers at ISCAC. In [14] and [15] it is possible to find a previous data analysis, made with Access and reporting to the first year of MOODLE usage at ISCAC, 2006-2007considering all the interactions between students and the platform, the contents available on courses as fora, chats, tests as interactive tools with course community, files and links. However, after this work no new studies about MOODLE courses, users and users-interaction with this platform were developed. The first time we use a similar data set was on the Master Course 1st edition. After it, we realize that this data set could join all the topics that we should focus on a course like this and, afterwards, this approach could be "exported" to other schools.

IDEA analysis about MOODLE main aims' is to explore several IDEA functionalities using real data and a familiar application context for all the students, using IDEA as the main tool for data extraction and analysis. As in a data mining tool, we focused on data understand, to define the questions considering the set we have, prepare and transform data attributes to have a real answer. During all this steps we reviewed some aspects from our 3.1 example and have the opportunity to learn several new others. Adding macros to this tutorial in order to "redo" all this analysis with new data extracted from MOODLE, was our last conquer.

1) Data description

In this study, we defined the following topics to do MOODLE usage analysis: courses created on MOODLE (courses table), Categories (courses_cats table), forum usage (forum_discuss table) and posts related to the forums created (forum_posts table) and users registered on the platform (users table). The Educational IDEA 8.04 had 10000 registers as limit for each database and this was the main reason why we didn't get any other indicators. However, we will refer that on topic "conclusions and future work": indicators such as courses and user activities and MOODLE platform logs are important data to know in a future work. For instance, we realized that the number of invalid logs on the system was very high and one explanation can be the new users' creation process: the can create their own account and choose their password, even if they don't belong to ISCAC. Sometimes, if they forgot the password, it is possible to create another user: the only criterion is that the email address can't exist already in MOODLE users' data base. However, this procedure can represent a lack of security and a risk to the organization. In Table 2, we have each database description with all the attributes and, in Figure 3, a COURSES_CAT database view. The data set was extracted from version 1.9+ .

Table 2. MOODLE Databases

Name	Contents	#records	Data preparation
COURSE	All MOODLE platform courses, each with this fields: ID (N), Category (N), Sortorder (N), Password (C), FullName(C), Shortname(C), Summary(C), Format (social, topics) (C), StartDate(N), Visible (N), Timecreated (N), Timemodified (N)0.24	395	Date and time in Unix Timestamp format: number of seconds since January, 1, 1970. New fields created to store this data in IDEA Date and Time format
COURSES_CAT	All MOODLE platform categories: ID (N), Name (C), Description (C), Parent (N), Sortorder (N), Coursecount(N), Visible(C), Timemodified (C), Depth (C), Path (C), Theme (C).	73	Beyond date and time in Unix timestamp format, understanding the hierarchical categories and dead with paths
FORU	All MOODLE platform	3113	New virtual date

M_DIS CUSS	top discussions: ID (N), Course (N), Forum (N), Name (C), Firstpost (C), UserID(N), GroupID(C), Assessed (C), Timemodified (C), TimeStart (C), Timeend (C), NewID (C)		type fields created to store date and time converted from Unix timestamp format
FORU M_POS TS	All MOODLE platform posts in every forum discussions: ID (N), Discussion (N), Parent (N), UserID (N), Created (C), Modified (C), Mailed (C), Subject (C), Message (C), Format (C), Attachment (C), TotalScore (C), MailNow (C)	6038	New virtual date type fields created to store date and time converted from Unix timestamp format to IDEA
USERS	All registered MOODLE platform users since the beginning: ID (N), Auth (C), Confimed (N), Deleted (N), Username (C), Firstname (C), Lastname (C), Email (C), Emailstop (N), Firstaccess (N), LastAccess (N), LastLogin (N), Currentlogin (N), LastIP (C), Mailformat (N), Maildigest (N), Maildispay (N).	4101	Last IP was only with 14 characters: all the ones which had 3 digits for each network were converted in mumbers. Also Date and Time format as before.

