Virtual Belonging in Web 2.0 Projects, and its association with Online Behaviour

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2016

Hamed Mohagheghi-Fard

School of Mechanical, Aerospace and Civil Engineering

List of Contents

(Word Count: 72,400)

List of Tables

List of Figures

List of Abbreviations

AEUPV	Average Extended Unique Poem View (see section 5.2.5)
AOB	Active Online Behaviour (see section 2.2.5)
APPV	Average Poem Page View (see section 5.2.5)
APVD	Average Page View Duration (see section 5.2.3)
ATBRPV	Average Time Between Recognised Participant Visits (see section 5.2.4)
ATBRUV	Average Time Between Registered User Visits (see section 5.2.4)
AUPV	Average Unique Poem View (see section 5.2.5)
AVCL	Average Visit Compressed Length (see section 5.2.2)
AVD	Average Visit Duration (see section 5.2.2)
AVL	Average Visit Length (see section 5.2.2)
BPR	Bounce Page Rate (see section 5.2.3)
BR	Bounce Rate (see section 5.2.3)
EL	Event Log (see section 5.1.1)
EP	Exit Page (see section 5.2.3)
EPR	Exit Page Rate (see section 5.2.3)
EUPV	Extended Unique Poem View (see section 5.2.5)
IC	Internet Cookie (see section 4.6.1)
LP	Landing Page (see section 5.2.3)
LPR	Landing Page Rate (see section 5.2.3)
NIV	Number of Incoming Visits (see section 5.3.4)
NOV	Number of Outgoing Visits (see section 5.3.4)
NPPV	Number of Poem Page Views (see section 5.2.5)
NPV	Number of Page Views (see section 5.2.1)
NRP	Number of Recognised Participants (see section 5.2.1)
NSV	Number of Single Visits (see section 5.2.1)
OB	Online Behaviour (see section 2.2.1)
OC	Online Community (see section 2.1.5)
POB	Passive Online Behaviour (see section 2.2.3)
PPV	Poem Page View (see section 5.2.5)
PVD	Page View Duration (see section 5.2.3)
RIR	Rate of Identified Returns (see section 5.2.4)
RIV	Rate of Incoming Visits (see section 5.3.4)
ROV	Rate of Outgoing Visits (see section 5.3.4)
RP	Recognised Participant (see section 5.1.3)
RPVC	Recognised Participant Visit Count (see section 5.2.4)
RU	Registered User (see section 5.1.4)
RUVC	Registered User Visit Count (see section 5.2.4)
SB	Sense of Belonging (see section 2.4.2)
SNS	Social Networking Site (see section 2.1.6)
SV	Single Visit (see section 5.1.2)
UPV	Unique Poem View (see section 5.2.5)
VB	Virtual Belonging (see section 2.4.6)
VCL	Visit Compressed Length (see section 5.2.2)
VD	Visit Duration (see section 5.2.2)
VL	Visit Length (see section 5.2.2)
VP	Visit Path (see section 5.2.3)
VPR	Visit Path Rate (see section 5.2.3)

Abstract

The behavioural changes which result from expansion and popularity of online platforms can be observed in day to day activities such as buying tickets, ordering food, and communicating with friends and family. From a technical perspective these platforms utilise similar technologies and infrastructures (e.g. the Internet, servers, and databases), however there are numerous conceptual and functional differences which can be identified between them. This research is associated with specific types of online platforms which are commonly referred to as Web 2.0 platforms. Popular examples of such platforms include Wikipedia, YouTube, and Facebook. These platforms rely upon user generated content and benefit from users' collaborative efforts.

Users' efforts can be considered as a resource, and a possible objective in associated projects could be to optimise the outcome of such efforts. In many cases, this can be accomplished by influencing users' Online Behaviour (OB). In Web 2.0 platforms, OB can be influenced by various factors, and among them Sense of Belonging (SB) has been highlighted by many previous researchers as a significant factor. While the literature suggests a significant association between SB and OB in Web 2.0 platforms, this research identifies a gap in the existing literature in regards to SB due to its qualitative nature and the dependency of associated investigations on human interaction. In addition to the subjective nature of such investigations, other practical challenges can be associated with measuring SB in Web 2.0 projects. Firstly, considering the demographical, cultural and lingual diversity of users in many Web 2.0 platforms, such investigations can be complex and therefore expensive. Secondly, investigating SB cannot be performed in real time which can result in extended timelines for gathering data. Thirdly, in early stages of Web 2.0 projects in which the concept is being defined and the platform is being developed, the users might not yet exist, and hence investigating their SB and OB may not be an option. Finally, without an existing platform and community, there is nothing to belong to, and hence SB might not exist to measure or investigate.

As an attempt to fill the identified gap and provide practical solutions for highlighted challenges in Web 2.0 projects, this research introduces a new factor, and investigates its association with OB. Virtual Belonging (VB) is proposed as a factor built upon suggested elements for SB, and based on a new perspective towards Web 2.0 features. Furthermore, in any Web 2.0 platform, VB can be quantitatively defined, compared, and analysed. This research suggests three dimensions for VB, and their associations with OB are hypothesised. To test the research hypotheses, a Web 2.0 platform was developed in association with a student poetry competition in the University of Manchester, and it was used as the vehicle for an experiment (available at epsPoetry.com). The platform includes a bespoke Tracking System that has been specifically designed and developed to assist with this investigation. Also, Ethical Approval was obtained from the University's Ethics Committees for the use of the data collected by the Tracking System from the individuals who have agreed to take part in this research and used the platform between March 2013 and July 2014.

VB's dimensions are presented in this research as a hierarchy with three distinct levels, namely Inclusion, Involvement, and Influence. The results show a significant association between every dimension of users' VB and their OB. On this basis, this research proposes VB as a predictive factor for OB in Web 2.0 projects. The findings can provide practical guidance and directions for Web 2.0 project managers, particularly in the early stages of such projects in which the existence of users and their interactivity with the platform can only be an assumption, and hence obtaining qualitative data (such as SB) may not be an option.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Dedication

This thesis is dedicated to my parents, who have given me the confidence to aim high, and to Maryam, whose smile makes possible the impossibles.

Acknowledgement

I would like to thank Dr Peter Fenn for being an inspiring mentor and a great supervisor. He gave me the freedom to explore, and the guidance to navigate. He also supported me in my personal life in addition to my PhD. Without his encouragement and support, this would not have happened. I would also like to thank my co-supervisor Dr Oscar de Bruijn for his guidance, expertise and support.

There are many people without whom this thesis might not have been written, but there are a few people in one's life who make success both possible and rewarding. I would like to thank my parents, Hassan and Shohreh, who have supported me all my life and have always been encouraging and inspirational. I also would like to thank my sister, Dina, for her worthy advices, and my brother and best friend, Navid, for teaching me the "Hello World" of PHP, and my niece, Ava, for inspiring me to think about the future generations, and the effects that technology may have on society and human race. Last but not least, I would like to thank Maryam, whose kindness, help and support during my research has been incredible.

1 Introduction

This chapter gives an overview of this research programme. Initially a background to the study is provided, and this continues by outlining the primary aim and objectives of the study, and highlighting the nature of the work undertaken and presented in the following chapters of this thesis. Finally, the structure of this thesis and an overview of the following chapters are described.

1.1 Research Background

The behavioural changes which result from expansion and popularity of online platforms can be observed in day to day activities such as buying tickets, ordering food, and communicating with friends and family (Ayeh et al., 2013; Filieri, 2015; He et al., 2013; Church and de Oliveira, 2013; Kim et al., 2007; Mellet et al., 2014). One may argue their influence is beyond that, and such platforms have influenced subjective norms in regards to more fundamental aspects of life such as buying a house, or even finding someone to love (Pizzato et al., 2013; Beracha and Wintoki, 2013; Finkel et al., 2012; Guadagno et al., 2012; Blackwell et al., 2014).

Some of these newly emerged platforms have also had a significant effect on many traditional businesses. For example, for more than a century "Michelin Guide Book" had been *the* guide for restaurants and hotels. However, its role is being somehow replaced by number of stars achieved on TripAdvisor or Google Maps (Michelin, 2012; Mellet et al., 2014; Duffy, 2012; Conrady, 2007). Similarly, television advertising is threatened by online advertising, and traditional shopping channels are losing their market share to their online rivals (Weltevreden, 2007; Limayem et al., 2000; Schultz and Block, 2015; Porter and Golan, 2006). While the emergence of these new technologies can be seen as a threat to some traditional business models, it can also open up new opportunities.

The Internet has been around since 1983, and from early stages the infrastructure required for user interaction has been available (see section 2.1.2). However, it took around 15 years for the viral adaptation of the associated technologies by non-technical individuals (Kaplan and Haenlein, 2010). The birth of Web 2.0 (or interactive web) can be dated back to early blogging communities. For example, in late 1990s a group of non-technical individuals started to use "Open Diary", and publish text-based content online. They voluntarily kept their blogs up to date without involvement of any editorial

or hierarchical decisions (Open Diary, n.d.). However, the division of web-based platforms into Web 1.0 and Web 2.0 was first popularised in 2004 by Tim O'Reilly (O'Reilly, 2007). He identified the adaptation of user interactivity among some of the websites. This interactivity was achieved by introducing features such as blogging, commenting, voting or allowing users to upload materials such as photos and videos (O'Reilly, 2007).

Web 2.0 platforms have been studied by many previous researchers, and from different perspectives such as education, health, and business (e.g. Greenhow et al., 2009; McLoughlin and Lee, 2007; Clara and Barbera, 2013; Koh and Kim, 2004; Maloney-Krichmar and Preece, 2002; Kim et al., 2008). As a complex phenomenon, these newly emerged technologies can be looked at from different angles, and consequently there are disagreements and sometimes opposition among scholars in regards to terminologies and proposed definitions. However, the fact that new innovative products and platforms are introduced and adopted by society, makes some of the exiting terminologies and definitions somehow out of date. For example, in 2007 the original *iPhone* was introduced (BBC, 2007), and since then the introduction of smart phones and wearable technologies have brought about new ways of interaction for the Internet users (Aldhaban, 2012). However, this has added a new layer of complexity into previously suggested terminologies. In this study any project which involves any form of online platforms, and enables users to interact with the platform has been considered as a Web 2.0 Project (see section 2.1.7).

In many Web 2.0 platforms (e.g. Wikipedia, YouTube, Facebook) the content is mainly generated by users, and this content is a fundamental part of the platform. Furthermore, many of such platforms are dependent on other forms of interactivity by their users (e.g. rating the content, sharing the content, reporting inappropriate content). Consequently, the outcome of associated projects could be dependent on users' interactions with associated platforms. Some previous researchers have categorised users interactions with Web 2.0 platforms to *active* and *passive* (Nonnecke and Preece, 2000; Koh et al., 2007; Benkler, 2002). On this basis, users are categorised as passive and active participants.

Passive Participants are the individuals who use Web 2.0 platforms, and although they are presented with functionalities which enable them to actively interact with the platform, they choose not to do so. On the other hand, *Active Participants* are the users who adopt the available features, and actively interact with the platform. For example,

Wikipedia is among the most visited online platforms, and it is reported that in February 2016 its web pages were viewed more than 16 billion times in total, and it is estimated to have had around 400 million unique visitors during this time (Alexa, 2016; Wikimedia Report Card, 2016). However, there are only around 130,000 users who have edited an article during this time (Wikipedia, 2016b). In other words, a majority of Wikipedia users are passive participants, and only view the content without contributing to it. While the ratio of active and passive participants can vary in different Web 2.0 platforms, van Mierlo (2014) suggests in many of such platforms around 1% of users create the vast majority of new content. However, in addition to *creating* content, there are other types of active participation (e.g. sharing or evaluating content) which can be important in Web 2.0 projects.

There are various types of active participations which can be observed in Web 2.0 platforms. Some of the popular ones are posting content (in the form of text or multimedia), commenting on available content, and evaluating the content by giving quantitative feedback. It should be noted that each platform can have their own specific representations and terminologies for similar features and functionalities which enable active participation (e.g. rate, rank, vote, like, love, plus one, thumbs up, etc.). Furthermore, there is another form of behaviour which can be regarded as active participation. This form of participation involves *sharing* the available content, and it can be seen as free advertising for many platforms (John, 2012). It should be noted that sharing content is usually executed by sending content (or Links to content) from one platform to another platform. Such behaviour can have an important impact on the popularity of each platform, and has been presented in this research among some others (e.g. Benkler, 2002; John, 2012) as active participation.

Active participation can be seen as a fundamental and important factor in many Web 2.0 platforms (Tedjamulia et al., 2005; Hsu and Lin, 2008). In some cases, the platforms' concept is so fundamentally associated with active participation, that without it any other success factor will not be achieved. For example, popular platforms such as Wikipedia, YouTube or Facebook are so dependent on user generated content that such platforms would not have existed without it. In many other cases, the success of the project is not solely but still largely depends on active participation from its users (Tedjamulia et al., 2005). Many previous studies (see section 2.2.5) have investigated active participation in a variety of Web 2.0 settings, and the outcome of these studies is reviewed in Chapter 2.

The reviewed literature suggests an important role for active participation in Web 2.0 platforms (see section 2.2.5), yet the relationship between users' active participation and their utilisation of provided Web 2.0 features has not been focused on. Such an understanding can present a new perspective towards why a minority of users are found in so many online settings who devote their time and resources towards freely maintaining and improving the platform. This can also provide practical guidance for choosing the appropriate Web 2.0 features in each platform. Furthermore, while investigating the Influencing Factors for users' Online Behaviour, many previous studies focus on qualitative and psychological measures such as trust, feeling of being included, and Sense of Belonging (see section 2.3). While these factors can be investigated and analysed in the academic literature, from a practical perspective such investigations can be subjective, complex and relatively expensive (see section 2.4.5). Furthermore, the dependency of such studies on *human interaction* can become challenging in early stages of developing Web 2.0 platforms in which the concept is shaping and the existence of users is only an assumption (see section 2.4.5). Therefore, this research takes a different approach and aims to provide an alternative method which is *not* dependent on human interaction and can be used in *any* Web 2.0 platform for discriminating, comparing and analysing users in real time.

1.2 Aim and Objectives

The aim of this research is to investigate Online Behaviour (OB) in Web 2.0 platforms, and to present a quantitative model in association with OB which can be used to discriminate and compare users in any Web 2.0 platform without human interaction. The following objectives were identified to achieve this aim:

- To improve understanding of OB (particularly active and passive participation in Web 2.0 platforms) by reviewing the associated literature
- To improve understanding of influencing factors for OB (particularly factors which are associated with active participation in Web 2.0 platforms) by reviewing the associated literature
- To revisit *behaviourism* (after a century from its introduction) by proposing and applying a quantitative method for investigating human behaviour in an online *experiment* based on a behaviourist approach

- To propose and test a quantitative model which can be used to discriminate and compare users in any Web 2.0 platform by investigating and categorising suggested Web 2.0 features in the literature

- To suggest *practical guidelines* for Web 2.0 managers (particularly in the early stages of development in which feedback from users might not be an option) by proposing a quantitative model in association with OB which can be used in any Web 2.0 platform

As part of this study a Web 2.0 platform (available at: *epsPoetry.com*) was developed, and those users who agreed to take part in this research had their activities tracked and investigated. The platform enabled students within the faculty of Engineering and Physical Sciences (EPS) of the University of Manchester to submit their poems, and take part in a poetry competition (see chapter 4). The generated content (e.g. accepted poems and comments) was publicly available on the Internet, and any other Internet user could access them. In addition to functionalities which enabled students to submit their poems, other Web 2.0 features (e.g. registration, voting, commenting and sharing) were included in the platform, and students were able to become active participants by using such features.

1.3 Thesis Structure

This thesis is presented in 7 chapters, and a summary of what is included in the following chapters is followed. It should be noted that while this study has a fundamental connection to technicalities of Web 2.0 Platforms, technical jargon has been avoided unless it is necessary.

Chapter 2 reviews the exiting scholarly literature about Web 2.0 Platforms (see section 2.1), Online Behaviour (see section 2.2), and some of the Influencing Factors for users' behaviour (see section 2.3). From the suggested influencing factors, it then focuses on Sense of Belonging which has been suggested by numerous studies as a significant factor for influencing different aspects of Online Behaviour (see section 2.4.4). Finally, a gap in the literature is highlighted (see section 2.4.5), and on this basis Virtual Belonging is defined (see section 2.4.6) in the context of Web 2.0 Platforms. In Chapter 3, initially the research paradigms are discussed and compared. Then the major criticisms toward the chosen paradigm are described and their implications on this research are considered (see section 3.1). This is followed by a brief discussion about

the psychological approaches towards investigation of human behaviour, and continues with further investigation of behaviourism as the chosen approach (see sections 3.2 and 3.3). The research design is then described (see section 3.7), and research hypotheses are outlined (see section 3.7). In chapter 4, experiment design is described. This includes technical and non-technical steps which have been taken to run the experiment. This is followed by describing the data collection procedure (see section 4.12) and data preparations procedures (see section 4.14) which are used during the experiment. The collected data has been statistically analysed, and the results are presented in chapter 5. This includes the process for quantifying Online Behaviour (see sections 5.2 and 5.3), and describing the statistical methods which are used in this research (see section 5.4). This will be followed by investigating research hypotheses in the experiment (see sections 5.5, 5.6, 5.7, and 5.8). In Chapter 6, the findings are explained and their implications are discussed together with the limitations of this research. Finally, in chapter 7, a summary of this research is presented and recommendations are made for Web 2.0 managers and future researchers.

2 Literature Review

This chapter reviews some of the exiting literature about Web 2.0 Platforms, Online Behaviour (OB), and influencing factors for OB. From the suggested influencing factors, it focuses on *Sense of Belonging* (SB) which has been identified by numerous studies as a significant factor for influencing OB. Finally, a gap in the literature is highlighted, and to fill the identified gap *Virtual Belonging* (VB) in Web 2.0 platforms is defined as a quantitative factor.

Initially, this chapter reviews the exiting literature to paint a picture of what can be regarded as a typical Web 2.0 Platform, and how it can be separated from Non-Web 2.0 Platforms. To do so, it starts by looking at the history of Web 2.0, and delves into the confused and sometimes contradictory terminologies and definitions which are available in the existing literature (see section 2.1). It continues by reviewing the categorisations which have been suggested in the literature, and thereafter Online Communities (OCs) and Social Networking Sites (SNSs) are looked at more specifically as the current most popular categories of Web 2.0 platforms (see sections 2.1.5 and 2.1.6). For these specific types of Web 2.0 platforms, some of the suggested definitions and associated literature are reviewed. Afterwards, Web 2.0 Projects are defined based on the reviewed literature, and the inclusion criteria are described for what exactly is regarded as a Web 2.0 Project in this study (see section 2.1.7).

In this chapter, OB is explained as part of human behaviour, and the relevant literature is critically reviewed in regards to what can be considered as influencing the behaviour, and more specifically influencing OB in Web 2.0 platforms (see section 2.3). Afterwards, the two main dimensions which have been suggested for OB (*active* and *passive*) are defined and described (see sections 2.2.3 and 2.2.5). The chapter continues by reviewing some of the major factors which have been suggested in the literature to have an influence on OB, and more specifically influencing factors for Active Online Behaviour (AOB) are investigated (see section 2.3).

After a general investigation of influencing factors for OB, *Sense of Belonging* (SB) is highlighted and investigated further (see section 2.4.4). SB has been suggested in the literature as a significant factor for influencing both dimensions of OB in many Web 2.0 platforms, in particular Online Communities. To achieve a better understanding of SB, the existing literature has been explored in association with *Belonging* both outside and

inside the online environments, and some of the relevant researches are described. Then within the context of Web 2.0, cause and effects of SB are described, and some of the existing instruments for measuring SB are reviewed (see section 2.4). Finally, this chapter highlights a gap in the literature (see section 2.4.5), and as an attempt to fill this gap it defines and describes *Virtual Belonging* (VB) as a quantitative factor in Web 2.0 platforms (see section 2.4.6). After defining VB, three dimensions are described for it, namely, *Inclusion*, *Involvement*, and *Influence*.

2.1 Web 2.0 Projects

The Web as we know it today is a relatively new concept and it was not invented until 1989 (W3, 2015a), Furthermore, the earliest machines which made possible such a revolutionary phenomenon were not built until the late 1940s (ENIAC, 1961). At that time, it would have been rather difficult to predict the influence such technologies would have on society and human race. Similarly, one may argue the future of these technologies, and their long term impact on society and human race can be as unpredictable today. With that in mind and also considering the possibility of future changes, later in this chapter (see section 2.1.7) an exact definition for Web 2.0 Projects will be proposed, and it will be used throughout this thesis.

2.1.1 Primary Definitions and Terminology

In the next sections the history of Web 2.0 will be described, and some of the popular examples of Web 2.0 platforms are investigated. Technical jargon has been avoided unless unavoidable. However, some of the primary terms which are used within the next sections are described in Table 1 without considering the associated complexities.

Term	Description
Computer Software	The programs and other operating information used by a computer (Oxford Dictionary)
Computer Hardware	Tools, machinery, and other durable equipment (Oxford Dictionary)
Computer Network	Telecommunications network which allows computers to exchange data (Kershenbaum, 1993).
Internet Cookie (IC)	Text-based data which is received from the server and stored in the users' web browsers (see section 4.6.1)
IP Address	Numerical identifier which is assigned to each device within a computer network that uses the *Internet Protocol* for communication (see section 2.1.2)

Table 1 - Primary Definitions

2.1.2 Background and History of Web 2.0

ENIAC (Electronic Numerical Integrator And Computer) was built in 1946, and it was one of the first programmable digital devices with high computational power. It was

mainly used by the United States Army, and it has been considered as the starting point for computers as we know them today (McCartney, 1999; ENIAC, 1961). Since then these machines have become more and more powerful in terms of computational power and data storage capabilities. On the other hand, during 1960s data transmission technologies started to provide a platform for these machines to digitally communicate with each other (Ryan, 2010). As a result computer networks were born and a new era for digital communication was started. These networks only allowed people within the same network to communicate with each other, and with increasing popularity of local networks the idea of a network of networks was considered (Ryan, 2010). As a result different standards and protocols were implemented and tested, and plans were made for a global switch over. Finally, in 1983 the current protocol of the Internet (TCP/IP: Transmission Control Protocol/Internet Protocol) was adopted as the standard by the Advanced Research Projects Agency Network (ARPANET) to mark the beginning of the Internet era (Network Working Group, 1981). However, the astonishing popularity of the Internet started in 1990 by creation of *the Web*.

