ABSTRACT

Psychometric questionnaires are a field of the Psychology very important in the psychological evaluation of a subject, these questionnaires, or analytic models, are frequently ministered used in sessions using the traditional method of collecting data, that is the paper-based questionnaires. Nowadays exist some applications that provide analytic models, however most of them are not focused in clinical practice or they are not validated by any study that proves that their results are as reliable and efficient as the paper based method. With this master thesis we develop a system that tries to copy the paper behavior by using some metaphors related to the traditional method, and evaluate the system performance comparing different environments in terms of time, error. This evaluation goal is to see if the purpose of creating a similar experience to the traditional is achieved.

Author Keywords

ACM Classification Keywords

INTRODUCTION

Psychological evaluation mechanism of nowadays are in general non-digital solutions, these master thesis work around digital solutions that produces as outcome an automatic report generation of a psychological evaluation. We are trying to solve two problems that are:

- Double Data Entry – is known by the double entry of answers in to the paper based survey and copying all these answers to the digital format manually
- Ecological – The amount of paper required to minister a session is significantly high.

To solve these issues we studied how a psychometric test is created, validated and ministered in psychological sessions, what dimensions are frequently evaluated in these analytic models (AM), we analyzed papers like “Touch Survey: Comparison with Paper and Web Questionnaires” and “Comparisons of Electronic Data Capture (EDC) with the Standard Data Capture Method for Clinical Trial Data” , those papers show how digital solutions that provide analytic models behave and how they compare between themselves, we want to see if the answers are as reliable as in a paper-based AM, does it yield the same results when comparing the analysis of Electronic Data Capture (EDC) and paper-based AMs, and if they are different is it possible to align these results. We also search for others digital solutions and made an evaluation to them taking in to account some requirements that we define for the system. The evaluation show that in fact not all these requirements are fulfilled. We design a solution that tries to cover all the requirements, since user interaction requirements, in these requirements we split the user type in three different types like the Interviewee, the Observer, and the Administrator. Our solution passes by the development of a progressive web application that allows the entirely work flow of a psychological evaluation, since the creation of analytic models, the ministration of these analytic models in a session, and the automation of the report generation to decrease the time needed in the calculation of psychological analytic model score. To ensure that the application is on the right track, i.e. the application is reliable & secure and has a user experience like the traditional method, we used some frameworks and cloud services to better handle the job like Amazon AWS and Google’s Material Design. At last to guarantee that the application is mimicking the traditional method with the best user experience we evaluate the system comparing three different environments like the Tablet, Computer and the Paper, in this evaluation we compare the environments in terms of time and error and we also retrieve from the users that test the system their feedback about what environment is more practical, dynamic, quicker and easier.

CONCEPTS

There is a couple of concepts that are necessarily to a better understanding how our digital solution was developed among them we described the meaning of Psychometric...
Questionnaires and our own meaning of Session in the system.

Psychometric Questionnaires
Psychometric measures and evaluate the theory of creating mental processes, giving a meaning to the respondents’ answers. [1] The standard definition is described as an assignment between the answers and the evaluation objects by following rules [2]. The definition also includes the four measurement typologies like nominal, ordinal, interval and ratio. [3][4] However inside of this theory there are a couple of concepts and models that evaluate the respondents’ answers differently, in our project we only took the Classical Test Theory [2] that evaluates a test by is global pack of items of evaluation, and the Item Response Theory [2] that evaluates the meaning of answering to the individual item of the test. Although these models are more than analyzed and are dissonant both share the same parameters of measurement like reliability and validity [2]

- Reliability refers to efficiency of the analytic method under evaluation
- Validity refers to the congruence between the instrument used for measurement with the dimension under evaluation

System Session
We define Session as a work space to Interviewees, Observers and Administrators execute tasks that are specific for each one of them. Each Session has a group of tools such as AMs, interviews and other analytic methods to extrapolate outcomes like reports, values, observations & annotations.