The attributes, considering the example of database COUSES on Figure 3, were hard to understand for the students before the first data transformation, mainly in what concerts to the date/time values: records ID=2 and ID=3 have STARTDATE=11498076000 and TIMECREATED =1149792990, which are unusual values, but possible and correct as we will see later. To understand the courses hierarchical places, we need to join this table with COURSES_CAT, to find the path attribute and the name of the category.

Figure 3. COUSES database

2) Questions about MOODLE data

As in "Data Preparation" column in previous Table 2, the need to do data transformation was detected in order to be possible to answer some important questions:

- number of registered (new) users each year (academic year);
- number of users with @iscac (teachers) and (@alumni.iscac (students) emails to find a pattern on users account and the connection to email servers at ISCAC. Only teachers have that email server, so it can be a distinction between teachers and students, however, a teacher can register himself with another email address. NOTE: at MOODLE, the same user can act as teacher in some courses and as student in other context. With these databases, there is no way to know if someone has a role as teacher or as student in a specific course because there is no way of join users and courses;
- Non-confirmed and deleted users;
- Users last logins: time last since last login, users' last IP access;
- Top users, categories and courses: in forum discussion and in forum posts.
- Patterns for forum activities and forum discussions (morning, afternoon, night, weekends);

After data analysis, several new data fields were add to store important information and IDEA functions were applied to create those new fields.

3) @Functions and IDEA: data transformation and some brief examples

All these time calculus are based on the principle that the day was stored in Unix timestamp as the number of seconds since January, 1, 1970. All these concepts were new for these students as they are present (or future auditors) but not computer science experts. Also all the concepts about using formulas on IDEA were novel: only one student which was expert on Dbase IV felt comfortable in this new context. Most students were also experts on Excel so, another approach could be: to do all the data transformation on Excel and then to import it at IDEA. However, that methodology was not followed because to be expert using IDEA formulas was one of our previous aims.

Some new fields were created using the menu **Data>Field Manipulation** and the option Append to compute data in a way that could be easiest to answer questions evolving dates in last topic:

- Date: all the attributes mentioned as date types (however, on IDEA and until then, this attributes were numbers type). New attributes (or fields) created were FIRSTACCESS_DIA, LASTACCESS_DIA, LASTLOGIN_DIA, CURRENTLOGIN_DIA. New field LASTLOGIN_DIA (and all the other new fields above), data type will be *Virtual Date* and the formula should do the conversion between the number of seconds that we have, as explained before, and we present it using IDEA. A possible formula to create a new field new *virtual data* beginning with, for instance, LASTLOGIN Unix Timestamp, 1246669569 seconds, should be converted to

04 Jul 2009, using this formula on Equation Editor like stated on figure 3:

@if(LASTLOGIN
<>0;@DaystoD(LASTLOGIN/(24*60*60)+DtoDays(@Ctod
("01-01-1970";"DD-MM-YYYY"))))

With this new fields will be possible to extract month, year and to do a ***Pivot Table*** presenting new users for each academic year. We suggest the students to check their formula results on http://www.timestampconvert.net/.
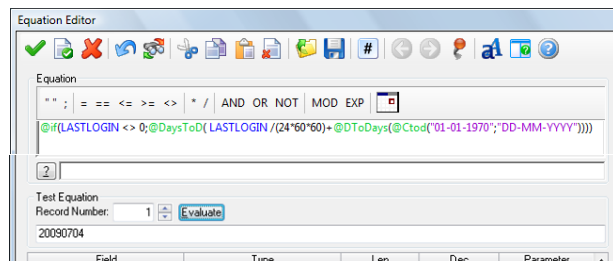

Figure 4. IDEA Equation Editor and Data Field Manipulation

- Time: like before, on date, several new time fields were created, converting the timestamp value to time.FIRSTACCESS_HORA, LASTACCESS_HORA, LASTLOGIN_HORA, CURRENTLOGIN_HORA. With this new attributes we can have a usage platform pattern taking the time of the day. For instance, if TIMECREATED =1149792990, this means 18:56:30.