The concept of *HyperText* was originally suggested by Ted Nelson around 1965 for a text which contains *Links* to other texts. He later expanded the concept by removing the emphasis on text and proposing the concept of *HyperMedia* which could include linked graphics, videos and sounds (W3, 2015b). Nowadays, the terms HyperText and HyperMedia are less common, but the same concept exists, and it is achieved by using *HyperLinks* or simply *Links* between the web pages and their containing elements such as text and graphics. Tim Berners-Lee, a British scientist at CERN, used this concept, and invented the *WorldWideWeb* (WWW) in 1989 (W3, 2015a). In November of the following year, Berners-Lee and Cailliau proposed to build a HyperText Project called WorldWideWeb as a web of HyperText documents, in which the HyperText allowed documents to be connected to each other. They requested 4 software engineers and one programmer, and they estimated that with this manpower the project would need 6 months for completion. According to their proposal, the documents would be viewed by browsers using *client-server* architecture. They also suggested that in 6 months time the users would be able to collaborate in document creation (W3, 1990).

In the WorldWideWeb *the Web* is created by *Links*, and on this basis the position of each website in the overall structure of the Web can be determined based on the Links which connects it to other websites. This can be an important factor when websites are compared with each other, and especially such comparisons are used by Search Engines

to sort the presented results (Page et al., 1999). The study of quantitative aspects of the web structure sometimes is referred to as *Webometrics*, and various measures have been suggested for it in the literature. For example, Ingwersen (1998) introduced the concept of the *Web Impact Factor* (WIF) for measuring online impact of a set of web pages. It is calculated based on the number of links to a set of web pages divided by the number of pages in the set (Ingwersen, 1998). Later research by Li (2003) shows that WIF correlate significantly with events outside the Web. Considering WIF is based on the received Links, this suggests that receiving Links from other websites highlights the credibility of the source. The number of received Links is one of the factors which are used by Search Engines to order their search results, however its importance has been shown in other fields. For example, Vaughan and Thelwall (2003) investigate scholarly use of the Web, and suggest that higher impact journals receive more Links to their website. Another research by Vaughan and Wu (2004) investigates commercial websites, and suggest the number of received Links is correlated with the companies' revenue and profit.

In 1996, Larry Page and Sergey Brin from Stanford University invented *PageRank*, as a quantitative measure of importance and reliability of the available content on the Internet. The Links between web pages were one of the main factors in their model, and they later used PageRank to sort the results in their Search Engine which was named *Google* (Page et al., 1999). PageRank is still one of the factors which is used by Google to sort the search results, however the algorithms which are used can evolve during time, and due to their sensitivity the exact measures are usually do not exist in the public domain (Langville and Meyer, 2011; Google, 2016).

While from the very early stages of the web, users could have been involved with content generation (W3, 1990), for years the content was mainly provided by websites' owners or managers. However, Web 2.0 technologies changed this forever. The web-based user generated content first became popular in its current form around 8 years after invention of the Web, and when the first blogging community was established (Kaplan and Haenlein, 2010). It is important to note that Web 2.0 is not anything conceptually different than WorldWideWeb as it was suggested originally (W3, 1990), and one can still look at all the content which is available online, and see them as documents which are connected to each other by Links. What distinguishes Web 2.0 is the way this content is *generated*, *sorted* and *distributed*. The phrase "Web 2.0" was first used by Darcy DiNucci in 1999, and years later it was highlighted and popularised

by Tim O'Reilly and Dale Dougherty in a brainstorming session. Their debate progressed to the birth of *Web 2.0 Conference* in 2004 (O'Reilly, 2007).

2.1.3 Definition of Web 2.0

O'Reilly (2007) describes Web 2.0 by looking at the differences in web services and categorising them as *Web 1.0* and *Web 2.0*. He describes Web 1.0 as software with regular updates. He argues that during its early years, the Web consisted of simple web pages with text, images and links between them. These documents could only be updated from the background and by managers of each website. The users could spend hours browsing different pages, but they could not interact with them. O'Reilly (2007) explains that Web 2.0 services started by using the Web as a platform. The users simply access the service through a website using a web browser, and any update to the platform happens without the users' involvement. Furthermore, Web 2.0 technologies are interactive and allow their users to influence the available content, and their inputs can usually be viewed by other users (O'Reilly, 2007).

Kaplan and Haenlein (2010) define Web 2.0 as a new way of using the Web, in which the content is continuously modified by all users in a participatory and collaborative fashion. From information transmission perspective, Web 2.0 provides a platform for participatory information transmission (Gu and Widén-Wulff, 2011). Although at the core of both Web 1.0 and Web 2.0 is the same technology, Chiang et al. (2010) suggest Web 2.0 represents a *paradigm shift* in design and use of websites. Different characteristics have been identified for Web 2.0 technologies among them: user-centred, participatory and interactive (Anderson, 2007; Stephens, 2007; Holmberg et al., 2009). Web 2.0 is also *open* in a sense that users could join an extensive variety of online platforms and publish their thoughts and ideas (Gu and Widén-Wulff, 2011).

Web 2.0 technologies can also be seen as communication tools. Birdsall (2007) explains that the available Web 2.0 technologies have made *easy communication* possible. It should be noted that despite the complex nature of the technologies behind Web 2.0 platforms, a majority of them are relatively easy to use, and usually their users are not officially trained to use these technologies. Users are usually able to identify and use common Web 2.0 features such as commenting, voting and sharing. Also, it should be noted that using Web 2.0 platforms can usually be achieved with almost no knowledge of what is going behind the scene (e.g. databases, internet protocols, data transmission, transmission security, etc.).

The Users are at the centre of Web 2.0 and *participation* is the main change in the context of Web 2.0 (Miller, 2005), and if they wish individuals are able to contribute to the available information. One can argue this is against some of the principles which have existed in publishing industry for decades. For example, The Michelin Guide is a series of annual guide books published by Michelin since 1900. The Michelin Red Guide is a hotel and restaurant guide, which awards *Michelin Stars*. Michelin also publishes Green Guides for travel and tourism (Michelin, 2012). In 2001 Michelin launched its main website, but despite having three million registered users in 2008, it was facing serious competition from newly emerged Web 2.0 competitors such as Google Maps and TripAdvisor. Therefore the site started to use Web 2.0 features to engage users in providing the content and building a virtual community (Schenker, 2008). Considering similar examples, one may argue these days public opinion is regarded as more credible than experts' judgement. In 2006, Time Magazine recognised this revolutionary phenomenon by choosing "You" (ordinary web users) as the magazine's *Person of the Year*. They explained: "... for seizing the reins of the global media, for founding and framing the new digital democracy, for working for nothing and beating the pros at their own game, TIME's Person of the Year for 2006 is you" (Time, 2006).

2.1.4 Examples of Web 2.0 Projects

Since their introduction, online platforms have changed enormously as a result of improvement in underlying technologies and more importantly as a response to wide spread acceptance in the society (see section 2.1.2). While based on terminology *Web 2.0* platforms seem to be dependent on *the Web* and subsequently the Internet, there are examples of platforms which existed prior to the Internet and followed the same principles as current Web 2.0 platforms. The earliest examples of such networks were created in late 1970s. For example, *Usenet* was created in 1979 by Tom Truscott and Jim Ellis from Duke University, and it was a worldwide discussion system that allowed users to post public messages (Emerson, 1983; Kaplan and Haenlein, 2010). While Usenet might not be considered as a Web 2.0 platform using today's standards and definitions, similar features and principles can be identified. This innovation was followed by other forms of digital communities including commercial online services, email lists, Bulletin Board Systems (BBS), and Internet forums (Kaplan and Haenlein, 2010). However, these communication tools did not become popular among ordinary people for at least two decades. Bruce and Susan Abelson founded "Open Diary" in

October 1998 (Open Diary, n.d.). It was the first web-based community for online diary writers, and the term "weblog" was first used at the same time. A year later, the term "blog" was coined after one blogger transformed the noun "weblog" into "we blog" (Kaplan and Haenlein, 2010). The growing availability of internet access, and user-friendly platforms such as "Blogger.com", added to the popularity of blogs and more individuals became bloggers. After the Internet *bubble* of 2000, once again the Web was becoming popular. By introduction of *MySpace* in 2003, the music industry and its celebrities were among the first to find a new communication channel. The unexpected success of MySpace was shortly followed by other successful Web 2.0 platforms such as *Facebook* which started in 2004 (Kaplan and Haenlein, 2010).

The evolution of Web 2.0 technologies has also been influenced by available technologies. For example, the original iPhone was introduced by Apple in 2007, and its success started a new era for smart phones (BBC, 2007; Chang et al., 2009). As a result of this innovation a new platform was created and the Apps were introduced for mobile phones. Consequently, many successful Web 2.0 platforms (e.g. Facebook and YouTube) created their own Apps. Furthermore, there have been examples of successful Apps such as Instagram and Tinder which achieved their success only as an App, and without using the traditional websites for online interaction. Hence, any definition should take into account similar changes which can happen in the future (see section 2.1.7).

It has been suggested that currently there are around one billion websites on the Internet (Berners-Lee, 2014; InternetLiveStats, 2016). It should be noted that the fact that these websites exist does not necessarily mean that they are active, and some of them might have very few visitors or no visitors at all. Websites' traffic is defined, measured and compared based on the number of individuals who visit each website during a fixed amount of time (see section 5.2.1). "Alexa Internet, Inc." is one of the credible sources of independent data about the websites' traffic (Gomadam et al., 2008). In February 2016, they report the following websites as the most popular in the UK in regards to their traffic: *Google*.co.uk, *Facebook*.com, *YouTube*.com, *Google*.com, *Amazon*.co.uk, *BBC*.co.uk, *eBay*.co.uk, *Yahoo*.com, *Live*.com, *Wikipedia*.org, and *Twitter*.com (Alexa, 2016). One can argue that Web 2.0 features are available in all of these websites. However, some of them (such as Google, Yahoo and BBC) are not solely dependent on user generated content. Furthermore, these websites existed before popularity of Web 2.0 technologies. On the other hand, websites such as Facebook, Wikipedia, and eBay

cannot be imagined without participation from their users. There are a variety of definitions and categorisations that have been suggested in the literarure for these user-centred websites. Some specific types of these websites have been defined and categorised in current literature as Online Communities (such as YouTube, eBay and Wikipedia) and Social Networking Sites (such as Facebook and Twitter).

It is natural to expect disagreement between academics, however one can argue that the fast pace of change in Web 2.0 technologies has aggravated the associated disagreements and contradictions. Furthermore, some of the scholars have updated their definitions in later publications, and in some cases have contradicted their own previous claims or definitions. With this in mind, initially Online Communities and Social Networking Sites (as the current most popular Web 2.0 platforms) are reviewed in the existing literature, and later in this chapter, Web 2.0 Projects are defined specifically for this study (see section 2.1.7).

2.1.5 Online Communities

As explained earlier (see section 2.1.4), Online Communities (OCs) are currently among the most popular forms of Web 2.0 platforms. However, there is a range of definitions which are available in the literature to describe this phenomenon, and different attributes have been suggested for such communities. Many scholars agree that computer mediated *communication* has a major role in defining OCs. However, the opinion about its dependency on the Internet is divided. Rheingold (2000) describes OCs as a group of people who exchange ideas and words through computer mediated tools. Similarly, Cothrel and Williams (1999) describe OCs as a group of people who use a computer network to interact with each other. Bishop (2007) adds the concept of *collaboration* to communication, and describes OCs as groups of people who collaborate through networked technologies. He also highlights the fact that by using such technologies differences in time zone and location can be disregarded as a barrier. As explained earlier (see section 2.1.2), this kind of communication has been available prior to existence of the Internet. However, some of the recent definitions for OCs are dependent on the Internet as the medium for communication. For example, Lin (2008) defines an OC as a cyberspace with internet-based chat technologies. He argues such communities depend on social interaction among their members, and the members of such communities use the provided platform to share interests and build relationships. Lee et al. (2003) explain that such communities facilitate informal sharing of *knowledge*. They suggest this knowledge is usually provided by more experienced and

skilled members of the community. The available functionalities in an OC enable the users to use the available information, and contribute to the existing knowledge if they so wish (Lee et al., 2003).

Although OCs have been mainly used to facilitate knowledge sharing, Preece (1998) suggests that such communities can also help their members emotionally. For example, an individual with an illness can use related communities, and in addition to receiving advice from the experienced members, they can also receive emotional support from other community members.

Different attributes have been suggested for OCs. Damsgaard (2002) identifies several attributes for such communities, including shared goals, interaction and ties among members, and a shared convention. Such *conventions* usually exist in any online environment, and in OCs the rules are often enforced by a group of community leaders. While the members are at the centre of OCs, Kim (2000) separates community members and community leaders, and highlights the distinctive role each have in the community. Balasubramanian and Mahajan (2001) define OCs based on people, electronic medium, interaction, and common interest. Alternatively, Lin and Lee (2006) suggest people and technology as the main attributes for OCs. Many researchers have included common purpose or shared interest as one of the attributes of OCs (Ahuja and Galvin, 2003; Porter, 2004; Lin, 2007). Although these communities are usually shaped around a common interest (e.g. computer programming, cooking, or democracy), the amount and diversity of available and produced information within the community can vary.

Some researchers have added more conditions and attributes for what they classify as an OC. For example, Jones (1997) identifies *sustained* membership as a fundamental attribute for an OC. Similarly, Butler (2001) highlights the fundamental role of *ongoing* interactivity in such communities. Ridings et al. (2002) also argue that the existence of online communication is not enough and the group members must communicate *regularly* and for some duration. Despite the difference between the content provided in each of these communities, Lin (2008) also suggests that they all must have some degree of stability and *growth*. It should be noted that while communication and interaction is usually achieved by utilising an electronic medium, members of such communities can also use other forms of communication to interact with each other (e.g. offline events and face to face conversations). Preece et al. (2003) argue that OCs can extend beyond the online environment. They suggest different attributes for such

communities, such as physical and software environments, common purpose, members, and culture.

One can argue the user generated content can be considered as a commodity for the owners of OCs. However, while active participation from users and higher volume of user generated content can be desired in many cases, too much content can produce information *overload*, and this has been identified as an issue for OCs (Barua et al., 1995; Finholt and Sproull, 1990). Although the available content is an important part of such communities, too much information can make it difficult to find the required information. To avoid this issue, the information provided in such communities should be organised and searchable (Finholt and Sproull, 1990). There are common features within such communities to help their member in finding relevant information. Usually this is achieved by providing search facilities and organising the content by threading, tagging, hyper-linking or other similar technologies (Lin, 2008).

The user generated content is usually stored in online databases, and there is a set of specific rules within each OC as to who can view or change the available content. In other words, for each user there is a level of accessibility defined for each piece of information (or Content) and any available functionality (such as Add, Edit or Delete). Information accessibility has been suggested to have an important role in OCs (Teo et al., 2003). Each OC is governed by its own set of rules or conventions. These rules are interpreted to algorithms which are embedded within the underlying software for each OC. If users choose to use an OC, they are bound by the existing rules and logics. For example, some websites only provide access to information (or port of the available information) to their registered members. Although *registration* is free in many OCs, it can be considered as a barrier (Teo et al., 2003). This can result from the fact that some of the users can be concerned about their privacy, or they might not have the required time or self confidence to go through with the registration process. While usually such access restrictions are used to increase membership, Teo et al. (2003) suggest such barriers can reduce user interaction and consequently the website's traffic. In some cases this barrier has been removed or the rules have been relaxed. For example, Wikipedia project is one the cases which have identified registration as a barrier. Subsequently, many of the articles in Wikipedia can be edited by the users even without registration (Wikipedia, 2016a). In this case, their IP Address would be logged and displayed in the change history of the edited article. However, to avoid vandalism and

negative impact from inexperienced users, registration is required for editing important, sensitive or controversial articles (Wikipedia, 2016a).

Following the growing popularity of OCs, business organisations have been trying to use these communities to achieve organisational goals, and in many cases they actively build and manage such communities to provide and share information, and enhance customer support (Teo et al., 2003). Furthermore, in many instances these communities have been used as a new tool to improve marketing (Teo et al., 2003). Many manufacturers provide such communities, and by using these platforms technicians collaborate with community members to help other users. It has been suggested by previous researchers that sustaining an associated OC can help businesses to promote knowledge sharing, and this can have a significant impact on their business (Igbaria et al., 1998; Teo et al., 2003).

2.1.6 Social Networking Sites

As explained earlier (see section 2.1.4), among the most popular websites in regards to global and national traffic, specific types of Web 2.0 platforms can be identified. Social Networking Sites (SNSs) are one of these types, and in regards to number of active users around the world, currently some of these networks are among the most popular Web 2.0 platforms (Duggan et al., 2015; Alexa, 2016). One may compare their effect on society and human race with revolutionary innovations such as printing press, telephone, and light bulb. Platforms such as Facebook, Twitter and LinkedIn have used Web 2.0 technologies, and have provided a new channel for communication, interaction, and collaboration. Many individuals all around the world use these platforms, and many of them use their services on a daily basis. In this section some of the definitions and attributes which have been suggested in the literature for this type of Web 2.0 platforms are reviewed. This is followed by more explanation about *Profiles*, *Connections* and *Accessibility* as some of the important features in SNSs. Finally, the business aspects of these platforms are described, and in regards to marketing, their capabilities are compared with *Sponsored Search Results* which is currently the alternative channel for online advertising.

Boyd and Ellison (2007) define SNSs as web-based services with three main features. Firstly, these platforms allow individuals to build a public or semi-public *profile* within the application. Secondly, the members are able to create some sort of *connection* to other members. Finally, members are able to *explore* their own list of connections and

the connections which are made by other members. Although these features can be identified in many SNSs, each platform can provide specific features or conventions, which differentiate it from the other platforms. Currently, there are numerous examples of such platforms, and their key technological features are very similar. Usually the users build a profile as part of their registration process and then connect to other members (Boyd and Ellison, 2007). Furthermore, these platforms usually offer their users the possibility to invite their existing friends to join the same network. For example, after registration Facebook persuades users to invite their friends using their email contact lists. Also, based on the existing connections, Facebook suggests other members to be added as a Friend. Boyd and Ellison (2007) insist on using the term "Social Network Sites" to describe these platforms. In their view "networking" emphasises relationship initiation, often between strangers which is not necessarily the case in some of these platforms. Despite their view, the two terms have been used in the literature interchangeably. It has been suggested by Haythornthwaite (2005) that although these platforms allow individuals to meet strangers, the connections are usually between individuals with some sort of offline connections. Furthermore, Beer (2008) argues that Boyd and Ellison's (2007) terminology and definition could be limiting the concept, and explains that the term "Networking" is better for describing the process.

There are several factors which have been suggested in the literature as the reason for using SNSs. Some individuals use these platforms to meet new people. They use these platforms to find people based on common interests or to find emotional support which is not available offline (Lampe et al., 2006; Ellison et al., 2006). On the other hand, some users make use of these platforms to maintain the existing relations, and utilise the available capabilities such as instant messaging to communicate with their friend, family and colleagues. Using such platforms allow many people all around the world to interact with others, especially with those not geographically close to them (Ellison et al., 2007). There are other factors have been suggested in the literature as motive for using SNSs. These include learning, searching for friends, and social involvement (Shu and Chuang, 2011).

SNSs can be different in size and diversity of their audience. While some do not have any specific type for their audience, others attract people with specific similarities (e.g. language, nationality, religion). Although these platforms usually have a preference for geographical regions or linguistic groups, due to the nature of these platforms, this does

not always determine the platforms' user group. For example, *Orkut* was one of the first SNSs, and it was launched by Google in 2004. Although it started in the United States with an English-only interface, Portuguese-speaking Brazilians quickly became its dominant user group (Kopytoff, 2004), and consequently in 2008 Google announced that it would be fully managed and operated in Brazil (Ellison, 2007). In 2011 Google launched its new SNS called *Google Plus*, and three years later Orkut was permanently closed (Landeweerd et al., 2013).

Due to the innovative and evolving nature of Web 2.0 platforms, in addition to common Web 2.0 features many additional features and functionalities can be found on these platforms. Private Messaging is one of the features which are usually provided by these platforms. This feature is similar to emailing and some of the popular platforms such as Facebook and Google Plus have recently moved towards combining messaging and emailing. There are many other features which can be found on different platforms. Some of the popular features include photo-sharing, video-sharing, blogging, liking, voting, and commenting (Kaplan and Haenlein, 2010).

2.1.6.1 Profile

Profiles are one of the key elements in SNSs (Boyd and Ellison, 2007; Haythornthwaite, 2005; Manago et al., 2008; Back et al., 2010). These are usually created during registration, and the displayed information in users' profiles is usually provided an updated by users themselves. One of the questions about user profiles is the accuracy of individuals' profiles in representing the profile owner. Manago et al. (2008) suggest that the profiles are used to communicate an idealised character. On the other hand, a research by Back et al. (2010) shows that people are not using their profiles to promote an idealised virtual identity. They explain that SNSs might be an efficient medium for expressing and communicating real personality. They even suggest that this efficiency might be one of the reasons for popularity of such networks.

2.1.6.2 Connections

Connections between users have been identified as one of the fundamental characteristics of SNSs (Boyd and Ellison, 2007). These connections can be used to categorise such platforms. Some of them (e.g. Facebook) support the maintenance of pre-existing Social Networks. These platforms can be seen as a new communication tool between friends, family members, or colleagues. On the other hand, the second category includes the ones which facilitate connections between strangers with common interest. For instance, Online Dating Websites are among the most popular examples of

such platforms (Boyd and Ellison, 2007). Usually these platforms only allow the connection between users to form by bi-directional confirmation. These connections are usually initiated by one party and confirmed by the other, and only after confirmation the connection appears in their profiles. However, there are some other examples (e.g. Twitter) for which the default settings allow connections to be created without bi-directional confirmation. Although the nature of these connections is similar, there is a variety of names which have been used by different platforms to describe them. Popular terms include Friends, Contacts, Connections, Followers and Fans. The reasons people connect are different, and hence Boyd (2006) argues that despite its popularity in this context the term Friend can be misleading and the connections do not always mean friendship.

2.1.6.3 Accessibility

The information accessibility and visibility of users' profiles is one of the differentiating factors for SNSs. Some platforms allow users to decide how much information they would like to share with others. For example, Twitter makes profiles visible to anyone, unless the users choose to make it private. In this case, the profile owner can choose who has access to their profiles and posts. On the other hand, Facebook uses a more complex system. It provides some default privacy settings (e.g. Everyone, Only Friends, Friends of Friends), and additionally users can set access level specifically for each post or photo they share (Liu et al., 2011).