We define a Session flow that explains how the session should be created and used, there are three major phases to do so:

1. Design – correspond to the AMs design, what analytic methods should be included, how they are evaluated, among other options.
2. Administration – appliance of the design phase outputted for the interviewees and observers to collect data.
3. Computation of the data collected in the Administration phase, and download of reports by the Administrator.

The tools that we make available to the administration of a session are the Psychometric Questionnaire/Analytic Model, Interviews, or any other analytic method that is an open classification tool that allows the Observer to give a value inside of a range to the Interviewee.

The reports are the output of a session, this is a PDF file that shows all the annotation and individual score of the interviewees in all sessions tools, it can include the global score of the session.

RELATED WORK
This section we split the area of concern in two major concepts, one regards how the AMs methods are administered and if it has any influence on the score and the other is to analyze related systems to ours.

Methods of administration
There are several ways of applying/minister the AMs and among the paper-based is the most used, it is the standard way of collecting data from people, it has a couple of advantages due the fact it does not distract the Interviewer of the system issues. However due the envelopment of mobile and ubiquitous computation the idea of using digital means to collect data is becoming more attractive, EDC allows data to be collected and analyzed while the interview occurs and can reduce the amount paper needed to do the psychological evaluation.

Touch Survey: Comparison with Paper and Web Questionnaires is a study to validate the development of a tool allowing the interaction comparison between different types of devices and paper based, they develop an application to evaluate different environments. In fact they evaluated and concluded that their prototype was as fast as the PC-based version, and they also describe that their application has higher values of satisfaction due to is usability on the tablet-version, and it is also the environment with higher values of accuracy.

Comparison of Electronic Data Capture (EDC) with the Standard Data Capture method for Clinical Trial Data, goal was to analyze the interviewee and respondent relations during the sessions by using different types of data capture methods like PC, Tablet, Mobile Communications, Paper and PDA. [5] In the overall the all the environments show similar results in terms of error and precision, meaning that in fact that the digital solutions results are highly correlated with non-digital solutions like the paper method.

Related Systems
In this section we explain characteristics and architecture of applications and services that use EDC methods. Applications like SurveyMonkey, QuickTapSurvey, Poimmaper, Akkro and OpenClinica provides to the user the possibility of writing surveys and ministered them globally. All of them use, except OpenClinica, use cloud service computation like Amazon AWS for the deployment of their services. SurveyMonkey [8] use a architecture of Model View Controller [7] to fulfill and solve scalability issues that they had in the first version of the application. Their model layer is responsible for the database schemes, and data management, the view is described as the interface to the client interact and the controller manage the data flow between the model and view layer. QuickTapSurvey provide a AMs management system. [9] [11] One of the features of this system is the possibility of recall offline data. There is a group of possibilities how a AM can be administered ranging from a face-to-face interview, “kiosk” mode used in retail stores like H&M, ZARA and others, by digital device frequently a tablet. The system administrator can add/delete and edit other users and delegate tasks to them, to support their services they use cloud computing in order to ensure
fault tolerance while distributing the work load for all the instances that they use among various regions of AWS [10]. Poinmapper allows the data collection using mobile devices and shares real time data. This application also works offline because it is more focus to collect data on the field and in adverse conditions that not always an internet connection is available. As the QuickTapSurvey they use AWS [13] to host their major features that are the form builder where the AMs are build, Web Service Interface to access information stored in the database manly these records are used to transmit electronic health records, Administration and quality assurance – define system authorization roles giving distinct responsibilities within their team, Mobile Application that allows the data capture on the field and a portal responsible for the data visualization and edition. [14] Akkro is a mobile software solution to collect data. [15] Each AM in the application is generated as an individual application with authorization models. The system has a dashboard control panels that allow users to collect and export the data collected, the forms are written in HTML, what can be a major benefit if the user wants to fit this system inside of their own website page. OpenClinica is an open-source software as a service with intention of clinical research [16]. It is described as lightweight and expandible application due the fact is fully controlled by the company that want to use it as collection tool. It is fully developed like a web-based application and it is possible to deploy in almost every platform as a service like AWS EC2 instance, Google Cloud Compute Engine, or any other cloud computing service on demand, or webhost. [6] Their architecture is described by the client company requirements and needs the only thing that must be ensured is the existence of internet communication with OpenClinica authentication server.