- Forum discussion and users: The process to calculate date and time for several accesses is the same as defined before, but, in this case, we want to obtain a graph with the top ten users that posted more on forums'. On *forum-post database*, we must summarize data by ID (*user ID*), using the option ***Summarization***, save the result in a new Database, *Total forum discussion by users* and join it with *Users* data to obtain all the data about users, in order to present it, for instance, as a dynamic graph, as in Figure 5.
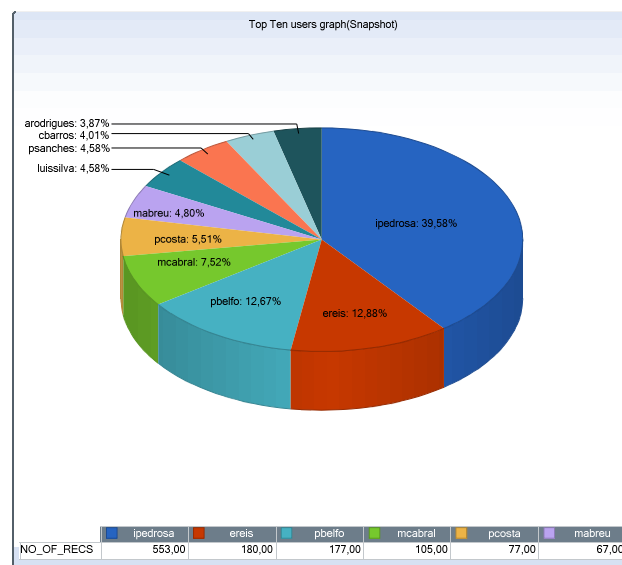

Figure 4. Users discussion top ten (all forum). Note that the top user has almost 40% of all top ten-forum discussions

4) Answers on MOODLE data set

After the data transformation and some new attributes creation, we can answer to the questions we stated before, using several IDEA functionalities and, in some situations, new attributes and calculus:

- number of registered (new) users each year (academic year): taking the USERS table, we created a new attribute, year. After, we used ***Summarization*** to group all the records (# 4101) by year of register. The result was 725 records in 2006, 1847 in 2007, 1203 in 2008 and, until 3-07-2009 (our last record on the USERS table), we have 304 for 2009 year.

- • number of users with @iscac or @alumni.iscac emails. A new data field was created only to store if the email had the pattern @iscac.pt, with the formula @Match(EMAIL;"@iscac.pt"), and another to the string "@alumni.iscac". The result was 134 records for the first and 453 for the last (only 11%). Therefore, we can conclude that most teachers use their ISCAC mail account but few students are connected with their institutional mail. This can mean that students don't check often their institutional mail –probably, they don't receive any Students Office messages or any other institutional ones. A solution was to do a compulsory update of all the users, with the institutional.

- • Non-confirmed and deleted users: criteria *confirmed=0* and create another database with that 25 records. All of them were Erasmus Students. As MOODLE is only available in Portuguese, this could be an important disadvantage. Therefore, as the school intend to proceed with its internationalization, to receive more students from abroad to attend graduation and master courses, there must be a MOODLE English version. Using the criteria deleted=1, there are 237 deleted users. All of them have a suffix with a timestamp in the *email* account

5) Discussion on MOODLE data evaluation

We could conclude from all MOODLE data:
- number of registered (new) users each year (academic year): this number has a significant grow in 2008. Many new courses were created during that year. We can predict that this number won't be higher in the next years: although there are new master and post-graduate courses, many of the attendees were already students at ISCAC and the MOODLE account was not erased;

- number of users with @*iscac* emails: There are 134 registers and the number has no significant grow since 2007/2008. Since September 2008 there are only 16 new registers. The main justification is the registration process for almost all the teachers in the first years of the project and there are almost no new teachers' admissions recently.