2.1.6.4 Business and Marketing

From the business point of view, SNSs can be seen as a new platform for online advertising and marketing (Royo-Vela and Casamassima, 2011). In many cases it can be relatively cheaper than traditional advertising channels such as TV Advertising, and could be more accurate than Keyword Targeting Techniques such as Sponsored Search Results (Tuten, 2008). Google is mainly known for its Search Engine (Vaughan, 2004), and for years it has produced most of its revenue from Sponsored Search Results. In simple terms, this can be described as an advertising system which targets potential customers based on their online searches. This kind of advertising has been used by business to increase their online presence. Small businesses and eCommerce websites also use these channels to promoted their product and services. In 2011 Google introduced its SNS (Google Plus), and one can argue this highlights the new competition from SNSs which provide more accurate techniques for advertising and targeting potential customers. Some SNSs (e.g. Facebook) offer features which can be

used by businesses to promote their brand, product, or services. This can be achieved organically by developing related pages and publishing content. Alternatively, some of these platforms such as Facebook offer targeted advertising facilities. These facilities target potential customers based on the information they have provided in their profiles. Royo-Vela and Casamassima (2011) suggest that targeting customers accurately based on their interests, makes this a powerful marketing tool.

2.1.7 Defining Web 2.0 Projects

By reviewing the existing literature and comparing the suggested definitions, one can argue there are a variety of definitions and assumed attributes even in more specific cases such as OCs and SNSs (see sections 2.1.5 and 2.1.6). Furthermore, there are currently platforms which share features from both of these types. For example, YouTube can be seen as an OC for users who have an interest in video content. However, almost all attributes which have been suggested for SNSs also exist in YouTube. Such overlaps in categorisations and diversity of definitions could have been resulted from the evolving nature of Web 2.0 technologies, and the way these platforms are adopted by their users. While OCs existed even before invention of the Internet or the Web, many of the underlying functionalities are similar to what is now regarded as Web 2.0 functionalities. Furthermore, innovation and technology can surprise the society once again by replacing the Internet or the Web with newer versions with more advanced capabilities. The popularity of smart phones and internet-based *Apps* can be seen as an example. Hence, this research proposes a specific definition for Web 2.0 projects, which can cover a wider range of online platforms and communication devices. The following definition is proposed, and it is meant throughout this thesis when refer to Web 2.0 projects:

> ***A Web 2.0 Project is a project in which software and electronic communication technologies (such as the Internet) are used, and the output of software can be affected by the input from its users.***

This is important to note that in many cases the users' input is affecting the outcome even without their knowledge. For example, based on proposed definition, a News Website which does not utilise any of the common Web 2.0 functionalities (e.g. posting, voting or commenting) can be still identified as a Web 2.0 Project if it has a section which shows popular items based on number of views. From this perspective, the proposed definition widens the inclusion criteria in comparison to suggested definitions

for OCs or SNSs in the existing literature (see sections 2.1.5 and 2.1.6). In other words, OCs, SNSs, News Websites with a Most Popular section, and even an eCommerce website which allows their users to rate their products, are all considered as Web 2.0 projects in this research. Furthermore, the *experiment* which is used in this research is associated with a Web 2.0 platform which lacks some of attributes which are suggested in the literature for OCs or SNSs (e.g. sustained membership, public profiles, connections), yet it shares many Web 2.0 features with them (e.g. sending content, evaluating content, sharing content).

2.1.8 Extended Inclusion Criteria

As explained earlier (see sections 2.1.5 and 2.1.6), online platforms are commonly categorised based on presented features and the purpose of their users. However, in this research the proposed definition for Web 2.0 Projects (see section 2.1.7) goes beyond such categorisations, and extends the inclusions criteria. In this section some of the successful Web 2.0 Projects are described and compared. They all use software and electronic communication technologies, and in all of them the output of software is somehow affected by the input from its users. Hence, they are all considered as Web 2.0 Projects in this research.

It should be noted that accessibility rules can be (and usually are) different for each platform. For example, in some cases the majority of content is publically available. Online platforms such as Wikipedia can be seen as "free for all" sources of unverified content. In these platforms, the collaborative efforts of volunteers have created a huge amount of content. Yet, the reliability of such content cannot be guaranteed. While in Wikipedia the content is mainly in form of text, YouTube can be regarded as an equivalent for video content. However, the access rules are more flexible in YouTube. It allows its users to categorise their videos as "private", "unlisted" or "public". Private videos can only be accessed by people who have been given the privilege by video owner, and unlisted videos are only accessible to the ones who have the direct link. On the other hand, public YouTube videos can be accessed by anyone and appear in search results.

The proposed definition (see section 2.1.7) also covers some new forms of eCommerce. For example, while Amazon or eBay cannot be categorised as SNS or OC based on the existing definitions in the literature (see sections 2.1.5 and 2.1.6), they satisfy the proposed definition in this thesis. These platforms allow individuals to easily become

sellers. Additionally, such platforms use specific feedback systems to represent the reliability and credibility of sellers. Furthermore, these platforms usually facilitate financial transaction and dispute resolution. However, once again these online platforms can be seen as software for which the output (e.g. auctioned items, comments, feedbacks) is affected by its users.

In addition to traditional forms of collaborative online platforms, the proposed definition includes online messengers and associated projects. For example, popular online messengers (e.g. WhatsApp, Telegram, Viber) combine software with internet technologies to provide a secure channel for communication. One can describe having a conversation on WhatsApp (or similar applications) as affecting output of underlying software. However, such changes to output are only visible to the parties who take part in that conversation. Such applications usually provide "group conversation" facilities in addition to one to one conversation. Group conversations represent content which is accessible to a specific group of individuals.

Based on determinants of accessibility of content, different factors can play a role in differentiating and defining new types of Web 2.0 Projects. A recent successful example is "Pokemon Go", which is a gaming application for smart phones. According to BBC News (2016) during its first week, there have been 15.3 million tweets about this viral application which is more than the number of tweets about Brexit in the week of the UK referendum. Based on the proposed definition (see section 2.1.7), such application are also regarded as Web 2.0 platforms. In this case the accessibility of content is determined based on location (i.e. to see and catch a pokemon users need to be close to it).

While in many Web 2.0 platforms the content is not verified, in some cases there are monitoring and approval processes which make the available content more credible. Such platforms can be used as Virtual Learning Environments in local or global scale. For example, Moodle is a Web 2.0 platform in which the teachers can create their own content and attract students from all around the globe. Consequently, they are responsible for the materials in their own courses (Moodle, 2016). There are similar platforms which are used locally in academic institutes. For example, in the University of Wolverhampton, WOLF (Wolverhampton Online Learning Framework) is a key system used by students and staff to support learning. It enables tutors to deliver learning material in support of classroom-based teaching (Wolverhampton, n.d.).

2.2 Online Behaviour

In the previous section (see section 2.1), the history of Web 2.0 was described, the suggested definitions and attributes were reviewed, popular examples were analysed, and finally a specific definition for Web 2.0 projects was proposed for the purpose of this study. In this section, the users' interaction with corresponding platforms will be described, and the existing literature in this regard will be critically reviewed. However, due to rather specific definition which is used throughout this study, the proposed definition for Online Behaviour and its boundaries are described before exploring the literature.

2.2.1 Defining Online Behaviour

Online Behaviour (OB) can be considered as an extension to general human behaviour. However, one may argue that OB is only an interpretation of how indivduals behave when using their tools. The first tools were made by humans around 2.5 million years ago (Semaw et al., 1997), and when corresponding human behaviour is investigated, the *offline-ness* of the behaviour is not highlighted. The same can be argued for what is known as OB. For example for a website user, clicking the mouse button can be analysed from a purely biological perspective, and without considering its value or meaning within the context of the associated website. In this study, a very specific and relevantly technical definition is proposed early in the section to help the reader during the rest of this chapter. It should be noted that the proposed definition is not specific to Web 2.0 platforms, and can be applied in other forms of online platforms such as eCommerce or News Websites.

In their original proposal for WorldWideWeb, Berners-Lee and Cailliau suggested a *client-server* structure for their HyperText Project (W3, 1990). In simple terms (and ignoring numerous additional complexities), it can be described as a structure in which the requests are sent from the individuals (*clients*) to central machines (*servers*) through an electronic network. These requests are then processed, and the software produces the response which is sent back to the corresponding clients. While this structure is currently the most popular structure on the Internet, other alternatives such as peer-to-peer networks (e.g. Torrents) have proposed less central structures (Schollmeier, 2001). However, in this study the proposed definition of OB is only applicable to client-server structures.

When the requests are received by the server, they contain information which can be used to identify the client. This information includes IP Address, Internet Cookies (ICs), and many other pieces of information which can be used for identification. Furthermore, in some cases the requests are sent to servers from other machines. Such requests are usually initiated by automatic software and for variety of reasons. These include Search Engine Crawlers (also known as robots, or bots), and many other forms of automatically generated requests. It should be noted that in some cases it can be rather complex to distinguish real humans from the machines only by analysing the requests. In sensitive cases (such as registration or online payment) array of additional techniques could be used to verify the client as a real human. For example, some platforms use CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). In more sensitive cases such as online payments, the banks or other organisations might telephone the user to verify their identity. In this study, the definition of OB only takes into account the requests which are initiated by *real humans*. It will be explained later (see section 4.14.6) that a variety of techniques were used during the experiment and data analysis of this research for identifying real humans and excluding the automated requests.

Many online projects use *tracking* techniques for different reasons such as gathering statistical usage information, and protecting the platform from vandalism and misuse. It should be noted that usage information is not limited to the number of page views, and it can extend to include more specific activities such as clicks, mouse movements, key presses, and so on. Such information can be valuable to understand users' interaction with the platform, and can help project managers in finding the deficiencies, and improving the platform. In this study, any information which is received by the server from the client, whether it is tracked and stored or not, is used as the basis for defining OB.

With the above boundaries in mind, the following definition is proposed and it is meant throughout this thesis when referring to Online Behaviour (OB):

> ***In a "client-server" environment, "Online Behaviour" is defined as the information which is received (or can be received) by the "server" from "real human clients", and is correspondent to any form of interaction with the underlying software.***

It should be emphasised again that such information might or might not be collected, stored, or analysed. Furthermore, the users may or may not have other forms of interaction with the project outside the online environment. However, in this study, information which is sent from real human clients to the server has been used as the only behavioural representative of users in any online platform.

2.2.2 Online Behaviour in Web 2.0 Projects

The Web started in 1990, but even before that users had experienced online communications, and interaction with each other and the machines through electronic networks (see section 2.1.2). However, during the last decade, online platforms have become more popular, and with introduction of Web 2.0 platforms, the *normal users* have become more influential in forming the content which is available and distributed on the Internet. The users' activities in association with Web 2.0 projects have been investigated by many scholars, and the existing literature suggests two major dimensions for OB in Web 2.0 platforms. These dimensions are sometimes referred to as *active* and *passive*.

Koh et al. (2007) categorise the users' activities in OCs. In their view most activities can be categorised as *posting* (active) or *viewing* (passive). One can argue that despite differences among Web 2.0 platforms and their functionalities, similar categorisation can be used for any Web 2.0 platform. In other words, user activities in a Web 2.0 platform can be divided into the ones which only include receiving content from the platform (passive), and the others which influence the available content on the platform (active). On this basis, the users' OB can be categorised as *Active Online Behaviour* (AOB) and *Passive Online Behaviour* (POB). It should be noted that while Web 2.0 projects provide the required tools for AOB, the users can choose to use any Web 2.0 platform as a traditional website (i.e. Web 1.0 platform), and their interaction with the platform can be limited POB (i.e. browsing its web pages and consuming its content). It also should be emphasised that usually POB is a prerequisite for AOB. For example, the users usually need to browse the available content before becoming an active participant and influencing the content.

2.2.3 Passive Online Behaviour

Within each Web 2.0 platform, there are usually a group of people who do not use the provided tools for interaction with the platform. For example, in OCs, some of the users do not participate in creation or improvement of knowledge, and only use the existing

knowledge which is provided by other member. For instance in Wikipedia project, while a majority of people only use the available content, there are some individuals who actively *generate* content, and *improve* the existing content. In February 2016 Wikipedia had around 400 million unique visitors, but there are only 130,000 users (around 0.03%) who have edited an article during this time (Alexa, 2016; Wikimedia Report Card, 2016; Wikipedia, 2016). While the ratio of active and passive participants can vary in different Web 2.0 platforms, van Mierlo (2014) suggests in many of such platforms around 1% of users create the vast majority of new content. It should be noted that in addition to creating content, there are other forms of AOB (e.g. sharing content) which can also be important in Web 2.0 projects.

While Web 2.0 functionalities enable users to interact with the platform, many users choose not to do so. In OCs, Nonnecke and Preece (2000) describe this group of users as *Lurkers*, and define them as community members who do not participate and do not contribute to their communities. Similar definitions can be applied to passive users in other forms of Web 2.0 platforms such as SNSs. However, based on the nature of each platform, the proportion of passive participants can vary. Furthermore, in some platforms membership might be *exclusive* to a specific group of people. For example, in the experiment which was used for this study, adding content is only available to the individuals with an official email address from the University of Manchester. For anyone else, apart from sharing the available content, any other form of active participation is not possible.

2.2.4 Passive Online Behaviour Measures

As explained earlier (see section 2.1.2), websites' *position* on the Web and their *impact* has been used to compare and evaluate websites (e.g. Web Impact Factor and PageRank). However, this evaluation is mainly based on the existing *Links* between the websites and their web pages, and does not take into account the size of their *audience*, and their *traffic*. The websites' traffic has been considered as one of the most important parameters for evaluating websites (Shen et al., 2006). An important factor which can be used for comparing online platforms is their traffic, and the associated measures (Phippen et al., 2004). These factors are usually linked to some sort of POB (e.g. visiting the website, browsing web pages, returning to the website). Different metrics have been suggested and used in the literature in this regard (Phippen et al., 2004). Some of the important factors include number of visitors, and the average number of web pages which are viewed in each visit. Also, the average duration for each visit, and

the average number of visits for each user within a fixed amount of time are also important factors (Burby et al., 2007; Gonçalves and Ramasco, 2008). It should be noted that in such factors users' AOB is not taken into account, and similar metrics can be used for Non-Web 2.0 platforms such as eCommerce or News websites. The web managers usually use third party tracking software to monitor the website's traffic and its users' POB. For example, *Google Analytics* is a service offered by Google, and can be used by web managers to generate different reports about the visitors. This is achieved by adding Google Analytics Tracking Code to the website. This code collects visitor data and sends it to Google. The web managers then can produce different type of reports on website's traffic, visitors, demographics and referrers. When Internet Cookies (ICs) are enabled on the client browser (see section 4.6.1), additional information will be collected. For example, new or returning visitors can be distinguished (Google, n.d.; Burby et al., 2007).

2.2.5 Active Online Behaviour

POB can be defined in any type of websites, including Web 2.0 and Non-Web 2.0 platforms (e.g. browsing web pages in an eCommerce website). However, Active Online Behaviour (AOB) is specific to Web 2.0 platforms, and in this study Web 2.0 projects have been defined as projects involving platforms which make use of such behaviours (see section 2.1.7). AOB and its attributes have been investigated by numerous researchers and in various fields. However, there seems to be a disagreement in what constitutes AOB in a Web 2.0 platform, and how it can be measured. This can be partly resulted from the changes in Web 2.0 platforms and the way users' interactions with these platforms have evolved. While there might be a disagreement among the scholar on how to define AOB, there seems to be no disagreement about the importance and significant role of such behaviour in many Web 2.0 platforms. Tedjamulia et al. (2005) identify AOB as a major determinant of success for OCs, and Hsu and Lin (2008) argue this dependency could be considered as major obstacle for such communities. In other words, if AOB does not exist, in many cases the associated Web 2.0 project cannot be successful.

As explained earlier (see section 2.2.2), some researchers assume two distinct types for OB, namely *posting* and *viewing*. On this basis, posting is associated with AOB, and viewing is regarded as POB. Posting content in form of text, image, sound and video is one of the primary forms of AOB in many Web 2.0 platforms. In the early years of Web 2.0, blogging platforms (e.g. Blogger.com) enabled normal Internet users to publish

content online. However, this content generation was mainly in form of text. This was later expanded and many Web 2.0 platforms provided the required infrastructure for other forms of content. For example, online platforms such as YouTube, Flicker and SoundCloud are among the popular platforms for user generated content in form of video, image, and audio respectively.

While posting content is fundamental in many Web 2.0 projects, there are other forms of AOB which are available in many Web 2.0 platforms. The conceptual diversity of Web 2.0 platforms provides a range of possibilities in how users can actively participate. For example, in many Web 2.0 platforms there are available features which enable users to evaluate the content. For instance, Web 2.0 features such as Facebook's *Like* provides a new method for AOB. In this type of participation, users do not add to the available content, but they can show their positive feedback towards it. It should be noted that the dependency of Web 2.0 projects on AOB (or specific types of AOB) varies. For example, in a video sharing platform such as YouTube the whole existence of the concept is based on AOB of users in form of uploading video content. On the other hand, a News Website with Comments Section is not entirely dependent on readers' comments, and in some cases it might benefit more from other forms of AOB such as sharing the content.

In many Web 2.0 platforms the content is solely provided by active members, and Cheung and Lee (2009) argue that the value of such communities is dependent on existence of *significant numbers* of active members. Wang and Fesenmaier (2004) suggest in such communities there needs to be a *large/balanced* proportion of active members, and Ardichvili et al. (2003) explain that AOB is a fundamental success factor for OCs.

2.2.6 Active Online Behaviour Dimensions

In many Web 2.0 platforms posting original content is the primary form of AOB. However, there are other forms which can be assumed for this type of behaviour. For example, sending comment is a common form of giving feedback in Web 2.0 platforms and can be categorised as AOB. Furthermore, in many Web 2.0 platforms features such as Vote, Rate, or Like are used to facilitate easer forms of giving feedback. Such feedbacks are usually displayed next to the original content and can be used to determine importance, popularity or credibility of the corresponding content. In case of Facebook, the number of Likes is also one of the factors which determine what items

are included in the News Feed, and in which order they are displayed (Luckerson, 2015).

Benkler (2002) assumes three dimensions for online collaboration, and these dimensions are used is this study to categorise different types of AOB. These dimensions are described by Benkler (2002) as *Content*, *Relevance/Accreditation* and *Value-Added Distribution*.

2.2.6.1 Content

Benkler (2002) describes the Web as a global library which is produced collaboratively by millions of people, and argues that its power comes from its *large number* of contributors. Web 2.0 features usually enable users to add different types of content such as text, image, or video. There are also other forms of Web 2.0 features which enable users to improve the available content. For example, in Wikipedia in addition to adding new articles, the users are also able to modify the articles which are already available and have been collaboratively produced by other users. Consequently, unlike articles of printed encyclopaedias, Wikipedia articles can be improved and extended during time. Furthermore, this dimension of AOB can also be regarded to cases which duplicated, wrong, or inappropriate content is removed or reported by active participants. In other words, the Content dimension of AOB consists of any activities which are directly linked to user generated content.

2.2.6.2 Relevance/Accreditation

Benkler (2002) argues that the existence of content on the Internet and the accessibility through the Search Engines in not enough. He regards the unaccredited and unorganised online information as *gobbledygook*, and suggests such information requires to be credited by real humans. In his view, even models such as Google's PageRank (see section 2.1.2) which use the structure of the Web for evaluating websites and web pages are ultimately based on human input. He implies that a *Link* to a website or its web pages can be seen as a *Vote*. Benkler (2002) identifies distributed production of relevance and accreditation as an important factor in organising the huge amount of user generated content on the Internet. There are different Web 2.0 features which are associated with this dimension of AOB. Many Web 2.0 platforms use *Ranking Systems* to allow their users to evaluate the available content. For example, Facebook and many other platforms provide Like Button or similar features, and users can use these features to show their interest or the importance of a specific piece of content. Some other websites (e.g. YouTube) use a combination of Like and Dislike buttons to provide both

positive and negative forms of feedback. Furthermore, such functionalities sometimes can be embedded into other websites. For example in the second stage of the *experiment* which is associated with this research (see section 4.5), two different systems were provided for users to evaluate the content (i.e. submitted poems). The first method was using a *Vote Button* which could only be used by registered users to vote for their favourite poems. Additionally, a Facebook *Like Button* was embedded next to each poem which allowed Facebook users to *Like* the poems. The number of Votes and Likes were displayed next to each poem.

2.2.6.3 Value-Added Distribution

The final dimension which can be assumed for AOB is associated with distribution of content. Benkler (2002) explains that after content is produced and evaluated collaboratively, it needs to be distributed. He highlights the underlying combination of software and hardware which provides this model of collaborative distribution. In this study distribution dimension of AOB for each platform is represented with any activity which involves using Web 2.0 features for online distribution of the Links to available content within that platform. Email Lists facilitated one of the first forms of online distribution of content on the Internet. By using email as a communication tool, the users have been able to send each other different forms of content or Links to interesting or important web pages on the Internet. Also, in many OCs the users have been able to post Links to other web pages. Furthermore, in recent years Web 2.0 platforms have introduced new features which allow users to easily distribute content. For example, Facebook utilises a *Share Button* which can be embedded into web pages by the site managers, and enables users to share those web pages with their Facebook friends.

2.2.7 Active Online Behaviour Measures

Web 2.0 platforms can be used for different purposes and within different industries. Despite the differences among such projects, users' OB has a fundamental role in success of these projects. Consequently, users' behaviour and in particular their AOB is very important for Web 2.0 project managers (Wang and Fesenmaier, 2003; Phippen et al. 2004; Bishop, 2007). However, there is a disagreement among the academics in regards to measuring AOB. This could be resulted from the changing nature of Web 2.0 projects and the diversity of available platforms.

As explained earlier (see section 2.2.2), some researchers have assumed two dimensions for users' activities in Web 2.0 platforms; namely *posting* and *viewing*. On this basis,

AOB for each user can be measured based on the amount of posting they have done. For example, AOB can be measured based on total number of posts during a fixed period, or during the length of membership within the community. Alternatively, average number of posts within a fixed duration can be measured and used to compare AOB among the users (Nonnecke and Preece, 2000; Chen, 2004; Preece et al., 2004; Nonnecke et al., 2006; Chen and Hung, 2010). While measuring AOB in this way can be used to evaluate and compare AOB, there are different criteria which have been suggested by previous researchers. Chen et al., (2004) argue that active participants are the ones whose posting activities are higher than average for the community. On the other hand, other researchers such as Preece et al. (2004) identify users with any posting activity as an active participant. It should be noted that while posting can be seen as AOB, the quality of posted messages might vary from one user to the other. Although it could be difficult to evaluate the quality of posted messages and take their quality into account, other alternatives have been suggested by previous researchers. For example, Arguello et al. (2006) only consider posting messages in an OC as AOB if the posted messages receive at least one reply. On the other hand, some researchers have concluded that posting activity is *not* necessary for AOB, and the users who spend *significant time* on the platform and read posts by others, can also be categorised as active participants (Wang and Fesenmaier, 2004).