RELATED SYSTEM EVALUATION
In this chapter we describe our evaluation of these related system accordingly to requirements that we need when designing and developing our system

• Work Online & Offline – it is possible that in some environments internet communication is not available or even the server is under maintenance, even so observers must be able to collect data while being offline.
• Repository – with this feature we want to see if the system allows group work, if an administrator can delegate tasks/roles between collaborators.
• Interface – in this metric we want to see if the interface is intuitive to achieve a better productivity, and if the interface shows all the information needed for the observer or administrator work be done.
• Security – We want to evaluate if their systems guarantee confidentiality, integrity and availability. Our model takes in to account if the systems use secure communication channels (SSL), and if the data is not being stored in clear on the servers.

In general most of the systems has some features that matches our requirements however there is not one which is fully compatible, so when we designed our solution and implementation we had these features in mind and tried to arrange them to a better fit to the system requirements.

FRAMEWORK & SERVICES
There is a bunch of frameworks that served our development like Google’s Material Design, that is a framework that allows the creation of interface with a visual language, and Amazon Cloud Service AWS, that hosts our servers and Databases.

Material Design
It is described as “framework for interface creation, the purpose is to create a visual language that synthetizes the classic principals of good design with the innovation and possibility of technology and Science”. This framework has goal to create a new technology framework that allows the correct design of interface, the conceptualization of models that were responsive in every device, since the mouse to the keyboard and even touch devices. For that they develop a bunch of metaphors like: [17]

• Material as a metaphor- the theory behind it give the sense of motion and individuality of the elements, has to feel “real” touch, like you would feel if you touch a piece of paper, or move a card form one position to another.
• Bold, graphic, intentional – There is a hierarchy between elements to emphasize focus and meaning.
• Motion provides meaning – All user interactions should force the system to give a meaning to transformations making available the continuity between user interactions.

In material design the physical properties of paper are translated to the screen. The background of an application resembles the flat, opaque texture of a sheet of paper, and an application’s behavior copy coats paper’s ability to be re-sized, shuffled and bound together in multiple sheets. This idea served our development, because one of the major concerns in our system is to develop a similar experience like the classic method that is already being applied. Since the classic method of AMs appliance is the paper, if we use this as metaphor, that was analyzed and studied by Google, we can use it to evaluate the similarity between these two methods, the classic against the EDC.

Amazon Cloud Service AWS
Cloud computation services allows companies, end users and other entities to use computational power without the need of servers’ maintenance. The price model for using a platform as a service like AWS depends on the tier of usage. For a common user that only uses one server to show his website this model is perfect since there is a free tier with license of one year to fully develop and deploy their web
There are advantages of using this type of hosting like:

- **Trade capital expense for variable expense** - There is no need to invest in data centers without knowing if the systems needs to scale.
- **Benefit from massive economies of scale** - Hundreds of applications made by users are reachable for each other achieving higher economies of scale.
- **Stop guessing capacity** - The system scales without the user interaction for that purpose.
- **Increase speed and agility** - Since hundreds of applications are reachable by each other it is very fast to deploy new IT services.
- **Stop sending money on running and maintaining data centers** - Application focus against the needs of data center maintenance.

This service benefit us in almost our systems requirements, since two of our main concerns during the development were the security of data, and the system availability. With the security data we used the VPC service form amazon to the machines communicate between themselves, like a controller server send data through a private channel to a database server, or even protect all of our machines with a pair of keys that only gives access to the user that has the public key. The availability is achievable by activating the ELB mechanism, that allows the system to with some metric measures to load balancing the work among the machines available in the system. In a case of use this allow to the servers respond a huge number of users to access to the system services without increasing the latency of the responses. To keep ahead of the current state of the EC2 instances amazon provide us a dashboard where is possible to control all the instances and tools to improve performance.