- Non-confirmed and deleted users: 237 deleted users and 25 non-confirmed users. Analyzing this last data, we found that these are new records, with non Portuguese names, so., probably, it could mean that, when we receive foreign students, we must create other solutions for this students.;

- Users last logins: time last since last login: calculating the number of days since last login until the data reference day, and then doing a first *stratification* using IDEA, we realized that 1986 users were on the platform at least on the previous 50 days. Doing a shorter stratification, 1830 users were on MOODLE at least on 10 last days. We must exclude also users that logged only once in MOODLE, 682: in this case, LASTLOGIN=0, 17% from total users. The users that left ISCAC were not erased from Server databases, which create a larger number for the ones who didn't log on for more than 200 days (since the beginning of this scholar year),

- Users last IP access: considering last logins and knowing which are the IP addresses belonging to ISCAC intranet, is possible to state that only 104 users logged on from ISCAC last time. However, to define if users use MOODLE, mostly out of school or preferentially "inside", we must have other details: we can forget that we have only each MOODLE user last access and to answer the previous question we must have all the logs. Again, the IDEA version creates the limit to our analysis, but we can define that the study will concern only with users' activities during last four months.

VI. CONCLUSION and Future Work

CAATTs are stated as important software in what concerns to Auditing and Data Analysis However, there are some problems when the focus is on CAATTs learning during a master course: real and interesting data sets with no specific additional background knowledge for the students are not easy to find. However, firstly, the focus should be on the data and not on the tool: students must understand the data and define the opportune questions to the data analysis and that process must be completely independent from the tool. A simple data set was taken to do the first approach to this method. After, we studied a real data set, to improve students competences on CAATTs as a data Analysis Tool and to understand the patterns concerning to MOODLE usage. Several indicators were presented using IDEA as a data extraction and analysis tool. The last focus was on using a CAATT, IDEA, in Information Technologies on Auditing course in a Master Degree graduation. It was relevant to use data that students could easy understand and manipulate. Additionally, it was important to define new approaches after data analysis mainly to improve MOODLE usage. We realized several new ideas to increase MOODLE users range and, concerning the application, doing that data analysis and report it with IDEA was easy to understand and with no significant time spare.

In a near future we intend to systematize this study with macros to reuse the analysis, to check the main results with new data, with some parameters to define the filters. To spare this practices that we followed, we can create partnerships with other schools to explain the method and analyze other new cases.

Many data on MOODLE can also be used as real data set to explore other CAATTs functionalities: data sampling is a good method to check the patterns on non succeeded logs attempts.

Another topic could be to study SAFT-PT, Standard Audit File for Taxes Purposes – Portugal, using tests and macros able to support data check, to cover all the topics not available on commercial analyzers. In this case, we should focus mostly on taxes issues; define a transversal study team, with specialists on taxes, taxes inspectors and supervisors. As this is a very important subject, in Portugal, probably the analyzers will be improved very soon.

Acknowledgment

References

[1] N. Gehrke, P. Wolf, "Towards Audit 2.0 - A Web 2.0 Community Platform for Auditors," hicss, pp.1-10, *43rd Hawaii International Conference on System Sciences*, 2010

[2] R. L Braun. et al, , "Computer-assisted audit tools and techniques: Analysis and perspectives" *Managerial Auditing Journal*, Vol. 18, Issue 9, 2003, pp. 725-731.

[3] S. A. Sayana, "Using CAATTs to Support IS Audit", "*Information Systems Control Journal*", Volume 1, 2003

[4] J. Kramer, "*The CISA Prep Guide: Mastering the Certified Information Systems Auditor Exam*", Wiley Publishing, Inc., Indianapolis, Indiana, 2003, pp. 43

[5] B. Jenkiens, A. Pinkney, "*An Audit Approach to Computers: A new practice manual*", The Institute of Chartered Accountants in England and Wales, 1978.