Considering the variety of Web 2.0 functionalities which are available to the users, both definition and measures can vary from one platform to the other. For example, in SNSs, the number of connections can be seen as a measure of AOB. However, such connections are not available in many other forms of Web 2.0 platforms. Also, there are other forms of behaviour which cannot be considered as posting to the investigated platform, but have been considered as AOB (Benkler, 2002). For example, when the content of one platform is shared in another platform, while the posting activity is only associated with the second platform, the user has been *distributing* the content from the first platform, and hence one can suggest their behaviour can be considered *active* in both platforms.

In this study, AOB is quantified and compared based on activities which can be linked to any of its described dimensions (see section 2.2.6). For example, in a Web 2.0 platform which provides *Voting* functionality, users' interaction with this feature is considered as AOB, and the associated metrics such as number of votes have been used to quantify this activity.

2.3 Influencing Online Behaviour

In many Web 2.0 projects, the success of project is unarguably dependent on how individuals use the platform, and whether they choose to actively interact with it. Consequently, it would be desired to influence users' OB towards achieving the project gaols. Numerous factors have been suggested in the literature to have an effect on different aspects of OB. As explained earlier (see section 2.2.2), OB itself have two important dimensions, namely AOB and POB. While there are disagreements among the researchers on what can influence each of these dimensions, there seems to be an agreement on fundamentally different nature of these dimensions, and accordingly different influencing factors for each. Furthermore, during last decades the technological changes such as availability of high speed connections and popularity of smart phones have changed users' behaviour, and it can be argued that such changes can also change the influencing factors for it. Also, the society is changing as a whole, and the subjective norms are being influence by technology. This too can have an impact on the factors which can influence OB (Ayeh et al., 2013; Filieri, 2015; He et al., 2013; Church and de Oliveira, 2013; Kim et al., 2007; Mellet et al., 2014; Pizzato et al., 2013; Beracha and Wintoki, 2013; Finkel et al., 2012; Guadagno et al., 2012; Blackwell et al., 2014).

In OCs, Koh et al. (2007) suggest different stimulation drivers for increasing users' AOB. These include leader involvement, offline interaction, usefulness, and IT infrastructure quality. They describe that in any given community, the posting activity stimulants are not the same as the viewing activity stimulants. The viewing activity in their view is associated with perception by community members of community usefulness. On the other hand, posting activity is mainly influenced by offline interaction and the quality of the IT infrastructure. Kim (2000) also investigates OCs, and identifies clear purpose, clear roles, leadership by moderators and online/offline events as the factors which can influence users' OB. For example, he suggests events which are organised by the OC, can improve members' identification within the community, and consequently influence their behaviour towards the platform. Furthermore, Rheingold (2000) highlights the importance of sufficient human feelings in shaping an OC, and the role it can have in influencing behaviour. In many Web 2.0 platforms, Knowledge Sharing has been suggested as the main motive for user participation (Constant et al., 1994; Koh and Kim, 2004; Liao, 2008; Hsu and Lin, 2008). The influencing factors for knowledge sharing activities have been investigated

by Chiu et al. (2006), and they suggest trust and shared vision as the main factors which can affect OB with regards to knowledge sharing. They also suggest that quality of shared information is significantly affected by social interactions between members. Similarly, Lu and Yang (2011) highlight trust and reciprocity as the important factors which influence quality of shared information. However, they argue that social ties between users have a greater influence on the quantity of shared information. On the other hand, Kankanhalli et al. (2005) have found a negative association between trust and AOB of users. They suggest reward, joy and self efficacy as the factors which positively influence AOB. Chai and Kim (2012) suggest in SNSs ethical culture, social ties, and SB as major influencing factors for users' OB. Hsu and Lin (2008) identify joy and reputation as important factors in forming users' OB.

In addition to cognitive factors (e.g. trust and ethical culture), technological aspects of Web 2.0 platforms have also been considered as influencing factors for OB (Koh et al., 2007). However, due to the diversity of available functionalities within Web 2.0 platforms, there are a wide range of technological aspects which can be identified in Web 2.0 platforms. Usability of such platforms, or in other words *Ease of Use* has been highlighted by many scholars as an important technological factor (Hsu and Lin, 2008; Lee et al., 2006; Chung and Buhalis, 2008; Wang and Fesenmaier, 2004). This is not specific to Web 2.0 projects, and similar findings have been found in other types of online platforms such as eCommerce. Hsu and Lin (2008) investigate a group of bloggers' behaviour, and in addition to cognitive factors they identify the effects of technological factors (such as Ease of Use) on OB. Wang and Fesenmaier (2004) identify different dimensions for functional factors, including information, efficiency, and convenience. Sangwan (2005) assumes more detailed dimensions for functional factors including objectivity of information, its high value, and its exactness.

2.4 Virtual Belonging in Web 2.0 Projects

Users' emotional connection to specific types of Web 2.0 platforms such as OCs and SNSs have been investigated by many scholars (see sections 2.1.5 and 2.1.6). Among the investigated factors, *Sense of Belonging* (SB) has been highlighted by many researchers as an important factor in shaping users' behaviour (see section 2.4.4). In this section the related literature is reviewed, and suggested attributes and effects of such a psychological connection to Web 2.0 platforms is explored. This will be followed by

highlighting a gap in the existing literature, and on this basis *Virtual Belonging* (VB) is defined as a quantitative representation of users' *connections* to Web 2.0 platforms.

2.4.1 Belongingness

Maslow (1943) explores human motives, and categorises them as different human needs. He suggests that the effort to satisfy these needs forms human behaviour. He argues when none of the human needs are satisfied, individuals will be dominated by their physiological needs, and all other needs will be pushed into the background. Furthermore, when physiological needs are satisfied, individuals search for safety. Maslow (1943) suggests when physiological needs are satisfied, and safety is achieved, individuals will be dominated by the need for love, affection and *belongingness*. He states that at this point individuals feel the "absence of friends, or a sweetheart, or a wife, or children" (p.381). At this stage Maslow (1943) suggests each individual's goal would be to have "affectionate relations with people in general, namely, for a place in his group" (p.381). He claims in many cases maladjustment and severe psychopathology could be associated with dissatisfaction of what he calls the love needs. He also highlights that the love needs involve both giving and receiving love. Maslow (1943) continues with adding higher levels of needs, namely, the esteem needs and the need for self-actualisation. What could be observed in Maslow's classification of needs is the position of "others" in each level. While self is at the centre of physiological and safety needs, satisfying the love needs would require other individuals. In other words, an individual can be safe and not hungry on their own, but they would need others to be able to find a place among them. To satisfy higher levels of need, the individuals may need to improve their status in the society. However, to achieve belongingness being part of a group seems to be more important than the achieved status. Anant (1966) defines belongingness as "sense of personal involvement in a social system so that persons feel themselves to be an indispensable and integral part of the system" (p.21). He also argues that there is a positive relationship between the sense of belongingness and mental health. On this basis, belongingness can influence both mind and behaviour.

2.4.2 Sense of Belonging

Belongingness is defined as "the quality or state of being an essential or important part of something" (Dictionary.com, n.d.). The feeling which is associated with being in such a state is sometimes referred to as *Sense of Belonging* (SB).

Bollen and Hoyle (1990) investigate perceived cohesion, and propose two dimensions for it; SB and Feelings of Moral. They state "Sense of belonging is fundamental to a member's identification with a group and has numerous consequences for behaviour" (p.484). Hagerty et al. (1992) present SB as a mental health concept and define it as "the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment" (p.173). They consider SB as a specific process of relatedness. Hagerty et al. (1993) also extend the concept of belonging by including *relatedness*. They argue that maintaining relatedness is a common human concern, and describe relatedness as the experience of individuals in all types of relationships to others, objects, environments, society and self. Based on this definition, they identify SB as a major process in establishing state of relatedness. Later research by Hagerty et al. (1996) identifies SB as a useful concept to explore social and psychological functioning. They also argue that SB can both influence the behaviour, and be influenced by the behaviour. Markus and Kitayama (1991) suggest that in some cultures SB can become so strong that instead of individual, relationships can be considered as functional unit of reflection.

Hagerty and Patusky (1995) outline two defining attributes for SB. The initial attribute in their view is *valued involvement* or the experience of feeling valued, needed, or accepted. This attribute is associated with the individual experience of being valued or important with respect to others. They suggest *fit* as the other attribute, describing it as the perception that the individuals' characteristics articulate with the system or environment. Bollen and Hoyle (1990) suggest cognitive and affective dimensions for SB. The cognitive dimension is based on accumulated information about the experience of being part of the group. On the other hand, the affective dimension is based on the feelings of the individual towards the group and its members as a whole.

Hurtado and Carter (1997) show that in a college environment some activities such as discussing the course content with other students outside class, and membership in religious and social communities are strongly associated with students' SB. Sedgwick and Yonge (2008) study the link between SB and learning among the nursing student in rural areas, and suggest it can be influenced by individual characteristics, interpersonal relationships, and the environment. Levett-Jones and Lathlean (2008) present belongingness as a deeply personal and contextually mediated experience, and suggest it is emerged from the following feelings: (a) feeling secure, accepted, included, valued

and respected by a defined group (b) feeling connected with or integral to the group (c) feeling professional and/or personal values are in harmony with those of the group.

Many researchers have shown the importance of SB in forming the feelings and behaviour of human beings. For example, stress and depression have been highlighted by many findings as some of the factors which can be influenced by SB. Choenarom et al. (2005) examine role of SB on symptoms of depression in individuals with history of depression, and they suggest SB as a significant predictor for depression. Another research by Bay et al. (2002) investigates the effect of SB on the survivors of traumatic brain injury. According to their findings, depression for these patients is negatively related with their post-injury SB. Furthermore, in regards to the effects of SB, the existing literature suggests similar findings in wide range of investigated groups. McLaren et al. (2007) examine a group of retired individuals (average age of 71), and show SB is negatively associated with depression and suicidal ideation. On this basis they identify SB as the protective factors for suicidal ideation in aging adults. McLaren (2006) investigate a group of women and show that SB is negatively correlated with dysphoria. Nuttman-Shwartz and Dekel (2009) investigate contribution of SB to stress responses among Israeli students in a conflict zone. Their results show that higher SB is associated with lower levels of distress. They also argue the existence of community can improve the ability to cope in the face of a continuous threat. Sedgwick and Yonge (2008) suggest SB can enhance education, and make the students feel safe and comfortable. On the other hand, the absence of SB can cause anger and confusion. Levett-Jones and Lathlean (2008) explain that for students SB can increase the feeling of being safe, comfortable, satisfied and happy. They also suggest it is directly linked to students' degree of self-efficacy, and even their future career decisions. Furthermore, they conclude that SB can significantly influence students' capacity and motivation to engage in learning opportunities.

2.4.3 Sense of Belonging Measures

Different instruments have been suggested in the literature to measure SB in different settings. For example, Bollen and Hoyle (1990) suggest a *Perceived Cohesion Scale*, for which the first dimension is SB. They measure SB to the community based on the following three questions answered in 11 points Likert scale: (a) I feel a sense of belonging to the community. (b) I feel that I am a member of the community. (c) I see myself as part of the community.

A later study by Hagerty and Patusky (1995) focuses specifically on measuring SB. They suggest a 33-item self-report measure as the *Sense of Belonging Instrument*. They argue that there are two distinct dimensions which need to be measure for SB. The first dimension represents the psychological state of SB, and includes valued involvement and fit. Furthermore, Hagerty and Patusky (1995) suggest another dimension which can be measured by their proposed instrument. This dimension represents antecedents of SB which includes desire and ability for developing SB.

2.4.4 Sense of Belonging in Web 2.0 Projects

When Maslow (1943) talked about *belongingness*, he suggested this could be achieved by finding friends, sweetheart, wife, or children. However, with advances in communication technologies and the introduction of Web 2.0 platforms, one may argue Facebook friends or similar virtual relationships could be added to his list.

SB has been suggested in the literature as an important influencing factor in Web 2.0 projects. Zhao et al. (2012) investigate OCs and explain that without SB the online discussion groups cannot be considered as OC. They suggest that the stickiness of an OC is reflected by its members' SB, and argue that it is a vital factor for success of any OC. Different psychological and functional factors have been suggested as determinant of SB in online platforms. Lin (2007) suggests perceived ease of use and perceived usefulness have a significant positive effect on SB in OCs. She also highlights offline activities as another significant determinant of SB. In a later study, Lin (2008) suggests that trust and member satisfaction can also positively influence SB.

Zhao et al. (2012) investigate the effects of SB on *getting* and *sharing* knowledge. In OCs, they define SB as the feeling of being integral parts of the OC, and demonstrate that SB has a positive effect on community members' intention to get and share knowledge. However, their findings suggest that the effect of SB on the intention to share knowledge is stronger than its effect on the intention to get knowledge. Another research by Teo et al. (2003) suggests that SB increases participation in virtual learning communities. Lin (2007) suggests that SB also positively affects users' intention to use OCs. Individuals' loyalty to an OC has also been suggested to be positively influenced by their SB to that community (Lin, 2008). A research by Roberts (1998) investigates newsgroups as an OC and shows that active participants have more SB towards the community. He implies the strong SB leads to strong intention for further involvement in OCs. Additionally, he claims that without SB no involvement or participation would

be forthcoming from users. The importance of SB is not limited to text-based communities. Kim and Zhang (2011) investigate the satisfaction and loyalty of 3D Virtual World users, and demonstrate that SB has a positive effect on both users' satisfaction and loyalty.

Teo et al. (2003) measure SB by asking the subjects to indicate if they felt strong sense of being part of the community. Additionally, commitment to the community and strong morale among the community members are used to measure SB in OCs (Teo et al., 2003).

2.4.5 Gap in the Literature

The existing literature suggests SB as an important determinant for OB in Web 2.0 platforms, and OCs in particular. Also, different determinant have been suggested for such a feeling towards online platforms (Teo et al., 2003; Lin, 2007; Kim and Zhang, 2011; Zhao et al., 2012). However, from practical point of view, two limitations have been identified in the existing literature. As described before (see section 2.4.3), SB is a qualitative measure, and it can only be measured if the platform is implemented, the community exists, and there are means to measure their SB. From a project management perspective, such data is not available in the early stages of projects, in which the concept is forming and the projects' viability is being investigated. In the later stages of such projects, which the concept is finalised and the platform is being implemented by programmers, there are still no users and nor any community. Hence, investigating SB in the early stages of Web 2.0 projects is *not* an option. Consequently, the existing perspective towards SB does not provide any practical suggestions for how to choose and implement Web 2.0 features to improve SB for future users. Furthermore, another practical limitation can be identified in the existing literature. Many Web 2.0 projects attract a global audience with different demographics, cultures and languages. Hence, collecting and analysing qualitative data in association with SB from such diverse subjects could be complex and relatively expensive. Considering the fact that many Web 2.0 projects are managed with relatively low budgets, an ongoing collection and qualitative investigation of such data might not be a viable option. Furthermore, such investigations cannot be performed in real time. In other words, the results can only be obtained if a sufficient number of users agree to provide qualitative feedback, and spare the time to complete the process (e.g. taking part in interviews or completing questionnaires).

2.4.6 Defining Virtual Belonging

In Web 2.0 projects, belonging (or ownership) can be discussed from a legal perspective. For example, when a Facebook user uploads a photo to their Facebook account, it is part of their profile and yet stored on Facebook servers, and one may question its owner. Such platforms usually ask their users to accept their terms and conditions by ticking a box, but in many cases users accept the presented terms even without reading them (Smithers, 2011). In case of disputes, pages and pages of terms and conditions will need to be analysed by experts to identify the rights which can be assumed, and their rightful owners. The legal investigation of content owners in Web 2.0 platforms is not within the scope of this study.

While the legal ownership of an upload photo in Web 2.0 platforms can be disputed, sometimes the users refer to such a photo as *their* photo. For example, they might say "20 people have commented on *my* photo". Furthermore, they might not even have the rights to the photo that they have published, and for instance they could have downloaded it from the Internet. Yet, they may regard it as their photo because it is in their profile. This study investigates such connections between users and Web 2.0 platforms, and *Virtual Belonging* (VB) is introduced in this research as a quantitative measure for users' connectivity to Web 2.0 platforms.

Hagerty et al. (1992) consider SB as a specific process of relatedness, and Hagerty et al. (1993) describe relatedness as the experience of individuals in all types of relationships, including relationships to objects and environments. On this basis the connections which exist between individuals and Web 2.0 platforms can influence their SB. Zhao et al. (2012) define SB as the feeling of being integral parts of the online environment. In many Web 2.0 platforms users can become an integral part of the environment by using Web 2.0 features. When a user sends a comment in an OC, their comment can potentially remain there *forever* and even after their death. While this does not necessarily produce a *feeling* of being an integral part of the community, it can be seen as being an integral part of the environment. On this basis, the following definition is proposed for VB and it is meant throughout this thesis when refer to Virtual Belonging (VB):

> ***For users of a Web 2.0 Platform, "Virtual Belonging" is defined as any connection between the users and the platform, which is initiated by the real human users, and by utilising any form of***

Web 2.0 features. Such connections are stored digitally in online databases or similar forms of storage, and may or may not influence the output from underlying software for other users.

It should be emphasised that due to conceptual diversity of Web 2.0 platforms, and variety of available Web 2.0 features, the type and significance of highlighted connections between users and platform may vary from one platform to another. However, in this study three dimensions are proposed for such connections in any Web 2.0 environment, and the association of each dimension on users' OB is investigated later (see sections 5.6, 5.7 and 5.8).

2.4.6.1 Inclusion

McMillan and Chavis (1986) suggest *membership* as an attribute for sense of community. They define membership as "feeling of being a part", and emphasise that it must have boundaries. In other words, there must be some individuals who are members and others who are not. In Web 2.0 platforms, membership is usually interpreted as going through a registration process. It should be noted that based on the McMillan and Chavis (1986) explanation of membership, registration might not be required for sense of membership. They explain that membership is the feeling of belonging or of sharing a sense of personal relatedness. Hagerty et al. (1992) argues that SB is associated with the feeling of being an integral part of a system or environment. Levett-Jones and Lathlean (2008) also suggest the feeling of being included is associated with SB. Bollen and Hoyle (1990) suggest three factors for measuring SB (see section 2.4.3). One of these factors is based on how much individuals feel to be a member of the community. While in offline communities membership can be a complex phenomenon and difficult to measure, in many Web 2.0 platforms membership can be easily obtained by registration. In some cases different types of membership are available (e.g. free membership and paid membership), and this can be seen as different levels for membership. Hagerty and Patusky (1995) present ability and desire as antecedents of SB. In many Web 2.0 platforms, engagement requires membership, and Web 2.0 features such as registration are provided to obtain membership. Such process usually enables users to participate if they have the desire to do so. In many Web 2.0 platforms, getting information does not require registration, however it can produce a sense of personal relatedness, and it can even be the starting point for becoming an active participant. Furthermore, in some Web 2.0 platforms such as Wikipedia, registration is not always required for AOB. In this case, for users who have not gone through

registration process, their IP Address is used instead of their username (Wikipedia, 2016a).

In this study VB is proposed as quantitative measure based on logical connections between the platform and its users, and hence no cognitive aspects are assumed for it. The following definition is proposed for the "Inclusion" dimension of VB and it is meant throughout this thesis:

> ***"Inclusion" dimension of "Virtual Belonging" includes any connection between users and Web 2.0 platforms, which gives them additional access to available information and/or Web 2.0 features, whether or not they choose to use these additional privileges.***

It should be noted that this type of connections between users and platform usually involves registration. However, in Web 2.0 platforms in which registration is not required for AOB, one can argue that any user who is aware of this fact has already achieved this dimension of VB. Furthermore, due to diversity of Web 2.0 platforms, different representations for membership can be imagined. For example, in some cases after registration users may go through an additional stage to verify their email address. Also, in some Web 2.0 platforms, there are both free and paid registrations, and in such cases the associated privileges are usually different.

2.4.6.2 Involvement

In a Web 2.0 platform, the Inclusion dimension of VB is defined in this study as any connection which enables users to perform AOB. As explained in previous section, the registered users (or members) may or may not use these privileges, and they can remain a passive participant. However, if they choose to become active participants, it involves creating new connections between the associated platform and themselves. Such connections can be created by utilising different Web 2.0 features, and users' involvement with each platform can be different. Hagerty et al. (1992) describe SB as the experience of personal involvement in a system or environment. Web 2.0 platforms usually provide the required infrastructure for users' involvement. Furthermore, one of the factors which are suggested by Bollen and Hoyle (1990) to be correlated with SB is individuals' perception of how much they see themselves as part of the community. In Web 2.0 platforms members who publish content usually are introduced to other

members with their username. Hence, using Web 2.0 features and influencing the available content can be seen as being part of the community.

When users choose to utilise Web 2.0 features, their activities create logical conections between them and the platform. On this basis, the following definition is proposed for the "Involvement" dimension of VB and it is meant throughout this thesis:

> ***"Involvement" dimension of "Virtual Belonging" includes any connections between users and Web 2.0 platforms, which is associated with knowingly using Web 2.0 features by real human users, and can influence the output of underlying software.***

It should be noted that in Web 2.0 platforms, different methods exist for creating such connections. These include posting content (such as text and multimedia), improving content (by editing or adding comments), and evaluating the content (by voting, rating or other similar features). The term "knowingly" is included purposely in the proposed definition because in some Web 2.0 platforms users might affect the output from underlying software without being aware of it. For example, many Web 2.0 platforms use content popularity as a factor for what is displayed on their homepage, and "number of views" can be among the factors which are used for calculating popularity. Furthermore, in many cases (Such as YouTube) number of views for each item (e.g. video) is displayed on the website. So, one can argue that each visitor of the website affects the output from the underlying software. However, in this study these connections are only considered as Involvement if the users are aware of the effects of their behaviour, and knowingly change their behaviour to use such features. An actual example of such behaviour is highlighted and further investigated in the experiment which is associated with this study (see section 4.14.6).

2.4.6.3 Influence

Hagerty and Patusky (1995) identify *valued involvement* as one of the attributes for SB. They describe such a feeling as the experience of feeling valued and accepted. While outside the Web it can be difficult to measure such an experience, in Web 2.0 platforms there are different features which can be considered as quantitative representations of being valued and accepted. For example, when a Facebook user shares a photo, the number of Likes they receive can be seen as a measure of acceptance or perceived value. Levett-Jones and Lathlean (2008) suggest that SB can result from the feeling that one's values are in harmony with those of the group. In Web 2.0 platforms, some of the

features such as voting or liking can be used to show acceptance or agreement. In other words, the number of votes or likes can be seen as the value which other users have collectively assumed for the content or its owner.