**SOLUTION**

Our solution was developed & designed within the range of the MAYS formal requirements for their business, within the constraints and requirements of the users of the system, and to support all of this, we defined a domain model, arrange interviews with Psychologists to understand how they annotate sessions, search for the best framework to create a most similar experience to the actual method, a flux diagram of the main processes of the system.

**Domain Model**

We define the domain model, which takes in to account all the requirements of MAYS (Figure 1), these model is composed by the following entities:

- **Interviewee** - is the entity responsible for participating in our system sessions and for so he must be identified by an ID, has to have a name, the date of his birth, gender and company that he is attached.
- **Observer** is the entity responsible for supervise a interviewee session. This identity has a unique identifier and password in the system to keep the data that he collects traceable and secure.
- **Administrator** - it administrates all the parameters of all sessions, have access to all sessions and because of that it must have an identifier and password to secure data stored in the system.
- **Session** - it is the major entity of the system, all the data that is being collected, analyzed, manipulated occurs inside of a session. So a session is defined by a identifier a pin to stop insecure access to a session, and a date that session will happen, can contain Analytic methods like questionnaires and interviews, has other entities that participate and manipulate data like Interviewees and Observers. It has also all the answers that are associated to Interviewees by their identification.
- **Questionnaire** -It is defined by having a type of answer associated to them, can be ordinal, nominal, ratio or interval. Has is supposed to the system evaluate psychological and social dimensions, each questionnaire has a group of dimensions and sub dimensions.
- **Dimensions** - this entity can include others dimensions has a list of sub dimensions, there is also another entity visible here that defines the way how this dimension is evaluated.
- **Computation** - Is the entity responsible for converting data stored in the system in something with real meaning to the system owner, it is the bridge between the input and the output. This computation can be the Average, Sum, Product or other statistical formula.

![Figure 1 - Domain Model](image-url)
• Question - This entity has the text of the question and a list of dimensions that it could be included
• Answers - This answers stores for each session and Interviewee all the answers that were collected through the system, this answer will serve as input to the Computation entity and had the particularity of including also the annotations of Observers.

User Interface Experience & Interviews
The user interface was split in two types one for the Interviewees and other for the Observers and Administrators the only thing that distinguish them is the fact that the Administrator has more privileges.

• Interviewees – the main concern with these users is the fact there is not only one type of user, this is a wide range group respecting the labor life. There are users that never touch this kind of interfaces and they are not familiar with new technology, as there are a very wide range of sample inside of this population who are in touch with information system all day long.
• Observers & Administrators – these users are very similar between each other, they have access to the Backend of a session and they have permission to write notes about the candidates.

So taking these two types of users we design a group of methods to better understand how can we develop and implement a solution that covers both users, for the interviewees since we had that problem we thought that better approach is to use a framework like Material Design that tries to give meaning to material as metaphor and we use several metaphors with this framework like Cards, Paper, Paper Stacks and post-it notes. To the other users we ran some semi-structured and structured interviews, to better understand what are the main issues for them, in fact we discover that Observers want a system that is flexible in the way that the notes are taken.

Then we wrote several prototypes for each user, but at the end we took in all the prototypes and mix them to became only one. For the Interviewees prototypes we use sticky notes to represent some information about the session, cards to represent the questions from the AMs. Observers & Administrators interface was highly correlated with the result of the interviews, although we still used Material Design metaphors we included text editors like WYSIWYG editor matching the flexible needs from the Observers.

Architecture
Our architecture must deal with sensitive data that is transmitted throughout the web and ensure that this data is not tampered, spoofed. The architecture model is more focused on two aspects - availability and security - the design is based on three layers of development, Application layer, Middleware layer, and Storage layer. Figure 2.

• Application Layer – Responsible to provides to users a GUI, where is possible to visualize stored data, make requests and other available services.
• Storage Layer – This model is designed to make possible the usage of different kinds of storage server, in our case we choose the MongoDB.
• Middleware Layers – This is the layer responsible for making the connection between the Application Layer and the Storage Layer, is at this level that all the services are available to the client and where is designed a general interface for the database request independently of the SGBD used.