[6] M. Popa et al, "Characteristics of the Audit Processes for Distributed Informatics Systems". *Informatica Economică*, Volume 13, Number 3, 2009

[7] D. Janvrin, D. J. Lowe, and J. Bierstaker, "Auditor Acceptance of Computer-Assisted Audit Techniques", Working Paper, Iowa State University, Iowa, 2008, URL: http://aaahq.org/meetings/AUD2009/AuditorAcceptance.pdf (accessed: 2010-09-15)

[8] R. S. Debreceny et al "Embedded Audit Modules in Enterprise Resource Planning Systems: Implementation and Functionality", *Journal of Information Systems*, Vol. 19, Nº. 2, 2005 , Pages 7-27

[9] M. B. Curtis, E.A. Payne, "An examination of contextual factors and individual characteristics affecting technology implementation decisions in auditing", *International Journal of Accounting Information Systems*, Vol. 9, Issue 2, Elsevier, 2008, pp. 104-121.

[10] H. Kim, M. Mannino, R. J. Nieschwietz, "Information technology acceptance in the internal audit profession: Impact of technology features and complexity", *International Journal of Accounting Information Systems*, Volume 10, 2009, 214–228

[11] B. N. Potter, et al, "The effect of interactive on-line learning systems on student learning outcomes in

accounting". *Journal of Accounting Education*, Volume 24, Issue 1, 2006. Pages 16-34.

[12] M. L. Ripley, "Teaching Better Electronically or: How I Learned to Stop Worrying and Love Internet Teaching". *Journal of Systemics, Cybernetics and Informatics*, Volume 7, Number 4 , 2009, Pages 27-31

[13] F. J. Coglitore et al, "The Use of Computer-Assisted Auditing Techniques in the Auditing Course: Further Evidence", "*Journal of Forensic Accounting*", Vol. VIII, 2007, pp. 201-226

[14] I. Pedrosa, C. Sousa, "The Actor Network Theory and Moodle in an High school community – A strategical tool and a life-long learning Approach". *IADIS International Conference on Cognition and Exploratory Learning in Digital Age*, CELDA 2007, 2007, Faro, Portugal

[15] C. Sousa, "*O Moodle: implementação e estudo de impacto no contexto da comunidade*", Graduation Project, ISCA, 2008, Coimbra

[16] R. C. Richardson, T. J. Louwers, "Using Computerized Audit Software to Learn Statistical Sampling: An Instructional Resource", *Issues in Accounting Education*, Volume 25, 2010, pp. 553

[17] M. B. Curtis, E. A. Payne, "An examination of contextual factors and individual characteristics affecting technology implementation decisions in auditing", *International Journal of Accounting Information Systems*, Volume 9, Issue 2, 2007 Research Symposium on Information Systems Assurance, June 2008, Pages 104-121

[18] Data Mining, predictive, modeling, Techniques, http://www.statsoft.com/textbook/data-mining-techniques/, (accessed: 2011-02-25)

[19] I. H. Witten, E. Frank, "*Data Mining: practical machine learning tools and techniques*", Morgan Kaufmann, 3rd edition, January, 2011

Author Biographies

Isabel Pedrosa teaches in the Mathematics and Computer Science Department of Instituto Superior de Contabilidade e Administração de Coimbra, ISCAC, since 2000. She is Master in Computer Science and has a Degree in Computer Engineering by Coimbra University, 1994. She teaches since 1993. Now, she is preparing her thesis about the main topic CAATTs, in Doctoral Program in Science and Information Technologies at ISCTE - IUL. She coordinates courses on Information Technologies on Auditing in Master Graduation Courses.

Carlos J. Costa teaches in the Computer Science Department of ISCTE, since 1993. He also taught in the Portuguese Catholic University, Portuguese Open University, IPAM and ISCEM. He worked in a IT consulting firm and in an Investment Bank. He has a Ph.D. in Computer Science - Information Systems and Data Bases, ISCTE, 2002, MsC in Corporate Strategy, ISCTE, 1995, and Degree in Business Administration from the Faculty of Economics and Business of the Portuguese Catholic University (Lisbon). 1990