As explained in the previous section, in a Web 2.0 platform, users usually need to become a member and use the available Web 2.0 features to get involved. In this study, the proposed Involvement dimension of VB does not take into account the effects of such behaviour on other users. However, this is represented as another dimension for VB. For example, when users post any form of content in a Web 2.0 platform, they are already connected to it. However, when their post receives comments or votes, a new form of connection can be assumed between the user who has posted the content and the associated platform. On this basis the following definition is proposed for the "Influence" dimension of VB and it is meant throughout this thesis:

> ***"Influence" dimension of "Virtual Belonging" includes any connection between users and Web 2.0 platforms, which is associated with both their use and other users' consequence use of available Web 2.0 features, in a way that can influence the output of underlying software.***

It should be noted that due to conceptual and functional diversity of Web 2.0 platforms, different representations can be identified for such connections. For example in Facebook, individuals need to go through a registration process and provide an email address to become a member. At this point, Inclusion dimension of VB can be assumed for them. After registration, they will be able to post content to their profile, which creates new connections between them and the associated platform. Such logical connections between Web 2.0 platforms and their users are considered in this study as Involvement. When other users start to comment on their post or Like it, it can be argued that their behaviour has influenced others' behaviour. At this stage the Influence dimension of VB is assumed for them. It should be emphasised that while similar features can be found in many other Web 2.0 platforms, in some cases identifying the proposed dimensions might be more complex. For example, in Wikipedia registration is not always compulsory for generating content. So, one can argue that Inclusion can be assumed for anyone who is aware of this rule, and is capable of using the available features for getting involved. Primary types of AOB in Wikipedia are creating new articles, or editing the existing ones. Any of these activities will create connections

between users and Wikipedia, which are identified in this study as Involvement. Finally, when a user sends content to Wikipedia, other users can edit it. In this case, the first user has influenced other users' behaviour, and hence in this study the Influence dimension of VB is assumed for them.

2.4.7 Investigating Virtual Belonging

In previous section VB and its dimensions were defined. In later chapters, these will be revisited again and again to make them operational, and investigate any association between VB and OB. In this section an overview of future references to VB and its dimensions is presented to help the reader in following the investigation process.

VB and its dimensions were defined in section 2.4.6. These dimensions will be repeated in chapter 3 (see sections 3.7.2, 3.7.3, and 3.7.4) and two hypotheses are presented in association with each dimension of VB. In chapter 4, the experiment design is described, and each dimension of VB is described within the context of the undertaken experiment (see sections 4.10.1, 4.10.2, and 4.10.3). In chapter 5, the results are presented, and each dimension of VB is investigated for possible association with OB (see sections 5.6, 5.7, and 5.8). In chapter 6, the results are discussed, and the findings are presented for each dimension of VB (see sections 6.1.2, 6.1.3, and 6.1.4). This is followed by combining all dimensions and introducing Virtual Belonging Hierarchy (see section 6.2). In chapter 7, this research is concluded and a set of practical recommendations is presented in association with each dimension of VB (see section 7.2).

2.5 Summary

This chapter started by reviewing the literature in association with OB in Web 2.0 projects. Thereafter, some of the influencing factors for OB were described, and SB was investigated in more details. Next, some practical limitations were identified and highlighted in regards to use of SB in Web 2.0 projects. Finally, to address the associated gap in the literature, VB was introduced and defined in Web 2.0 projects.

Initially, in this chapter the history and evolution of Web 2.0 technologies were described, and some of the major milestones were highlighted (see section 2.1.2). Thereafter, the categorisation of web based services into *Web 1.0* and *Web 2.0* was described based on the proposed definition by O'Reilly (2007). This was followed by focusing on current most popular forms of online platforms based on traffic reports by

Alexa (2016). Among these platforms some specific types of Web 2.0 platforms were identified, and therefore OCs and SNSs were further investigated as some of the most popular forms of Web 2.0 platforms (see sections 2.1.5 and 2.1.6). This was followed by defining Web 2.0 projects specifically for the purpose of this study (see section 2.1.7).

The chapter continued by presenting a rather technical definition for OB (see section 2.2.1), and reviewing the associated literature in this regard. While there are a diverse array of Web 2.0 platforms, two dimensions have been suggested by previous investigations for OB, namely AOB and POB (see sections 2.2.3 and 2.2.5). Hence, to improve understanding of OB, some of the previous researches for each of these dimensions were reviewed in this chapter, and the suggested definitions, attributes, measures and causes were described. In regards to AOB, user activities were categorised in Web 2.0 projects, based on dimensions which are suggested by Benkler (2002) for online collaboration (see section 2.2.6).

Among many other factors, SB has been suggested in the literature as a significant influencing factor for OB, and it has been highlighted by some of the previous researchers as an important factor in Web 2.0 platforms (see section 2.4.4). However, by reviewing the existing literature, a gap was identified in regards to use of SB in Web 2.0 projects (see section 2.4.5). The shortcomings are mainly associated with the *qualitative* nature of SB, and the fact that it cannot be investigated without interacting with users. As an attempt to fill this gap, Virtual Belonging (VB) was proposed as a new factor which can be investigated in Web 2.0 platforms. VB was defined and described as a *quantitative* factor which is built upon suggested elements for SB, and based on a new perspective towards Web 2.0 features (see section 2.4.6). Three dimensions have been suggested in this research for VB, namely *Inclusion*, *Involvement*, and *Influence*. Each dimension has been defined and described in this chapter together with practical examples in some of the existing Web 2.0 platforms (see section 2.4.6). The association between each dimension of VB and each dimension of OB is hypothesised and investigated in this research program, and will be described in the next chapters (see section 3.7).

In the next chapter, research paradigms are reviewed (see section 3.1), and some of the proposed psychological approaches towards investigating human behaviour are discussed (see section 3.2). This will be followed by further investigation of behaviourism as the adopted approach (see section 3.3). Thereafter, a research design is

proposed to investigate any associations between different dimensions of VB and different dimensions of OB in Web 2.0 projects (see section 3.6). This will be followed by outlining research hypotheses (see section 3.7).

3 Methodology and Hypotheses

In this chapter, initially *research paradigms* are described, and some of the paradigms which have been used by previous scholars for investigating OB are reviewed and compared (see section 3.1). Thereafter, major criticisms towards the chosen paradigm are described and their implications on this research are considered (see sections 3.1.5 and 3.1.6). This chapter is followed by a brief discussion about psychological approaches towards investigation of human behaviour (see section 3.2), and it is continued with a further investigation of behaviourism as the chosen approach (see section 3.3). The research design is then explained (see section 3.6), and finally the research hypotheses are outlined for this study (see section 3.7).

3.1 Research Paradigms

Understanding *research philosophy* has been suggested as an important step in any research (Guba and Lincoln, 1994; Saunders et al., 2007). This can guide the scientists to achieve their scientific aim and objectives. One of the important aims in science is to predict and control. In natural sciences, such as physics and chemistry, usually the effort is to find a causal link between quantifiable variables, and ideally present it as a mathematical formula. This formula then can be used to predict the results in future experiments. Although there might be different approaches towards finding the facts, it is usually taken for granted that there is a fact. The facts are seen as the objective truth about reality which can be discovered. These sciences which are closely linked to mathematics are sometimes called *hard sciences*, and on the other hand social sciences, are referred to as *soft sciences*. Social science is associated with human behaviour and human mind. Even if hypothetically an objective and causal link exits between independent variables and human behaviour, it would be difficult to define and measure these variables globally, independently, and objectively. For example it is rather easy to define and measure acceleration in physics, while it is much more complex to objectively define and measure creativity in social sciences.

Guba and Lincoln (1994) define paradigm as "the basic belief system or worldview that guides the investigator, not only in choices of method but in ontologically and epistemologically fundamental ways" (p.105). Similarly, Weaver and Olson (2006, p.460) define paradigms as "patterns of beliefs and practices that regulate enquiry within a discipline by providing lenses, frames, and processes through which

investigation is accomplished". Saunders et al. (2007, p.108) explain that research paradigms should be the basis for assumptions, strategies, and methods which are used in scientific investigations. Hence, it can be regarded as guidance throughout every stage of the study, including choosing the appropriate method, data collection and analyses, and last but not least the way the results are interpreted. Johnson et al., (2007) suggest such guidance can have a significant impact on research outcome.

Gage (1989) suggests a *paradigm war* among the existing paradigms. However, one can argue research paradigms present different *pathways* for investigation, and these different pathways may guide the investigator to the same findings. Hence, there might not be a right or wrong paradigm, only different ones. There are a variety of paradigms which have been suggested and implemented by previous researchers. Some of the popular paradigms include positivism, postpositivism, constructivism, interpretivism, realism, and pragmatism (Guba and Lincoln, 1994; Fitzgerald and Howcroft, 1998, Mackenzie and Knipe, 2006). It should be noted that one can observe overlaps on how these paradigms are defined and categorised. Guba and Lincoln (1994) categorise the competing paradigms in social sciences based on their ontological, epistemological and methodological questions. They suggest *Positivism*, *Postpositivism*, *Critical Theory*, and *Constructivism* as the main paradigms, and argue that in social sciences there has been a shift from objectivism to subjectivism. The most commonly used paradigms in researches which are associated with information technologies are positivism and interpretivism (Orlikowski and Baroudi, 1991, Fitzgerald and Howcroft, 1998, Walsham, 1995). Fitzgerald and Howcroft (1998) identify these as *hard* and *soft* research approaches respectively. In addition to these paradigms, Smith (2006) suggests *critical realism* can overcome some of the existing insufficiencies, and it can provide a better explanation for users' behaviour in information systems research.

3.1.1 Positivism

The society and human behaviour is complex and sometimes it might be difficult to analyse. However, positivism uses scientific principles to study the society and human behaviour (Webb, 1992). In other words, positivism is the scientific study of human behaviour, which usually involves quantitative measuring of different variables, and finding the relationship between these variables (Guba and Lincoln, 1994). Objectivism is one of the main principles in positivism, and it suggests empirical approaches to test the hypotheses. Mackenzie and Knipe (2006) argue the aim in this paradigm is to test the theory based on empiricism and rationalism. Positivism assumes the existence of

truth, and the possibility of realising it objectively and independent of human interpretation (Orlikowski and Baroudi, 1991). In such a paradigm, data collection and analyses should not be affected by researchers' beliefs and values (Niehaves, 2007). Furthermore, Krauss (2005) highlights the fact that in such approaches, the observation or data collection should not affect the data, and hence the researchers must distance themselves from the objects they study. Guba and Lincoln, (1994) also suggest in positivism researchers must be natural observers, and must not influence their research outcomes. Positivism assumes reality exists, and it can be known by rigorous empirical investigation (Creswell, 2007).

Karl Popper (1963) was among others who criticised positivism and its fundamental assumptions. Popper argues that "the historicist discovery that all standards are after all only historical facts, leads to the deification of facts" (Popper, 1963, p.346). He regards positivism and behaviourism as *secularised religions*, and argues that human behaviour includes verbal behaviour, and on this basis the facts are at least limited to language. He suggests the ultimate wisdom of philosophy in our time would be to "appeal to the logical and moral authority" of these facts (Popper, 1963, p.346). Postpositivism can be considered as the result of such criticisms towards positivism. Postpositivism shares its basic beliefs with positivism however it tries to respond to some of its critiques. One of the requirements in positivist research is the objective and scientific observation of the examined variables. However, postpositivism recognises the subjectivity of research, and aims to reduce it by multiplism and triangulation. Such techniques usually suggest if different methods and/or subjects lead to the same results, such results are more credible (Guba and Lincoln, 1994).

3.1.2 Interpretivism

In social sciences, interpretivism assumes the existence of truth, but denies the possibility of objectively knowing it (Creswell, 2007; Saunders et al., 2007). Interpretivism also suggests multiplism in social investigations; however it recognises the effects of interpretation and hence suggests use of many different interpretations for finding the truth. Furthermore, interpretivism differentiates investigation of humans and investigation of objects (Saunders et al., 2007), and in human investigations allows the development of subjective meanings (Creswell, 2007). Consequently, the researchers do not need to distance themselves from the research subjects, and may interact directly with the research participants. Mackenzie and Knipe (2006) highlight the effects of researchers' background and participants' perspectives. Interpretivism is usually

associated with qualitative data collection and inductive reasoning (Mackenzie and Knipe, 2006).

3.1.3 Realism

Realism shares some of its principles with positivism and interpretivism. It assumes the existence of reality independent of humans mind, but highlights the importance of humans' sense of truth (Krauss, 2005; Saunders et al., 2007). Saunders et al., (2007) suggested two types of realism; namely direct (or naive) realism and critical realism. Direct realism assumes humans' sensations of reality as the truth. On the other hand, Critical realism identifies these sensations as humans' perception of truth, but not necessarily the actual truth (Saunders et al., 2007). Krauss (2005) suggests that both qualitative and quantitative methods can be appropriate for investigations which follow realism. The term *Critical Theory* is used by Guba and Lincoln (1994) to cover a group of paradigms such as poststructuralism and postmodernism, which highlight the value-determined nature of investigation. Such paradigms recognise the subjectivity of reality, and the influence of investigators' beliefs and values on ultimate findings. While realists believe in mind-independent reality, there are other paradigms which suggest different versions of reality. For example, Guba and Lincoln (1994) suggest *Constructivism* as an alternative to earlier paradigms, in which truth is a belief system within a context. They consider *knowledge in a specific field* as a construct which is built by scientists to explain the natural world. In other words, Constructivism suggests that there are different versions of reality which exist in individuals' minds (Guba and Lincoln, 1994).

It should be noted that while any of these paradigms can be used to explain human behaviour, no particular paradigm can be seen as the only acceptable paradigm. For example OB can be investigated through any of these paradigms, and while there will be ontological and epistemological differences, the investigation can improve our understanding of human behaviour.

3.1.4 Basic Beliefs in Research Paradigms

Ontology is concerned with the "form and nature of reality". Positivism (Naive Realism) assumes an apprehendable reality exists, and the knowledge can be presented in form of time and context-free generalisation. Postpositivism (Critical Realism) also assumes a reality exists, however it suggests it is only imperfectly and probabilistically apprehendable. These paradigms argue that reality can never be perfectly apprehended, but critical examination of claims about reality can improve our understanding of the

reality. Critical Theory (Historical Realism) also assumes an apprehendable reality exists, but it suggests over time the reality has been shaped by social, political, cultural, economic, ethnic, and gender factors. Finally, Constructivism assumes local and specifically constructed realities (Guba and Lincoln, 1994; Krauss, 2005; Saunders et al., 2007).

Epistemology is concerned with "what can be known about reality". On this basis, positivism and postpositivism are objectivist. While positivism aims for finding the objective truth, postpositivism aims for finding the probable truth. In postpositivism objectivity is treated like an ideal, yet not always possible to achieve. Furthermore, in a postpositivistic view, findings are always subject to falsification. On the other hand, Critical Theory and Constructivism are both subjectivist, and focus on the role and values of investigator in the findings. In Critical Theory, the investigator and the investigated object influence each other, hence the findings are always associated with and affected by the interaction between a particular investigator and a particular object or group. Finally, Constructivism views knowledge as a construct, and suggests the findings are created as the investigation proceeds (Guba and Lincoln, 1994; Krauss, 2005; Saunders et al., 2007).

Methodology is the study of methods, and the ways which these methods can be best applied in an investigation (Guba and Lincoln, 1994; Saunders et al., 2007). In other words, when an investigator believes something can be known, methodology explains how they can get on with finding it. The answer to this question must be in line with ontological and epistemological views in each paradigm. In positivism questions or hypotheses are proposed based on theory and verified by empirical tests with appropriate controls to prevent subjectivity. Webb (1992) argues that for a positivist methodology only empirical quantitative measurements should be used, and the conclusion should be based on the identified relationships between measured variables. On the other hand, postpositivism suggests critical multiplism as a way of falsifying hypotheses (Popper, 1963). Falsifiability of a hypothesis is the possibility that it can be proven false. In such post positivistic paradigms hypotheses are not proven, and instead null hypotheses are rejected. In other words, if all observed swans are white, it increases the probability of all swans being white but it does not prove it. However, observing a single black swan can reject this as a truth. This is sometimes referred to as "Problem of Induction" which explains that no number of confirming observations can verify a universal generalisation. The possibility of highly improbable is sometime referred to as

Black Swan events (Taleb, 2007). In postpositivism emphasise on quantitative measurement has been reduced, and the effect of biased observations have been considered (Webb, 1992). It also recognises the importance of situational information, including the emic viewpoint (how local people think) in describing meanings and purposes associated with actions. While positivistic methodologies concentrate on objectivism, the methodologies in both Critical Theory and Constructivism highlight the interaction between the investigator and subjects, and the social and environmental factors which can influence the findings.

3.1.5 Received View and Critiques

Positivist and postpositivist paradigms as some times referred to as *Received Views* (Guba and Lincoln, 1994). This is based on the principle assumption of these paradigms in existence of *objective truth*. The investigator only receives this truth, and should be impartial to the truth. Guba and Lincoln (1994) explain several critiques towards these paradigms, and categorise them into *internal* and *external* critiques. In this section these critiques are discussed and their implications towards this research are considered.

3.1.5.1 Internal (Intraparadigm) Critiques

In quantitative approaches, usually a limited number of variables are examined. Other variables which exist in the context are usually stripped from consideration by using appropriate controls. Guba and Lincoln (1994) call this *context stripping* and argue this can jeopardise the generalisability of research. In their view this can only be improved by using qualitative data and providing contextual information. Moreover, they suggest that there is a disjunction between theories and local context. Hence, the research needs to be empowered by view of the studied individuals and contextual information. On this basis they once again argue theories need to be qualitatively grounded. They also highlight the complexity of human behaviour and suggest that meanings and purposes which are attached to the human behaviour can be better explained by qualitative data. Finally, they argue that statistical generalisations might not be applicable to individual cases, hence qualitative data is required to resolve the ambiguity. The internal critiques describe some of the problems embedded in the received view which can affect the objectivity and reliability of findings. However, most of these shortcomings can be improved by additional use of qualitative data. None of these critiques question the basic assumptions of these paradigms, and only suggest qualification of approaches.

3.1.5.2 External (Extraparadigm) Critiques

Unlike internal critiques, the external critiques towards received view challenge the basic assumptions of positivism and postpositivism. Guba and Lincoln (1994) challenge the received view's interpretation of facts, and point out the interdependency of theories and facts. They explain facts are facts only within some theoretical framework. They also point out the dependency of values and facts by suggesting facts are viewed through a value window, and rejecting the possibility of providing value-free theories in social sciences. They also explain "underdetermination of theory" (or problem of induction) as another philosophical flaw in received view. They argue while larger samples can improve confidence, there can always be a contradictory example (such as a black swan) which has not yet been observed. This is why philosophers such as Popper (1959) suggest theory falsification should replace theory verification. One of the principles in received view is that of independency of investigator. In other words, the investigator should not influence the phenomena or vice versa. Guba and Lincoln (1994) suggest that in social sciences findings are created through the interaction of investigator and the observed phenomenon (usually people). This contradicts the objective observation principle which is necessary in received view.

3.1.6 Received View and Online Behaviour

Two set of criticisms (internal and external) have been highlighted towards use of positivist and postpositivist approaches in social sciences (Guba and Lincoln, 1994; Saunders et al., 2007). However, while concentrating on OB some of these problems are resolved, and other ones can be considered as limitations which are ultimately associated with any investigation. The internal criticisms highlight the importance of contextual information and suggest the use of qualitative data for improving this shortcoming. However, both OB and VB are defined in this study as quantitative measure which can ideally be recorded without investigators' influence, and in Web 2.0 platforms it can be ultimately presented in quantitative form (0s and 1s). They also highlight the complexity of human behaviour and suggest only qualitative data can fully explain the behaviour. While this can be correct in principle, the proposed definition of OB is much less complex, and by focusing purely on observable behaviour it can be explained quantitatively and objectively. It should be noted that scope of this study does not include all aspects of human behaviour in association with Web 2.0 platform. It only focuses on the relationship between quantitatively defined representation of behaviour (see section 2.2.1), and quantitatively defined connections between users and the

associated platform (see section 2.4.6). As another criticism, Guba and Lincoln (1994) argue that statistical methods cannot explain the individual cases. Once again this is correct in principle, but from practical perspective, statistical results can be as important as understanding the individual cases.

The external criticisms towards received view raise questions about the fundamental assumptions of such paradigms. While Guba and Lincoln's (1994) argument about dependency of facts on theories and values is understandable in social sciences, in this study OB and VB, are presented quantitatively, and ideally can be defined and measured independently. They also explain the problem of induction and raise questions about generalisability of experimental findings. This research recognises this flaw and concentrates on falsification rather than verification (i.e. rejecting null hypotheses). Finally, Guba and Lincoln (1994) challenge the objectivity of investigation when it involves interaction between investigator and humans. However, while this study is associated with human behaviour, one can argue that the proposed quantitative measures (see sections 5.2 and 5.3) ideally can be investigated and analysed without affecting the investigated subjects.

3.2 Psychological Approaches

The British Psychological Society defines *psychology* as "the scientific study of people, the mind and behaviour" (BPS, 2014). Based on this definition, and considering the fact that this research aims to investigate users' online behaviour, different psychological approaches can be considered. From pure concentration on behaviour to analysing the brain functions, there are different psychological approaches (and underlining perspectives) to explain and predict the human behaviour. Some of these approaches are so different in principle that can be considered contradictory. However, this wide range of approaches can be justified by recognising the complexity of human behaviour and its causes.

The psychology of human behaviour can be looked at from a biological perspective. Biological psychology or behavioural neuroscience is the study of the biological reasons behind human behaviour (Rosenzweig et al., 1996). It is a very positivistic view of organisms which assume any behaviour has its root in the brain (or body). Although this view can be ultimately true, it would involve such a complex procedure and methods which is beyond the scope of this research. While biological psychology concentrates on brain functionalities, cognitive psychology takes them as granted and concerns itself

with mental processes such as perception or attention (Anderson, 1990). Alternatively, developmental psychology concerns itself with development of mind, and the way it is influenced (Bornstein and Lamb, 1999). Psychodynamic psychology concentrates on the development of mind, by trying to find a relationship between present and past, and also investigating conscious and unconscious motives (Jarvis, 2004). Freud, who can be considered as the father of psychoanalysis, compares human mind to an iceberg, in which conscious is the only visible part. Humanistic psychology concerns itself with understanding human behaviour through the eyes of the individuals (Schneider et al., 2014). Maslow can be considered as one of the earliest humanistic psychologists, and he explains human behaviour based on hierarchy of needs and motivations (Maslow, 1943).