Communication
All the communication between the clients and servers is made with HTTP requests, but to have confidentiality, integration and authentication between entities we use transport layer security (TLS/SSLv3), using upon that the server authentication (1-Way authentication). To conclude this solution we have our own Certificate Authority (CA) that sign all the server’s certificates. So for each HTTP request to the REST servers ensure that certificates are exchanged in order to validate the communications between entities. The communication between the servers and the storage servers is also with TLS and it includes client certificates to ensure that no other entity with no authorization can access and modify the storage.

IMPLEMENTATION
The implementation of the code took all the information collected and the solution designed was split in two areas, frontend implementation and backend implementation.

Frontend Implementation
For all of the interfaces and logic behind of the implantation of the frontend we followed always the same rules, like the use of Material Design in each screen that was used, like in figure 3.
All of the frontend was made with resource of HTML5, CSS3 and Javascript6, and bootstrap framework to have screens that are resizable for every screen device, we also use D3js to print the report that is the output of a session.

Backend Implementation
The backend was entirely developed using the Node.js as technology of development. In the following subsections we describe how the routing was developed, what is the algorithm behind the report generation and evaluation of scores and the technology that we used to render our interface pages.

The routing as we described before is the core of our RESTful API, and that has multiples endpoints links like the "/admin/login" or /admin/home" and in order to maintain a healthy code we follow a strict rule of everything that is supposed to render something to the user will provided by a GET method and everything that will store data in order to ensure security will be communicate through the use of a POST endpoint.

EJS is an embed JavaScript that will take our pages and populate them with data provenience from the database, we use it to render our list of cards, list of users among other things in a dynamic way.

The computation algorithm starts by iterate among each questionnaire that a interviewee answered in the session and gives a value of zero at beginning. This value is being updated depending on what:

- if the questionnaire has sub dimensions or not, if the questionnaire does not have sub dimensions the value is calculated at the dimension level
- the computation mode that a question has, i.e. if a questionnaire as a range of values [1,5] if the question is on the "normal" mode the answer "1" has a value of "1" however if the mode is "reverse" the value of "1" turns out to be "4", this is necessary because the sentence of a question can be written with the negative form.
- At the end the global score is evaluated by the weight of each dimension multiplied by the average or any other math operation.

EVALUATION
We split the system interface evaluation in two phases because we want the system to be evaluated by usability experts, that will appoint major issues of interface, that could be untraceable. With this evaluation we proceed to develop the notes token by the experts. The other phase concerns if the system is somehow alike the traditional methods of AM appliance. In this evaluation we will ask several kinds of users to do tasks and measure their time and error and in the end, we asked them which one of the methods is more practical, easy to learn, quicker among other characteristics.

Nielsen Heuristic Evaluation
This evaluation served to see major usability errors on our interface, we asked 5 usability experts to run our first phase of the prototype and evaluate it accordingly with Nielsen heuristics.

Paper vs Digital Analytic Models Appliance Comparison
In this study we evaluate our systems against the traditional method of AMs, that is the paper method. As far as we know this is the most common and consensual method of appliance. However with the development of new technologies and new methods of creating similar experiences to the real world, like Material Design tries to mimic the paper and everything that has material, we develop a system that has Material Design as backup framework and the aiming of this evaluation is to compare and validate our system against traditional methods and see if it is coherent among the studies already made.

As there are several types of user in our system we split the tasks among these types however most of tasks of the Administrator does not exist in the traditional method or not are valuable of measure & comparison. In this evaluation we will take in to account the Observers and Interviewees tasks. Observers in the traditional method write their notes in paper, some write them like it was a task, others write matrices to check some psychological dimensions. To the Interviewees we want that sample to cover all work life range, so in our evaluation we have few people that are not used to interact with technology to people that used it everyday for their work, social, gaming and others. These users in our evaluation answered to paper and in digital AMs to compare these two environments, we provide to them a validate AM in paper and a similar one in the digital method (Web App in Tablet and Computer).