3.3 Behaviourism

Unlike most of the approaches mentioned above, *behaviourist approach* concentrates primarily on observable behaviour. Some behaviourists accept the importance of internal factors, but describe them as complex and difficult to measure. Hence, they concentrate on external and measurable factors (Watson, 1913; Boakes, 1984). Some other more radical behaviourists claim that people (and animals) are controlled by their environment. They claim the human being is the result of what it has learnt from its environment. The learning process is viewed by behaviourists as classical conditioning (learning by association) or operant conditioning (learning from the consequence). Behaviourism is also concerned with the effect of environmental factors (Stimuli) on behaviour (Responses). It is a reductionist approach and works on the principle that behaviour can be reduced to "Learned S-R (Stimulus-Response) Units" (Skinner, 1974; Boakes, 1984). Although the use of behaviourist approach has been criticised (see section 3.3.2), the Web can bring a new platform to explore behaviourism. This research is trying to revisit some of the principle assumptions of behaviourism (see section 3.4).

Behaviourism can be seen as the result of positivism on psychology, and it was very popular among both psychologists and philosophers during the first decades of twentieth century (Smith, 1986). It can be argued that it started around a century ago when John Watson (1913) published "Psychology as the Behaviourist Views It.", which can be seen as the Behaviourism Manifesto. He states "Psychology as the behaviourist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behaviour." (Watson, 1913, p.158) Watson argues that

psychology is too concerned with the meanings, and he doubts the possibility of ever finding generally agreed statements which can describe simple human feelings, hence he suggests a psychology without terms such as consciousness, mind and imagery. Instead he argues psychology should concern itself with stimulus and response, habit formation, habit integrations, and simply more tangible terms. Watson (1913) claims that the problem with human psychology is its dependency on introspection which is not open to experimental treatment. As an alternative, he describes behaviourism as a purely objective and experimental branch of natural science, which does not require any introspection. Based on experimental study of animal behaviour and without considering consciousness, he argues that for a general understanding of human behaviour there is no need to involve consciousness. He claims the scientific control of behaviour can only be possible by investigating response range, and the determination of effective stimuli, habit formation, habits persistency, interference and reinforcement of habits. He insists on such an approach despite possible lack of generality in results.

Watson (1913) suggests a shift of focus from consciousness to behaviour, and his view is based on two main assumptions. Firstly, organisms (humans and animals) use hereditary and habit equipments to adjust themselves to their environment. Secondly, certain stimuli lead the organisms to make the responses. Based on his idea, when a system of psychology is completely worked out, the response can be predicted based on stimuli, and vice versa. He describes his method as a tool for obtaining an accurate knowledge of adjustments and the stimuli. He makes clear that his goal is not "description and explanation of states of consciousness", but instead "finding general and particular methods to control behaviour". The ideas presented by Watson were followed by different psychologists, such as Skinner. In 1977, Skinner published "Why I am not a cognitive psychologist", and described environment as the ultimate reason for human behaviour, and regarded cognitive psychology as misleading (Skinner, 1977).

3.3.1 Types of Behaviourism

There are different types of behaviourism, and they differ in their principles. *Methodological* behaviourism defines psychology as science of behaviour. It does not dismiss the influence of mental events and processes (internal), but it simply chooses to ignore them by concentrating on behaviour (external). On the other hand, *psychological* behaviourism claims behaviour can be described and explained without considering mental events and internal processes (Staats, 1996). It emphasises on the significant influence of environment, in comparison to mind. *Analytical* (or Radical) behaviourism

claims all internal processes and mental terms can be and should be described and translated into behavioural concepts (Chiesa, 1994). While Skinner suggests that organism cannot be treated as a black box, in his radical view he doubts the power of mind in changing reinforced behaviour, and his views have been sometimes regarded in contradiction with individuals' freewill.

Based on these definitions, the approach followed in this research can be identified as methodological behaviourism. In other words, this research does not concern itself with the mental events and processes, and instead it concentrates on the observable behaviour. It should be emphasised that this study does not disregard the existence and influence of internal factors in shaping OB, but their role in forming the behaviour is beyond the scope of this study.

3.3.2 Critiques of Behaviourism

Behaviourism has received major criticism during last decades (e.g. Stich, 1998; Chomsky, 1959), and has lost popularity among researchers. Its suggested laboratory routines have been criticised as too distant from natural and social environments, hence it has been suggested that the observed behaviour could be different from the behaviour in natural environment. Additionally, even if it is accepted that individuals' reinforcement history can have an effect on their behaviour, its superiority to internal factors cannot be investigated without recognising their existence and investigating their effects.

Another downfall with behaviourism is its inadequacy in fully explaining the experimental data without referencing mental terms. Stich (1998) suggests the experimental data cannot be explained without assuming an "information processing mechanism" within the organism. He argues behaviourism even lacks the power to fully explain features within this mechanism which are associated with qualities. Behaviourism has also been criticised by Chomsky (1959) for its lack of power to explain learning of language.

3.4 Online Behaviourism

The early findings about behaviourism were associated with experiments on animals in very restrictive and unnatural environments. Some of these experiments can be seen as though animals are treated as machines. A machine usually contains a processing system, a limited number of inputs, and a limited number of outputs. However, it can be

problematic to assume humans (or even animals) as a machine with a limited number of inputs and outputs. If they are viewed as a machine it could not be forgotten that there might be other inputs or outputs which we are not currently aware of. Also, there are a huge range of environmental factors which needs to be considered to enable behaviourism to fully explain the behaviour. However, the proposed definitions for OB and VB (see sections 2.2.1 and 2.4.6) present a quantitative perspective, and their association can be investigated without considering other factors. This should be noted that this investigation does not aim to paint a full picture of what might be considered as human behaviour in association with online platforms. This study investigates the relation between electronic connections (which can be measured quantitatively), and the digital signals which is received (and can be measured quantitatively) from humans through electronic tools and networks. Additionally, the digital nature of online world makes it easier to explain and compare the online environments.

Another benefit with online experiments is their similarity to non-experimental environments. It would be ideal for a behaviourist to run an experiment in which a large number of participants are willingly connected to a machine which records their every behavioural output. However, one of the challenges with such a machine would be not to affect the corresponding behaviour. Also, it should ideally work in natural environment, and record the environmental factors in real time. While such a machine might sound impossible in real world, it is easily achievable online. In other words, combination of tracking systems and actual online environment can ideally provide a platform for observing OB without influencing it.

While behaviourism may not be able to explain the full extent of human behaviour, *Online Behaviourism* can be built upon the principles of behaviourism and achieve its goals in a different context. Watson (1913) suggests the theoretical goal of behaviourism is to predict and control the behaviour. While this has not been achieved in its full extent through behaviourist approaches, Online Behaviourism can be able to fulfil this goal. In fact it can be argued that Online Behaviourism has already been tried and proven successful by giant online companies such as Google and Facebook. Google search results, or Facebook's News Feed can only be appealing to their audience if these companies understand their audience, and this is partially achieved by analysing their OB. One might suggest an online user can be seen as one of *Skinner's birds* in a cage which is presented with a limited number of buttons to click on. The power of Online Behaviourism would be to predict with a high level certainty which button would be

clicked. Finally, it should be emphasised that undoubtedly by interacting with real users, and collecting qualitative data about their online experience, many other aspects of behavioural motives can be discovered. However, this is beyond the scope of this study.

3.5 Online Environment

The *mentalist* approach in psychology highlights the importance of *internal* factors (such as perceptions, satisfactions, and expectations) while investigating human behaviour. On the other hand, in behaviourist approach the importance of *environment* is emphasised, for explaining human behaviour. For example, Watson suggests the individuals use hereditary and habit equipments to adjust themselves to their environment (Watson, 1913). More radical behaviourists such as Skinner (1974) explain behaviour as a function of present and historical environment.

Online environment (e.g. the website) and its attributes (e.g. visual and logical aspects of the website) can be considered as part of the overall environment which is described in behaviourism. It can also be argued that in regards to involving factors, the online environment is less complex than the natural environment. The website is part of the environment in which the OB takes place. Any website consists of the visual design of website and the logical functionalities which have been embedded within the underlying software. One can argue that websites are defined by their visual and logical aspects. While the visual differences can be easier to identify between different platforms, it can be argued that the logical differences (e.g. algorithms which are represented with code in the underlying software) differentiates the websites more fundamentally.

3.5.1 Visual Environment

The way web elements are placed in a website is sometimes referred to as *Website Design*. For example, each website's design is the combinations of its layout, fonts, colours, and so on. Website Design should ideally combine art and technology with human psychology to prepare and optimise the visual environment (Webesignerdepot, 2010). There are different factors such as Ease of Use, Entertainment Factor and Information Overload which have been suggested in the literature to have an influence on behaviour (Papacharissi and Rubin, 2000; Davis, 1989; Barua et al., 1995), and it can be argued such factors have their roots in Visual Environment of each website.

3.5.2 Logical (Non-Visual) Environment

In each website, a *logical* structure dictates the boundaries and attributes for its users' OB. These set of logics can be seen as the rules for users' interaction with underlying software. Also, such rules (or logical responses from the software), create the culture and conventions for the associated platform and its users. These rules might be simple or complex, and also they can be known or unknown to the individuals which are bound by these logics. For example, the Facebook's famous *Like Button*, has a particular colour, font size, and all together a visual design. This together with the details of every other element and visual layout which exists on Facebook create the visual environment for Facebook. However, the Like Button also represents a set of logical and mathematical functions. When an individual clicks on the Like Button these logical functions are triggered. For instance clicking on a Like Button shapes the updates and notifications which are received from Facebook in the future. However, the actual logics (or algorithms) which are used to filter and sort the updates are not necessarily known by Facebook users. These logical rules are simply dictated by the website, and when an individual chooses to use a website, by default they are bound by its logical environment as well as its visual environment.

3.6 Study Design

The aim of this research is to investigate OB in Web 2.0 projects, and to present a quantitative model in its association (see section 1.2). In chapter 2, VB was proposed (see section 2.4.6) as a quantitative factor in Web 2.0 projects, and based on a gap which was identified in the literature (see section 2.4.5). After reviewing some of the suggested research paradigms (see section 3.1), *postpositivism* was chosen to guide this study. On this basis, *behaviourist* approach and *quantitative* method were chosen for this research. As discussed earlier (see section 3.1.6), with proposed definitions for OB and VB some of the major criticism towards quantitative and positivistic approaches can be defended.

Following a postpositivistic paradigm and a behaviourist approach, *experimental* investigation is appropriate. For such an approach, experimental data with high level of control over the environment and research participants was required. Consequently, it was decided to design and implement a web based experiment, and collect behavioural data through a bespoke tracking system. The experiment involved technical and nontechnical steps which are described later (see chapter 4). The data which is collected

from this experiment has been used to investigate the research hypotheses, and rejecting associated null hypotheses (see chapter 5).

3.6.1 Sampling

When investigating a population, it might not be possible or required to investigate one and every individual in that population. In such cases, usually sampling techniques are implemented and a subset of the actual population is investigated. The aim is to choose and investigate a sample from the population and obtain information about the entire population. Fricker (2008) suggests two categories of methods for web and e-mail surveys; namely "probability-based sampling" and "non-probability sampling".

In probability-based sampling, the respondents are selected by probabilistic mechanisms. These mechanisms include "Simple Random Sampling", in which any two groups of equal size are equally likely to be selected, or less random techniques such as "Stratified Random Sampling" or "Cluster Sampling" in which the chosen samples must satisfy additional criteria (Fricker, 2008). On the other hand, non-probability sampling, is used when the probability of respondent inclusion cannot be determined, or if individuals have a choice to participate in the investigation. In such scenarios different techniques have been suggested, such as "Quota Sampling" in which desired number of respondents are specified, or "Judgement Sampling" in which the researcher selects the sample based on his or her judgement (Fricker, 2008). In non-probability sampling, rigorous investigations should minimise the number of non-participants. Fricker (2008) explains two types of bias which are associated with this type of sampling. Firstly, those who take part may not represent of the whole population. Secondly, because generally there is no information about the determinant of participation, magnitude of the bias cannot be assessed.

In this research, one of the requirements from the Ethics Committees was for users to give prior consent for taking part in this research. Consequently, probability-based sampling was not an option. Based on the requirements of the Ethics Committees, and the scope of undertaken experiment (see chapter 4), non-probability sampling is used in this research. To obtain the investigated sample and increase participation rate, every visitor of the website was presented with a form to verify their age, and also the option to take part in this research. The raw data from the experiment shows that this form has been submitted 9,162 times from which 5,843 instances are associated with individuals who have agreed to take part in this research (around 64% participation rate). In other

words, from each 3 individuals in the actual population, 2 of them are in the investigated sample. Although both types of bias can still exist, considering the size of sample and rate of participation, the findings about the sample have been generalised for the entire population, and any bias resulting from this technique is considered as limitation for this research (see section 6.4).

3.7 Research Hypotheses

In this section the research hypotheses are outlined. These will be investigated later (see chapter 5) by investigating probability of null hypotheses. This research follows a behaviourist approach, and as explained earlier (see section 3.3), environment has been highlighted by behaviourism as an important factor in investigating behaviour. Hence, in this section initially the effects of Online Environment on OB are hypothesised. This study proposes a new perspective for classifying and comparing users' connections to Web 2.0 platforms. Such connections are presented in this study as VB, and three dimensions have been identified for it; *Inclusion*, *Involvement*, and *Influence*. Additionally, in this research OB is categorised into POB (e.g. browsing the website) and AOB (e.g. sending content). On this basis, the associations between different dimensions of VB with each dimension of OB are hypothesised.

3.7.1 Online Environment

As explained earlier (see section 3.5), Online Environment is defined in this research as the combination of *visual* and *logical* aspects of online platforms, which are presented to the users through web browser. The following hypotheses are assumed in this research in regards to association between Online Environment and each dimension of OB:

- H1. Online Environment is significantly associated with POB.
- H2. Online Environment is significantly associated with AOB.

While visual differences among websites are usually easy to spot, distinguishing logical differences can be much more complex. For example, while Facebook users would be able to see and recognise change of colours or fonts on the website, it is more difficult to realise if there are any changes in the logic which sorts their News Feed. In this research two different stages (see section 4.8) of the experiment have been used to investigate these hypotheses (see section 5.5).

3.7.2 Inclusion

As explained earlier (see section 2.4.6), the logical connection between users and Web 2.0 platforms are described by in this research as VB, and one of the proposed dimensions for it is Inclusion (see section 2.4.6.1). The following hypotheses are assumed in this research in regards to association between Inclusion dimension of VB and each dimension of OB:

- H3. Inclusion dimension of VB is significantly associated with POB.
- H4. Inclusion dimension of VB is significantly associated with AOB.

In many online platforms active participation is exclusive to members, and membership is usually achieved by registration. In some Web 2.0 applications (e.g. Wikipedia) registration is not always required for active participation and instead the IP Address can be used for identification. However, online registration is still one of the most accepted and widely used forms of becoming an active participant. Furthermore, it is often required for the users to verify their email address by clicking on a link or entering a verification code which has been emailed to them. In this research Inclusion is measured based on users' access levels in an exclusive Web 2.0 platform (e.g. unregistered, registered, and verified). It should be noted that the access exclusivity is in regards to use of Web 2.0 features and active participation. Use of such functionalities requires providing an official email address from the University of Manchester. Other features such as browsing the website and consuming the available content are publicly available. In this experiment, three different levels have been assumed for users' Inclusion. The lowest level is associated with the individuals who have not gone through registration process. This can be either because they do not have access to a valid Manchester University Email, or just because they are not interested. The second level of Inclusion is assumed for the individual who have gone through registration, and have not verified their email address. The highest level of Inclusion is assumed for the individuals who have registered and verified their email address. The users are discriminated based on their level of Inclusion, and their OB is compared.

3.7.3 Involvement

In a Web 2.0 platform, while in many cases registration can be free and simple, for some users it still can be a barrier, and they might never register to become a member. Furthermore, access to some platforms might be exclusive and the membership can only be available to a specific group of people. However, after registration the next barrier

would be to get actively involved, and some members in an online platform might never pass this second barrier. As explained earlier (see section 2.4.6.2), in this research *Involvement* dimension of VB is assumed for users who have achieved *Inclusion* and have also used any of the available Web 2.0 features, to actively participate in the platform. Such activities create logical connections between users and online platforms. The following hypotheses are assumed in this research in regards to association between Involvement dimension of VB and each dimension of OB:

- H5. Involvement dimension of VB is significantly associated with POB.
- H6. Involvement dimension of VB is significantly associated with AOB.

As explained earlier (see section 2.2.5), a variety of different methods are used by Web 2.0 platforms to facilitate active participation. For example, in this experiment, the registered users are able to *Send* content in form of *text* (poem or comment) and *Link* (to associated audio version of their poem). They can also actively participate by *Evaluating* (voting) or *Sharing* the available content.

3.7.4 Influence

In any Web 2.0 platform, when the users pass the first two barriers, and achieve Inclusion and Involvement, they may or may not affect other users' OB. For example, when a piece of content is posted by a user, it may or may not receive votes or comments from other users. If it does, a new type of connection can be imagined between the user who posted the content and the associated platform. While different representation of such connections can be observed in different Web 2.0 platforms, they all are quantitatively stored in some sort of database as a link between the original user and the associated platform through the feedback from other users. Such logical connections are presented in this research as *Influence* dimension of VB. The following hypotheses are assumed in this research in regards to association between Influence dimension of VB and each dimension of OB:

- H7. Influence dimension of VB is significantly associated with POB.
- H8. Influence dimension of VB is significantly associated with AOB.

There are a various methods and representations for Web 2.0 features, and each online platform utilises their own representations. In the experiment which is associated with this research, Influence dimension of VB is assumed for the users who have actively participated in the platform (by sending poems and audio links), and have received

feedback (votes, comments, Facebook Likes). Furthermore, Influence is investigated in the experiment as both *Nominal* and *Scale* measure. In other words, in addition to comparing individuals who *have* and *have not* received comments, number of received comments have also been analysed.

3.8 Summary

In this chapter, initially research paradigms were described, and some of the relevant paradigms for investigating OB were identified and compared (see section 3.1). This was followed by describing ontological and epistemological aspects of research paradigms (see section 3.1.4), and the important role these basic beliefs can have in guiding scientific investigations. Thereafter, internal and external criticisms towards Received View were described (see section 3.1.5), and their implications for this research were considered (see section 3.1.6). The chapter was followed by describing some of the psychological approaches towards investigation of human behaviour (see section 3.2), and it was continued by reviewing behaviourism as the chosen approach (see section 3.3). Based on the aim of this research (see section 1.2), and the reviewed approaches and methods, the research design was described for this study (see section 3.6), and the research hypotheses were outlined (see section 3.7).

In the next chapter, the technical and non-technical steps which were taken to run the experiment (which is used for investigating the research hypotheses) are described. This will be followed by describing the steps which were taken to collect and process data.

4 Experiment Desgin

To test the research hypotheses, an experiment was designed and implemented. This experiment is shaped around a poetry competition in the University of Manchester. It is a competition for students in the Faculty of Engineering and Physical Sciences (EPS), and it started in 2012. The competition is organised by Dr Peter Fenn, and it is motivated by *The Two Cultures* lecture by C. P. Snow, in which he criticised the gap between sciences and humanities (Snow, 2012). The competition is open to English poems written in any form and by any student in EPS. The poems are expected to show the creative and human aspects of science and engineering, and the final winner is chosen by renowned poets Wendy Cope and Lachlan Mackinnon. This competition has been used as vehicle for an experiment which is described in this chapter.

4.1 Vehicle for Experiment

In 2012, the competition started, and EPS students received an email inviting them to take part in this competition and submit their poems by email. As a result 164 poems were received, and they were judged by a panel of judges. The competition's winners were later announced in a ceremony, and received cash prizes. Following the success of the first year of competition, the author suggested to use this competition as the vehicle for an experiment for investigating OB. The organisers accepted the proposal, and an online platform was developed by the author.

In 2013, and for the second year of this competition, the implemented platform was used for submitting the poems. This time the students received an email encouraging them to take part in the completion and submit their poems via the website available at *epsPoetry.com*. Consequently, 171 poems were submitted, and this time the submitted poems were publicly available and anyone could read them via the Internet. The data available from the website shows during the first 150 days of competition the "Poem Page" was requested 17,608 times. Although a small prize (£50) was devoted to the most popular poem, the final winners were still chosen by the panel of judges. In 2014 for the second time the online platform was used for submitting the poems. However, this time the website was redeveloped to improve engagement. Also, the competition rules were altered, and the users were given the possibility to improve their chance of winning the competition by engaging others, or adding an audio version to their poems.

This time 163 poems were received and the "Poem Page" was requested 41,222 times during the first 150 days of competition.

4.2 Experiment Scope

The experiment is designed to differentiate and classify users based on their connections to the platform. Such connections are represented in this research as different dimensions of VB. Also, users' OB is collected and stored based on quantitative data which is obtained from an online tracking system (see section 4.6). A quantitative approach has been used both in data collection and data analyses. Consequently, this research presents a quantitative perspective towards VB and OB in Web 2.0 platforms, and it investigates their statistical association. The factors which are associated with the users' VB have been represented with quantitative and traceable factors, and have been highlighted in the analysis results. Also, OB is tracked and recorded in real time and in a quantitative form by utilising a bespoke tracking system. This system is designed to store a wide range of behavioural variables, and will be discussed later in this chapter (see section 4.6).

The *mental* or *psychological* factors which are associated with OB can be identified and investigated, but this is beyond the scope of this research. However, in Web 2.0 projects, users' logical connections to associated platforms are investigated in this research, without considering psychological factors. To reject the null hypotheses, mathematical associations and correlations between such logical connections (VB) and the different dimensions of OB is statistically analysed. For instance, participating in a poetry competition can have enormous number of deciding factors, such as the ability to write a poem, the confidence to allow people to read it, and the knowhow to use the system and submit a poem. These among so many other factors have their roots in the conscious and unconscious mind of the individuals who take part in this competition and participate in this study. However, in this research the individuals' courage and state of mind is reduced to and represented by an independent Boolean variable (HAS.POEM). The association between this Boolean variable and OB is then statistically analysed, without considering the associated feelings or state of mind.

4.3 Subject Group

Although the competition is targeting the EPS students, this experiment looks at a broader audience. As explained earlier, the website was publicly available, and the

poems could be seen by anyone via the Internet. Every visitor of the website was invited to take part in this research, and their role in the research was described in an online "Participant Information Sheet" (see Appendix 1). The users were required to give prior consent by ticking a box. The only exclusion criteria were the age of participants, and their unwillingness to take part in the research. The minimum age of participants is 16 years, but it should be noted this is only based on users' declaration, and the investigator cannot independently verify the participants' age.

4.4 Ethical Approval

Any research within the University of Manchester which is associated with human beings is expected to receive Ethical Approval, and online research about human behaviour is not an exception. So, the first step before using the implemented Web 2.0 application was to obtain Ethical Approval from the University's "Committees on the Ethics of Research on Human Beings". Hence, a detailed specification was prepared and the application form for Ethical Approval was completed and submitted in January 2013 (Ethics Application 12377). Also the committees' meeting was attended in February 2013. In this meeting, the Web 2.0 platform was presented, and the committees' questions were answered. Finally, the committees' requirements were met and the ethical approval was granted. According to this approval, for users who have given prior consent, the data collected from the platform between 1/3/2013 and 31/5/2013 can be used for this research. The application was later extended to cover the data collected till 31/7/2014. The "Participant Information Sheet" is provided in Appendix 1.

Online privacy and the ethical debates about tracking OB is a popular and controversial topic at the moment, and the academic community is using the Ethics Committees within the universities to monitor and regulate the online research. Although the procedure was reasonable and straightforward, the author believes the current system can be improved. For example, medical research has a lot of additional regulatory and monitory steps, some involving National Health Service (NHS), and it seems like the investigation and research protocols are well established. However, one might suggest online research has not yet found its ethical position, and it can be argued that the current system lacks the technical dimension which should be in place when technical researches are ethically investigated. Although one can argue this research among many other online researches are benign and the current restrictions from the ethics committee

are too restrictive, there can still be potentially harmful online researches passing through the current system purely because of their technical complexity.

4.5 Online Platform

The Web 2.0 application which is used as the experiment is located on a Windows machine within the University of Manchester Network, which acts as a server, and runs XAMPP software to utilise a server-side scripting language (PHP) and database (MySQL). Like many other Web 2.0 applications, the technical core of this experiment in regards to data is implemented with an online database. In this case MySQL was chosen due to the fact that it is an open source database and can be used on a server without the costs which are associated with some other databases such as Microsoft SQL Server. The database is designed in a tabulated format. In each table different fields were defined to store data, some of which are only associated with operation of the website and running of the poetry competition. On the other hand, some of the available database tables and the associated field have been specifically designed for this investigation, and used to store a quantitative representation of users' OB. PHP is an open source server-side scripting language. It simply acts as a bridge between the end-user and the database. In this research PHP was used for both database communication and form validation. Furthermore, JavaScript was used as the client-side scripting language, and it was mainly employed to improve usability. All validations and database communications were placed on the server to insure the communication security. Cascading Style Sheets (CSS) was used for styling the web pages. Use of CSS facilitates separation of design and the core functionalities.

One of the unexpected yet very time consuming technical challenges during the experiment was automated emails. To meet the expectations of the Ethics Committees, it is required to use the University of Manchester's computer network. While this network allows PhD students, who are given a static IP Address, to run internet servers (such as the one which has been implemented for this study), there are restrictions in place which stops students from using email protocols. One can only assume such limitations are in place to avoid spamming. However, as a result of the existing regulations and the expectations from the Ethics Committees, all verification emails had to be sent manually via outlook. This also imposed a delay between registration and receiving the verification email. The University's IT Support was contacted several times in this regards, but unfortunately they could not make an exception for this

experiment. While the University's Ethics Committees insist on using the university's internal infrastructure for online research, experience in this regards shows there is little support from the University's IT support.

4.6 Tracking System

Following a behaviourist approach, and to investigate the association between VB and OB, it was required to collect a wide array of behavioural data. There are different methods which are currently used in the industry to obtain such a data. Using *Server Logs* is one the basic available methods to track users' OB (Ingram, 1999). Server Logs are text based files which are automatically created on internet servers, to log the requests which are received by the server. It is traditionally used to monitor the server activities, and contain the minimum required information (Liu and Kešelj, 2007). Also, based on browser settings, some of the files might be locally cached and these cached viewings are not logged on the server. Although Server Logs have been previously used by some researchers to analyse OB especially in eCommerce settings (Kwan et al., 2005), due to its limited capability for tracking AOB, it was decided that this technique is not adequate for a thorough investigation of OB. An alternative and more sophisticated technique for obtaining behavioural data is to use of Google Analytics (Clifton, 2012; Google, n.d.) or other similar applications which provide statistical and demographic reports for web managers. These systems are primarily designed for eCommerce websites, and produce different type of reports on websites' traffic, visitors, demographics and referrers (Google, n.d.). Additionally, these tools can be used to optimise online marketing campaigns by combining behavioural data and advertising costs (e.g. Google AdWords).

4.6.1 Internet Cookies and PHP Sessions

Internet Cookies (ICs) are small pieces of information, which are stored on local machine of the client. This helps in identifying requests from the same clients, and has also been used for identifying returned visitors. In simple words, ICs can be considered as notes which server leaves on the browser. For example, when a user visits a website, they can be given a unique identifier, and this identifier can be stored on both server (i.e. in the database). Additionally, this number can be written to an IC and stored locally on the client machine. So, next time this user visits the website, this IC still exists and will be sent to the server containing the unique identifier from previous visit. This identifier can be checked in the database, and previous activities can be retrieved.

Most web browsers automatically accept ICs, but users usually can prevent this if they so wish. However, in some cases this may affect how the websites work. ICs are specific to the server that have created them and cannot be accessed by other servers. It should be noted that ICs identify users' browsers, and they cannot be used to access personal information or passwords.

PHP Session is a technical term, associated with a group of requests to the server from the same browser. It is based on ICs, and has been used for simplifying the identification process. In other words, when a PHP Session starts, it is given a unique identifier, and this identifier is stored locally as an IC. In this experiment PHP Sessions are used in addition to ICs. Furthermore, in this experiment each PHP Session is stored in the database with a unique identifier and technical details (e.g. IP Adress, referrer, etc.) about the client.

Although use of ICs is a common practice for tracking OB, there are some limitations which are associated with it. These limitations are mainly due to the fact that ICs are not stored on the server, and hence users have control over it. The following scenarios outline examples of limitation in identification of unique visitors based on tracking data and ICs:

- Using multiple devices by the same user (e.g. Laptop, PC and Smart Phone)
- Using multiple browsers (e.g. FireFox, Internet Explorer)
- Disabling or removing ICs in the browser
- Using incognito browsing

In addition to tracking users' behaviour, ICs are also used for finding their browsing habits. Such information can be valuable for marketing purposes. However, privacy issues resulting from the growing use of ICs have led to new laws and legislations. For example, since May 2012 all UK websites must obtain "informed consent" from visitors before saving ICs on a device (ICO, 2011). This was also requested by the University's Ethics Committees, and hence in this experiment ICs are not stored before users' consent. It should be noted that while in regards to use of ICs, prior consent has been expected from this study by the Ethics Committees, the University's own official website is yet to comply with new regulations (last checked in January 2016).

4.6.2 Event-Based Behavioural Tracking

While reports generated by third party tracking systems can be very useful for web managers, they present some problems and shortcomings for this research, and hence it was decided to develop a bespoke tracking system. The first problem with third party tracking systems is lack of detailed event reports, while such information can be argued as an important factor in categorising the users. For example, it is important to know how many people visited the "Registration Page", but one can argue it is even more important to track the way it has been used. An "event" in this case would be "to submit a form" or "to receive an error", and due to different technical specification of each website, it is difficult for third part systems to track the internal events. To address this problem, a detailed list of events was produced and embedded in the bespoke tracking system. The other problem with third party tracking systems is the limitation to pre-produced reports rather than detailed behavioural logs, and hence the focus on accumulative reports, and averages rather than user-centric and user-specific reports. This was also addressed in the bespoke tracking system by use of ICs and session variables (see section 4.6.1) to base the tracking logs around each visit and each user. Additional mathematical algorithms were used to detect access from different devices and represent an accurate behavioural profile of the participants. Last but not least of the problems with the third party tracking systems is their incapability to meet the expectations of the University's Ethics Committees. The inadequacy of such systems is due to lack of control over data, and the conflicts which exist between the expectations of Ethics Committees and the terms of use and privacy policy of such systems. This was addressed and resolved in the bespoke tracking system, by implementing it on a web server within the University Network to provide full control over data.

4.7 Publicising the Platform

In regards to this experiment, one of the questions by the University's Ethics Committees was the sample size. It is a common practice for these committees to authorise the number of participants in investigations which are directly linked to humans. However, this experiment utilises an online platform, and anyone can access the platform through the Internet, and take part in the research if they so wish. The total number of expected visitors, and the proportion of them who would take part in this study could not be determined. Hence, the number of research participants (sample size) could not be determined prior to running the experiment. The Ethics Committees were satisfied with this justification, and they agreed not to put a limit on number of research

participants. Consequently, any number of participants was allowed as far as they had given prior consent to take part in this research.

The website was popularised through different methods. The first and most effective method for attracting the traffic was the invitation emails and the follow up reminders which were sent to the three main email lists within EPS. These email lists target undergraduate students, postgraduate students and staff who currently study or work in EPS. Additionally, for each competition around 200 posters were printed in different sizes, and these were attached to notice boards in busy buildings within EPS, and some of the common buildings (such as the Students' Union) and some of the libraries, and computer clusters. Also, the University's advertising screens were used to publicise the competition, and social media was employed to attract traffic to the site. In 2013 this was solely done by using prominent "Share Buttons" to Facebook, Twitter and Google Plus. However, based on the usage pattern during the first year, it was obvious that Facebook has a much higher potential in increasing the traffic. Hence, in 2014 the utilisation of Facebook was increased by using its famous "Like Button" next to the "Vote Button" for each poem. Additionally, in 2014 the competition was publicised in the University's Online News Bulletin, and all poems were shared in the associated Facebook Fan Page with around 290 followers.

Based on the data which is accumulated in the database, in total there have been more than 16,000 visits to the website. The associated OB is represented in the database with more than 240,000 records each associated with a *timestamp* and additional information about the activity which has been recorded.

4.8 Experiment Stages

As explained earlier (see section 4.5), an experiment was used in this study to investigate OB. To run the experiment, a Web 2.0 platform was developed and publicised in association with a poetry competition in the University of Manchester. Although the competition was exclusive to the students of the Faculty of Engineering and Physical Sciences (EPS), the website and its content were publicly available on the Internet. The website has been online since February 2013, and it is still available at "epsPoetry.com". Furthermore, two different implementations of the website were used in 2013 and 2014. These versions were both visibly (e.g. layout, colours, button, etc.) and logically (e.g. Web 2.0 features, rules, validations, verifications, conventions, etc.)

different. This difference has been used in this study to investigate the association between OB and Online Environment.

It is important to note that environmental factors are not quantified in this research. The visual environment is described verbally, and screenshots have been used for illustration. The logical environment has also been described verbally, and in some instances it has been further explained mathematically. The environmental factors have been investigated in this research by comparing OB in different Online Environments. It should be noted that dedicated cash prize for competition's winner was the same in the two competitions which are associated with each stage of the experiment. For the second stage, the website was redesigned, and the procedure for choosing the winner was also changed. These have been considered as differences in Visual and Logical Environment, and their association with OB has been investigated. It should also be noted that the subject group for different stages of the experiment is not exactly the same. These samples are associated with poetry competition in 2013 and 2014. However, considering the fact that students and staff who leave the EPS are usually replaced by newcomers, it has been assumed there is not a significant difference between the individuals who received the invitation emails in 2013 and 2014. Hence, as explained earlier, the environmental factors are accounted for the differences in OB.

4.8.1 Stage 1: EPS Poetry 2013

The empirical data which was gathered during the first stage of the experiment includes the online behavioural data from 1,851 participants. They all have given prior consent to be part of this research. The visual environment included a black and white design with old fashioned fonts, and a gray picture of Sackville Street Building as the header (see section 5.5.2). While some of the common Web 2.0 features (e.g. sending content, commenting, voting, and sharing) were implemented in the platform, the logical environment was distant from Web 2.0 principles. In a sense that winner was chosen without considering users' feedback (or poems' popularity). The judging process was accomplished by utilising two layers of judging by experts. There was no emphasis on user opinion rather than a £50 cash prize for the most voted poem. The number of votes was hidden and any registered user can vote for as many poems as they wanted. An official university email was required for registration; however email verification was not required for voting. In other words, the users could have registered with a non-existing email address (e.g. dummy123@manchester.ac.uk), and still use the voting feature. However, email verification was required for submitting poems and sending

comments. An email address was considered as official university email, if it finished with ".manchester.ac.uk" or ".man.ac.uk". The competitors were also invited to send an audio version of their poem, but only one person was willing to do so, and even they changed their mind later in the progress, and requested their poem's audio version to be removed from the platform.

4.8.2 Stage 2: EPS Poetry 2014

The empirical data which was gathered during the second stage of the experiment includes the online behavioural data from 4,016 participants. They all have given prior consent to be part of this research. The visual environment included a colourful design with an image of the spectrum in the website's header. Although the winner's cash prize was the same as previous years (£500), the rules were modified. A pointing system was introduced which allowed users to take part in judging process. In short, each poem could have a higher chance of bypassing the first layer of judgement by receiving more votes or views. Submitting the audio version also was possible, and the rules stated that it would provide three additional points. Unlike the previous year, this feature was adopted by some of the users, and 58 poems were linked to their audio version. The rules about points were slightly complex, and it was designed in a way that achieving a high number of points was not required for winning the competition. In other words, the first judging panel could add worthy poems to the shortlist despite their achieved points or number of votes. Also, like the previous year a cash prize (£100) was allocated to the poem with the highest number of points. Another complexity (or ambiguity) in the rules was about the points which were achieved from the number of views. The rules stated that "up to 5 points can be achieved based on number of views", but it did not explain how it is calculated. In practice it was calculated based on a linear function of number of views.

4.9 Online Behaviour

OB is defined in this research based on what can be objectively observed and recorded from the activities which take place while users visit a website (see section 2.2.1). During the experiment, OB was recorded by a bespoke tracking system, and it is represented with a *series of events* for each user during each visit. Each record can be seen as a *Click* by the user, and it is associated with a *timestamp* showing the exact time in which the Click has happened. The tracking system also stores additional information which can be used to explain the events' consequence. For example, when the

"Registration Form" is submitted, it is recorded as an event and considered as OB. Furthermore, the result of this action (Success or Failure in registration) is recorded by tracking system in the database, and can be used to further explain the original event (Submitting the "Registration Form").

As explained earlier (see section 2.2.2), in this research two dimensions have been assumed for OB. Passive Online Behaviour (POB) is mainly associated with browsing the web pages. Such activities are recorded for all research participants. On the other hand, Active Online Behaviour (AOB) can have different representations in each Web 2.0 platform. This can include sending or editing content, or voting and commenting and so on. It should be noted that not all Web 2.0 features are available in all Web 2.0 platforms. Furthermore, different representations can be found within different platforms. There are different dimensions which have been suggested for AOB. Although generating content is one of the most important dimensions of active participation (Koh et al., 2007), Web 2.0 applications also provide a platform for other forms of active participation. As explained earlier (see section 2.2.6), this research categorises active participations based on dimensions which are suggested by Benkler (2002). Benkler categorises online collaboration into "Content", "Relevance/Accreditation" and "Value-Added Distribution". In other words, the content which is produced collaboratively needs to be sorted or evaluated and finally virtually distributed. In a Web 2.0 platform, all of these could be done voluntarily by the platform's users. However, not all Web 2.0 platforms benefit from users' intention to use the provided features. Benkler's (2002) categorisation has been used in this study for categorising different forms of AOB for the participants of the research experiment.

The data which has been accumulated in the database represent OB in quantitative form, and this is for all the website users. However, according to the obtained Ethical Approval (see section 4.4), the data could only be used for visitors who have given prior consent to take part in this study. Hence, the investigated sample contains the OB of individuals who have given prior consent to be part of this research, and have used the website either actively or passively during the experiment. The active use of website in most cases requires going through the registration process, while passive use of website is open to the public. Also, to create an exclusive community, only the official university emails were accepted, and verification was required.

At the end of experiment, a copy of MySQL was exported and processed to produce the dataset which is required for this research's statistical investigation. There are some

restrictions which have been considered prior to analyses. The initial limitations are dictated by the Ethical Approval which is obtained for this research. According to this approval, only the data which has been collected between 1/3/2013 and 31/7/2014 and with prior consent from the users can be used and analysed for this research. Also, to remove potential bias from the findings, any behavioural data associated with website's administrators were identified to be removed from analyses. The administrators' role involves operating the website, approving submitted poems, and monitoring submitted comments. Additionally, the website was used to *rank* the poems by the first panel of judges, and their OB has also been removed. Finally, to satisfy the requirements of the obtained Ethical Approval (see section 4.4), any records from the users who had not confirmed their age or had not agreed to the privacy policy was excluded from analyses.

4.10 Virtual Belonging

As explained earlier (see section 2.4.4), the literature suggests SB as an influencing factors for OB in Web 2.0 platforms (Roberts, 1998; Teo et al., 2003; Lin, 2008; Kim and Zhang, 2011; Zhao et al., 2012). These suggestions are based on qualitative definitions which are associated with an *internal feeling* among the users. For example, Hagerty et al. (1992) define it as an "experience of personal involvement". Similarly, Zhao et al., (2012) define SB in OCs as "the feeling of being integral parts of the community". Different instruments have been suggested in the literature to measure this internal feeling (e.g. Hagerty and Patusky, 1995). However, due to its qualitative nature it seems somehow impossible to define and measure it quantitatively and objectively. Furthermore, based on these definitions one might argue the only way for measuring SB would be to interact with the involving individuals. As explained earlier (see section 2.4.5), in this regard this study identifies possible gaps in the existing literature. These shortcomings are mainly associated with the qualitative and subjective nature of SB, and the way it is measured. In early stages of Web 2.0 projects, in which the concept is shaping and the platform is yet to be implemented, the existence of users and their interaction with the platform can only be assumed. Hence, investigation of users' SB at early stages of Web 2.0 projects might not be an option, and the existing literature does not provide any alternative measures or practical suggestions for web managers. The suggested measures (see section 2.4.3) are subjective, and they need to be interpreted for each Web 2.0 platform. Such qualitative investigation might not be an option for small sized enterprises. Furthermore, due to factors such as "privacy concerns" and "lack of interest among the participants", the feedback obtained from online

questionnaires and so on can be biased. Consequently, scientific collection and analysis of such data might be too expensive for some of the managers. After all, some successful Web 2.0 platforms started in a university accommodation room or a garage (e.g. Facebook and Google), and there might not be any budget for extensive qualitative investigations. This research is based on a postpositivistic paradigm and ideally it aims for an objective observation. Hence, as explained earlier (see section 2.4.6), this study is proposing a new metric based on users' connections to Web 2.0 platforms. This metric is referred to as VB, and three dimensions are proposed for it. These dimensions are inspired by the elements which have been suggested in the literature as attributes of SB (see sections 2.4.2 and 2.4.3). It should be noted that while SB is usually defined and explained based on *internal feelings* and *perceptions*, this research does *not* concern itself with these feelings. For example, the empirical data which is used in this research is obtained from an online platform which is associated with *poetry*. Consequently, one may argue a personal (emotional) attachment can be assumed between a poet and their poem. However, nor this nor any other factors have been investigated from the *mentalistic* perspective. In other words, while any of the measured factors can be associated with a set of very complex and qualitative internal feelings, such feelings are not investigated, and are beyond the scope of this study. For example, the fact that during one stage of the experiment the users could increase their chance of winning the competition (even after submitting poems) can produce a completely different perception towards the competition. Furthermore, the visibility or invisibility of number of votes can be associated with individuals' confidence. Moreover, the effect of self-satisfaction can be influenced by a functionality which means the popular poems can be seen in the homepage and even affect the colour of corresponding school. While each of these internal factors can be qualitatively measured and investigated, it is beyond the scope of this research.

This study is concerned with *logical connections* between individuals and the platform. These connections differ from one platform to another, and are indentified in this research based on available Web 2.0 features for each platform. Such connections are referred to in this research as VB. In undertaken experiment, the users can be linked to the platform by registration, submitting poem, sending comments, voting and so on. These connections and their association with different dimensions of OB are investigated in the associated experiment. As explained earlier (see section 2.4.6), three dimensions have been assumed for VB. In the following sections, these dimensions are identified and explained in the corresponding experiment for this research.

4.10.1 Inclusion

As explained earlier (see section 2.4.6.1), the first dimension of VB is *Inclusion*, and it is associated with any connection which enable users to become active participants. In both stages of the experiment, registration is required for active participation. Furthermore, registration is only available to individuals who have an official email address from the University of Manchester, and hence active participation can be regarded as exclusive to these individuals. Furthermore, two different levels are identified for Inclusion in this experiment. The first level is obtained by going through registration process. At this stage, a verification code is emailed to the provided email address. The second level of Inclusion is assumed in the experiment only for individuals who have verified their access to the provided email address, by going through verification process. In the first stage of the experiment, first level of Inclusion was enough for using one of the available Web 2.0 features (voting). However, in the second stage of the experiment any active participation is only available to verified users.

4.10.2 Involvement

As explained earlier (see section 2.4.6.2), the second dimension of VB is *Involvement*, and it is associated with any connection which is built between the users and the platform as a result of utilisation of available Web 2.0 features. In this experiment different Web 2.0 features were provided. Furthermore, to represent a variety of these features, at least one Web 2.0 feature was implemented to represent each of the dimensions which are listed by Benkler (2002) for online collaboration. On this basis, the available Web 2.0 features in the experiment can be categorised in three categories: "Sending Content" (Poems, Comments, Audio Links), "Evaluating Content" (Voting), and "Distributing Content" (Sharing). The utilisation of each of these features, is represented as Involvement dimension of VB, and its association with OB is investigated.

4.10.3 Influence

As explained earlier (see section 2.4.6.3), the third and final dimension of VB is *Influence*, and it is associated with any connection which is built between the users and the platform as a result of other users' utilisation of available Web 2.0 features. In this experiment, the Influence dimension of VB is represented by connections which can be imagined between the users who have submitted their poems, and their poems have received any form of feedback from other users. These include receiving "Votes",

"Facebook Likes" or "Comments" from other users. These connections are represented quantitatively, and their relation with users' OB is investigated.