We split the results in the two types of users – Interviewees and Observers. In the overall the system easier to use was the paper method, however this was not the system where the user “felt” they were more fast and dynamic since it was a split between the Tablet and Computer. The most practical
as we can see in the figure 4, were the paper and computer. The figure 3 is one of the outcomes that we manage to ask to each interviewee after the interaction with the system.

**Figure 4 - Environments Comparison in terms of quickness, practical, dynamic, fluidity**

**METHODOLOGY**

Since the beginning of the second phase of the master thesis we applied the Scrum framework as way to manage the software development. Scrum was designed to break the work in one-week to maximum four-week cycles, called” sprints”, we check the progress daily with the customers and my co-advisor Dr. Miguel Rocha Martins in 15-minute stand-up meetings, and deliver workable software at the end of every sprint. And in every Thursday, we define the next sprint tasks to be done. In the beginning of the development we made a split between what it is supposed to occur in each sprint, we define the system in three distinct phases:

- **Interviewee Interface** - this corresponds to the final user interviewee interface, tasks to do in this layer focus how to improve the user experience and was the major issue to deal with among this semester, because there was a concern that this what MAYS’ clients will see and interact to answer the AMs
- **Backend Interface** - The Administrator and Observer backend are here concerned; these tasks meet more information visualization of the data collected report generation.
- **Backend Services** - here were described tasks to store and retrieve data from the backend, the algorithm responsible for computing data collected in the system

Each time that the tasks were described, we developed and ask questions directly to the client to ensure that we were on the right track.

**FUTURE WORK**

There are a group of tasks that we decided that are less relevant given the master thesis scope but they are important for the future of this system.

**User Interface Experience**

We want to improve the experience in the future of this in particular to give a more immerse experience to the final user, we want to develop gamification techniques, so that the users does not see this only a way to answer to questionnaires, but also to feel that he is part of something important, like a journey, a quest, or even for the user feel more relax when using the platform without putting too much pressure on it. Since we are dealing with social and psychological dimensions there has to be limits for what can been done in this area without manipulating the scores of what dimensions are supposed to evaluate - reliability and validity.

**User Emotion Tracking**

For the future we want to include a way to track emotions using the device camera to associate an emotion to an answer and extrapolates some value from the person. This is a huge feature that will help us to increase the sense of creating a more similar experience with the most classic interview face-to-face.

**CONCLUSION**

Through all the project there was a couple of problems that we somehow knew that were meant to be solved likewise, when we were analyzing the user interface it was difficult for was to see because all of them apart from the OpenClinica, follow a business model of” freemium” or even premium turning their interface evaluation unlikely to be the most accurate. Other problem concerned about, how we design a solution that has the most similar experience with the paper based methods. We turn out to discover Material Design that has in fact all the theories that we want to ensure in our system, using metaphors like cards, paper to describe the interface elements. As the development went by another problem was facing us, MAYS has a client that they wanted to test the system and for that to solve the issue of any scaling problem, we had to stop awhile the development of user interface in the system. As the development went by another problem was facing us, MAYS has a client that they wanted to test the system and for that to solve the issue of any scaling problem, we had to stop awhile the development of user interface in the system. The test went well and the platform scaled and we also test our user interface in the real life, receiving from part of the client company very good feedback in the user interaction experience. In the overall this is a full stack project with very high levels of functional and non-functional requirements and we decided to develop according with these requirements and the MAYS’ needs, reaching a similar experience as it is possible to seen in the

**REFERENCES**

1. P. Kline, A Handbook of Test Construction, 1st ed. 27 Church Road, Hove BN3 2FA: Routledge, 2015.


   https://docs.openclinica.com/3.1/technical-documents.


8. SurveyMonkey, “The architecture behind surveymonkey,”


10. I. o. i. a. Amazon Web Services, “Overview of amazon web services,”

11. QuickTapSurvey, “Data encryption,”

12. ———, “How quicktapsurvey works for online and offline data collection,”


14. “Poimapper: Real-time field data,”

15. A. MMXVI, “Trust & security,”

16. O. O. S. for Clinical Research, “Rest web services,”

17. Google, “Material design,”