4.11 Experiment Data

In many web-based projects, and Web 2.0 projects in particular, the outcome of projects is dependent on OB. In many cases, if OB can be influenced, the project success factors can improve (see section 2.3). There are many factors which have been suggested by previous researchers to have an influence on OB. This research introduces VB as a new factor (see section 2.4.6) to investigate possible association between users' logical connections to the platform and their OB. For this investigation, an experiment has been designed and implemented (see section 4.5). Empirical data (representing users' OB) has been collected and accumulated for over 19 months, and in the next sections the steps which are taken to make this data operational are described. Initially data collection procedure is described in the experiment (see section 4.12). After data collection, the raw data has been prepared for statistical analyses. The general steps which are involved in such conversions are explained (see section 4.13), and it is followed by detailing the procedure which was used in this particular experiment for preparing the raw data (see section 4.14).

4.12 Data Collection

In this section, the data collection process in the experiment is described. As explained earlier (see section 4.1), the collected data is associated with a Poetry Competition among the students in the Faculty of Engineering and Physical Sciences (EPS) of the University of Manchester. A Web 2.0 platform was proposed, designed, and developed by the author, and in 2013 and 2014, the students were asked to use this platform to submit their poems (see section 4.5). Furthermore, the approved poems were available on the platform, and could be accessed by anyone through the Internet. Also, every visitor was asked in their first visit to participate in this research, and Ethical Approval was obtained from the University of Manchester's Ethics Committees to collect and analyse tracking data in association with users' OB (see section 4.4). It should be noted that in addition to required functionalities for submitting and reading the poems, some of the common Web 2.0 features such as commenting and voting were embedded into the platform. Moreover, various links were created between the platform and popular SNSs. This combination of features was designed to cover all proposed dimensions for VB, and facilitated the investigation of its association with different dimensions of OB.

4.12.1 Hypothetical Scenario (Janet's Scenario)

In this research a bespoke Tracking System has been used to collect information about the visitors and their OB. To explain the system, and the type of data which has been collected throughout this process, a typical scenario is considered and the corresponding collected data is explained. The following hypothetical scenario is used as an example and has been referred to in this thesis as Janet's Scenario:

In February 2014, a student of the University of Manchester checks her university emails and finds an invitation to take part in a poetry competition within the EPS Faculty for a chance to win £500. She clicks on the provided link and a web browser appears. She is immediately presented with a form to verify her age, accept the terms of use, and also to take part in this research. There are also links to "Terms and Conditions" and "Participation Information Sheet". She goes to the next stage by checking three checkboxes and submitting the form. On the next page she is presented with a simple homepage with a prominent button for registration. She clicks on the button and fills the registration form. However, after submission she receives an error indicating that only official university emails are acceptable. She changes her email address, finishes registration, and closes the browser. Later that day, she receives another email containing a verification code, and she is asked to verify her email address by entering this code. She follows the link, verifies her email address, and in the same visit she submits a poem to take part in the competition. For the sake of the argument, and to reference to this scenario later in this chapter, this student is called Janet, and this hypothetical scenario will be referred to in next sections as Janet's Scenario.

4.12.2 Database Structure

The above scenario has been used to describe the quantification process of OB in this research. However, this is only one hypothetical example among an unlimited number of imaginable scenarios. During the Janet's Scenario, different records are stored in MySQL database tables. These tables and the records which are logged during Janet's Scenario are explained in this section. The majority of records are stored in **`eventlogs`** table, which in less than two years have accumulated more than 250,000 records. These records are the raw data which is used to quantify visitors' OB. In this table each record is associated with a unique identifier (**`eventlogID`**), a timestamp (**`time`**), and an event identifier (**`eventID`**). Also, there are two other fields to identify and distinguish the visitors based on ICs (see section 4.6.1) and login history

(**viewerID** and **userID**). Finally, there are three extra fields to store the events' additional properties (**int1**, **int2**, **int3**). MySQL representation of **eventlogs** table is provided in Figure 1.

Field	Type	Collation	Attributes	Null	Default	Extra
eventlogID	int(10)		UNSIGNED	No	*None*	auto_increment
time	int(10)		UNSIGNED	No	*None*	
eventID	int(10)		UNSIGNED	No	*None*	
viewerID	int(10)		UNSIGNED	No	*None*	
userID	int(10)		UNSIGNED	No	*None*	
int1	int(10)		UNSIGNED	No	*None*	
int2	int(10)		UNSIGNED	No	0	
int3	int(10)		UNSIGNED	No	0	

Figure 1 - MySQL representation of "eventlogs" table

In Janet's Scenario, when she lands on the homepage of the website, 5 records are stored for her in **eventlogs** table (see Figure 2). There are also 2 more records stored in other tables which will be described later in this section. As explained earlier (see section 4.12.1), the invitation email which is sent to the Faculty's Email Lists contains a link to the website. This link is specifically created to assist with this research and it contains an identifier. Here is the link:

http://epspoetry.com/start

The "**/start**" part of the link is an identifier. Hence, as soon as the visitor is connected to the server, the first record of the visit is logged in the database. This record contains the time (e.g. Linux Timestamp) and a unique identifier associated with the invitation email (**eventID**: 67). Although technically visitor identifier information could have been collected at this stage, to follow the expectations from the Ethics Committees, no identifiers have been used for the visitor prior to their consent. The second record stored in Janet's Scenario is associated with visiting "Age Confirmation Page" (**eventID**: 19). As soon as Janet ticks the boxes and presses the button, she is redirected to the home page, and three more records are stored in **eventlogs** table; age confirmation (**eventID**: 21), participation agreement (**eventID**: 22), and home page view (**eventID**: 1). In Janet's Scenario, she has accepted to participate in this research, and hence a unique identifier for her as a participant is generated automatically (**participantID**), and it is stored as additional information. In this case, **int1** field of the record stores **participantID**. At this stage they have also accepted the use of ICs (see section 4.6.1, and Appendix 2), and hence a **PHP Session** is created. Also, a unique identifier is given to this session (called **viewerID**), and will be used to

identify and connect the next events by this visitor. The 5 records stored for Janet in **eventlogs** table up to arriving at Home Page are shown in Figure 2.

eventlogID	time	eventID	viewerID	userID	int1	int2	int3	eventID	desc
103587	1392886175	67	0	0	0	0	0	67	2014 First Email
103588	1392886176	19	0	0	0	0	0	19	Age Confirmation Page
103589	1392886179	21	0	0	0	0	0	21	Age Confirmation - Success
103590	1392886179	22	0	0	1893	0	0	22	Participation Agreed
103591	1392886179	1	5857	0	0	0	0	1	Home Page

Figure 2 - First records stored in "eventlogs" table in "Janet's Scenario"

The **viewerID** field represents an identifier for the visit, and it is the primary key for **viewers** table. The tracking system uses ICs to identify different events which occurred during the same visit. Figure 3 shows MySQL representation of the **viewers** table.

Field	Type	Collation	Attributes	Null	Default	Extra
viewerID	int(10)		UNSIGNED	No	*None*	auto_increment
time	int(10)		UNSIGNED	No	*None*	
vip	int(10)		UNSIGNED	No	*None*	
ref	text	utf8_unicode_ci		No	*None*	
agent	text	utf8_unicode_ci		No	*None*	
participantID	int(10)		UNSIGNED	No	0	

Figure 3 - MySQL representation of "viewers" table

The **viewerID** is used to create a sequence of events within each visit. The obtained Ethical Approval expects prior consent for the use of ICs, and hence ICs are not stored until the age is verified and the privacy policy is accepted (see section 4.6.1 and Appendix 2). To identify the sequence of page views and events which are associated with each visit, at the start of the visit a record is created in the **viewers** table, including the start time for this session and some additional technical information. For example, the visitor's IP Address is stored in the **vip** field as an equivalent integer. The page which has referred the visitor to the website (**ref**), and the browser's details (**agent**) are also stored. From this stage of the visit, any record which is stored in **eventlogs** table is associated with **viewerID**, and this is used throughout this research to identify distinguishable sequence of events by participants, to represent their OB in each visit (look at the last record in Figure 2).

In Janet's Scenario, in addition to age confirmation and privacy policy acceptance, Janet agrees to take part in this research by ticking the corresponding box. At this stage additional information is stored in the database, and also locally as ICs (see section 4.6.1). Initially, in **participants** table a new record is created for Janet as a participant. Figure 4 shows the MySQL representation of this table.

Field	Type	Collation	Attributes	Null	Default	Extra
participantID	int(10)		UNSIGNED	No	None	auto_increment
time	int(10)		UNSIGNED	No	None	

Figure 4 - MySQL representation of "participants" table

The unique identifier for Janet (**participantID**) is also recorded on their computer as an IC. Additionally this identifier is recorded in **viewers** table for her. This will help to identify her in future visits, and create a sequence of visits for her.

In Janet's Scenario, in her first visit she manages to register after one unsuccessful attempt. Consequently, there are 2 more records which are stored in **eventlogs** table for Janet's first visit. These records indicate the failed attempt for registration (**eventID**: 41) and the successful registration (**eventID**: 40).

After registration, the associated information is stored in another MySQL table. Such details are stored in **users** table, and its structure is displayed in Figure 5.

Field	Type	Collation	Attributes	Null	Default	Extra
userID	mediumint(8)		UNSIGNED	No	None	auto_increment
xuid	varchar(20)	utf8_unicode_ci		No	None	
time	int(10)		UNSIGNED	No	0	
isAdmin	tinyint(3)		UNSIGNED	No	0	
password	varchar(200)	utf8_unicode_ci		No	None	
email	varchar(200)	utf8_unicode_ci		No	None	
nickName	varchar(50)	utf8_unicode_ci		No	None	
isParticipant	tinyint(3)		UNSIGNED	No	0	
vrfTime	int(10)		UNSIGNED	No	0	
vrfCode	int(10)		UNSIGNED	No	0	
facultyID	tinyint(3)		UNSIGNED	No	None	
schoolID	tinyint(3)		UNSIGNED	No	None	

Figure 5 - MySQL representation of "users" table

In Janet's Scenario, after successful registration, a random number is created and stored in **users** table, to be used as their verification code. This code is emailed to Janet, and hence for verification in addition to providing an official university email, she also must have access to it.

4.12.3 Secondary Visits

So far, the first visit and the corresponding data collection procedures in Janet's Scenario have been described. To explain the procedure for creating a sequence of visits, the post registration process in Janet's Scenario is described in this section, and associated data collection procedures are explained.

After registration, Janet checks her emails and finds an email from EPS Poetry, asking her to verify her email address by entering a provided Verification Code. A link is

provided for verification, and she clicks on the link. A browser appears and she is asked to enter her email address, password, and the Verification Code. She enters the details, and successfully verifies her email address. As described in her hypothetical scenario, in this visit she continues by submitting her poem using the provided Web 2.0 features.

Form a technical perspective, as soon as Janet clicks on the link and connects to the server, her **`participantID`** (which is stored locally as an IC) will be sent to the server. Additionally, there is another IC associated with age confirmation and privacy policy acceptance which are received by the server. The **`participantID`** for Janet is previously stored (during her first visit) against the record which is created in the **`viewers`** table. On this basis, this new visit is linked in the database to her first visit, and any other visits which are made in the future from Janet. It should be noted that there are some technical details which can limit this type of identification. For example, for this to work Janet must use the same browser, and she must not remove ICs. The procedure for event based data collection continues, by recording a list of events for each of her visits to the website.

4.12.4 Database Tables

In addition to MySQL tables which were explained earlier, other tables have been used to store data for the application. These include tables for storing details of submitted poems, received comments and votes, and so on. The relationship between all the tables which are used by underlying software is represented in Figure 6.

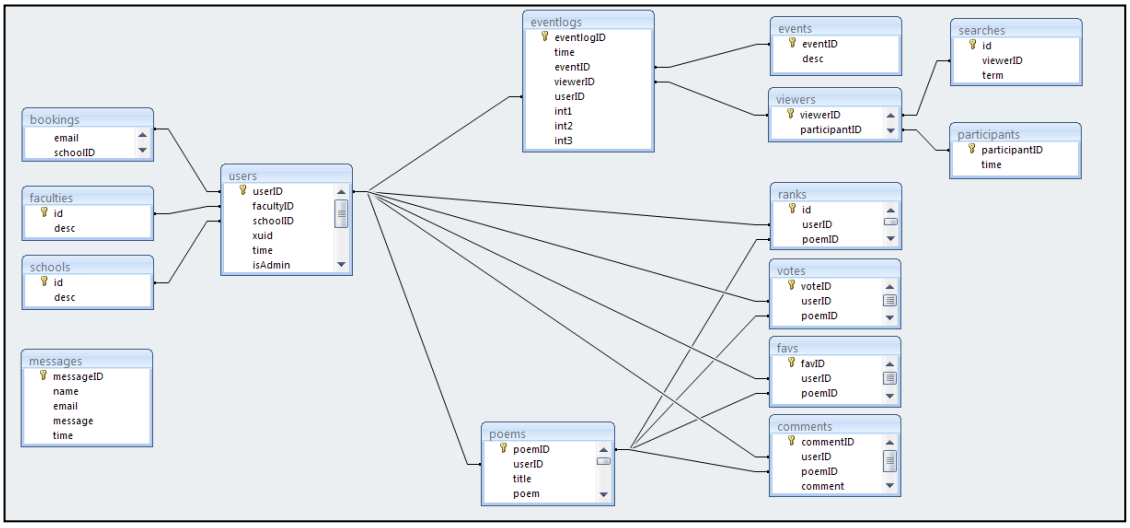

Figure 6 - The relationship between MySQL database tables in the experiment

Although behavioural data is stored in **`eventlogs`** table, additional information about the users and their involvement with the application is stored in other tables, and used in

this research to investigate the research hypotheses. For example, **`users`** table is used to store information about users' registration and verification. This information is used to investigate Inclusion dimension of VB. Other measures have been mathematically defined and will be described later (see sections 5.2 and 5.3).

4.13 Data Preparation

During the design and implementation of the web application which is used for this research, it was tried to maximise the amount of data which is collected in association with online experience of the users. Also, ICs and PHP sessions were used to identify and distinguish the visitors and their visits. As a result more than 250,000 records have been collected in the database. The records which are stored in **`eventlogs`** table include the majority of empirical data which is used for this research. However, there are constrains which need to be considered before moving to the next step. This includes the removal of the records from non-participants or the ones outside the time period which has been approved by the Ethics Committees. Additionally, any records associated with online robots (sometimes referred to as "bot"s) needs to be identified and removed (Middlebrook and Muller, 2000). Furthermore, any behavioural data in association with the site administrators and administrative tasks should be removed from the records.

The literature suggests (Cooley et al., 1999; Srivastava et al., 2000; Lee et al., 2001) different methods to convert the raw data (sometimes referred to as server logs) into datasets which can be analysed. In this section some of the tasks which are suggested by Cooley et al. (1999) in association with data preparation are explained, and where relevant the procedure is explained for this study's experiment.

4.13.1 Requester Identification

From the technical perspective, each page view is resulted from a request which has been sent to the server. There are different techniques which are used to verify the identity of the person who has initiated this request. However, in some cases it could even be difficult to prove whether the request has been initiated by human or automated software. One of the common ways for identifying users on the Web is the utilisation of passwords. Passwords can be transferred from one person to another. Hence, using one's password in a web-based platform (or other access details), does not necessarily prove the request is from the same person. Additionally, there have been numerous reports that even the most secure websites have been hacked, and sensitive information

has been leaked. For example, in November 2014, Sony Pictures computer system was hacked in an online attack, which forced the company to shut down its systems (BBC News, 2014). In a later incident in November 2015, TalkTalk which is one of the major telecommunications companies in the UK was hacked and nearly 157,000 of its customers' personal details were accessed (BBC News, 2014; London Stock Exchange, 2016). There are also other challenges in respect to identifying visitors based on registration details. For example many website users prefer to use the website without registration, and they also may provide invalid information due to privacy concerns (Cooley et al., 1999).

To avoid invalid details, in this experiment only official emails from the University of Manchester are accepted for registration. Furthermore, the users are also asked to verify their email address to use a majority of available Web 2.0 features. Also, in addition to registration details, ICs have been used in this experiment to identify and distinguish the visitors (see section 4.6.1). This is used to connect the events within the same visit and create a sequence of page views and events. ICs have also been used to identify returning visitors. Although the combination of verified registration and ICs can improve user identification, there are a few more limitations with these techniques. The bias can be resulted from use of multiple devices, or even multiple browsers in the same device. Additionally, by clearing the browser history the users are able to remove the existing ICs. There are other enhancing techniques which could be used to improve user identification. For example information such as IP Address can be used in association with the ICs to improve user identification. However, in this experiment the use of ICs has been the primary technique for user identification. Also, it is important to realise that identification techniques based on ICs distinguish devices (more accurately web browsers) rather than users. For example, if the same user access the website once from their laptop and later from their mobile, it can be impossible to identify them as the same user. However, in this experiment for registered users when they sign into their account, the ICs are combined with the login history to identify their previous visits (even made from different devices). However, such a technique is only applicable to registered users, and consequently in this research, ICs have been used as the only technique for identification of non-registered users.

4.13.2 Visit Identification

In addition to identifying the users and their behaviour, it is important to identify the group of activities which together build a single visit. Markov and Larose (2007) have

suggested different techniques which can be used for visit identification, and some of these techniques are explained in regards to the experiment for this research.

In this experiment PHP sessions (see setion 4.6.1), which are based on ICs, have been the primary techniques for identifying visits. It works by utilising temporary ICs to represent each PHP session, and these sessions are then analysed to identify each visit. The session identifier is stored locally as an IC, and each secondary request within the same session is associated with this identifier. Although use of PHP sessions can be sufficient in an ordinary scenario, they introduce a new set of limitations for this research. One of the problems with PHP sessions is the fact that it is based on temporary ICs, and this is sometimes reflected by interpretation of the device which has been used by the visitors. For example, it is expected for the sessions to be closed at the end of each visit by closure of the browser. However, when browser tabs are used, the users can continue to surf the web by opening and closing the tabs without closing the actual browser. In such scenarios the session can last for multiple visits. In this case the session identifier will be the same, and hence on the basis of PHP sessions alone, these will all be considered as one visit. There are also more questions which can be asked of "what a visit actually is". For example, if a visitor browsed the website for a while, and then took a 10 minute lunch break before continuing to browse the site, should this be considered as one or two visits? In this research specific definition is used for each visit to the website in the associated experiment. This is based on mathematical definitions which will be discussed later (see section 4.14.5).

The techniques which are associated with visit identification are usually based on navigation or time (Cooley et al., 1999). The navigational based identification of visits uses the referrer of the request (i.e. the previous request or page), to create a sequence of requests (or page views). This technique is useful in absence of ICs, and hence not useful in this experiment. On the other hand, time-based identification of visits utilises a cut-off point, and the time between requests are analysed and whenever it exceeds a time limit a new visit is assumed (Cooley et al., 1999).

In this experiment a cut-off point is used in addition to PHP sessions for enhancing visit identification. While Catledge and Pitkow (1995) suggest a cut-off point of 25.5 minutes based on empirical data, many commercial products use 30 minutes as a default cut-off point (Cooley et al., 1999). This should also be noted that the cut-off point should ideally be defined based on the type and content for each specific website. In this experiment a cut-off point of 30 minutes is chosen to separate the visits. In other words,

if two sequencing requests within the same PHP session are more than 30 minutes apart, they are considered as part of two separate visits.

4.13.3 Data Unification

In practice, data which are associated with OB can be stored on multiple tables and databases, and can be even located on different servers across the globe. To draw a complete picture of each visits, any piece of data associated with the investigated visits needs to be merged prior to visit identification and user identification (Tanasa and Trousse, 2004). However, in this experiment the web application was only connected to one MySQL database on one machine (acting as server) with a static IP Address throughout the process. For practical reasons some of the database tables had to be emptied before re-starting the poetry competition in 2014 (second stage of the experiment). These tables were backed up, and as part of the preparation process they were merged with the second year data to create the final datasets.

4.13.4 Robot Identification

One of the main resources of data for Search Engines are the data which are gathered automatically by software. The process which is initiated with such software includes accessing different websites and indexing their content by sending automated requests to the servers. They may be called spiders, crawlers, robots, or bots, but technically they are simply automated software for data collection. Although these pieces of software are usually harmless to the servers, they can also be used in some instances to attack the Internet servers, or collect confidential data by hacking into databases. While collecting behavioural information about the website users, it is likely to also collect considerable amount of data from these robots (Markov and Larose, 2007). It is very important to identify associated records within the collected data, and separate them from requests made by real humans. It should be noted that from a technical perspective, it is possible for these robots to look exactly the same as human users. Hence, more and more websites are using “Completely Automated Public Turing test to tell Computers and Humans Apart” (or CAPTCHA) to identify and stop these robots from accessing specific content or functionalities (e.g. registration). However in the case of Search Engines, these programs usually introduce themselves by using known signatures as their user agent. There are also mathematical techniques which can be used to identify these automated requests (Tan and Kumar, 2002).

In this experiment, the age verification stage requires use of ICs, and submission of a form. This combination usually stops a majority of robots (including the ones from Search Engines). However, it will be explained later (see section 4.14.6) that additional measures were also used during data preparation procedure to identify the records which are associated with these robots.

4.13.5 Path Completion

In traditional static websites, local caching is sometimes used to reduce the load on the server. In other words, when a user visits a page for the second time, no requests are sent to the server, and instead the previously displayed page is displayed again. In such settings the behavioural data which is generated from the server requests needs to be analysed and the missing links needs to be identified (Cooley et al., 1999). In this experiment, the server is configured to avoid caching, and this should generate a more accurate picture of the investigated visits. Any possible bias resulted from browser caching cannot been taken into account, and this is considered as a limitation for this investigation.

4.13.6 Data Integration

After taking the above steps, ideally the server logs can show a quantified representation of OB for real humans. However, there is usually more to be learnt about these individuals and their visits. For example in many online platforms (e.g. eCommerce), it is very important to identify the source of traffic. This can include Search Engines, Marketing Campaigns, or SNSs. Additionally, it can be important to know whether each visit is associated with any particular activity (e.g. online purchase), and if so what the outcome was (e.g. revenue, margins, etc.). The IP Address can also be analysed to produce relatively accurate information about the location of the visitors. Extended information about registered users (e.g. the information they provide during registration) can also be an important source of revealing information about them. Also, User Agent is a piece of information which is sent to server together with any request, and it can be used to identify the technology which is used by the users. This can include the type of device they are using (e.g. mobile, tablet, or PC), the operating system (e.g. Windows, Linux), and even the type and version of the browser they are using (e.g. Google Chrome version 48).

Data integration process consists of combining the behavioural data with other available forms of data which can be obtained for each visitor. This can be a vital step in

understanding the real individual behind the computer, and predict or influence their behaviour accordingly (Kohavi et al., 2004; Buchner and Mulvenna, 1999). In this experiment additional data includes quantitative measure such as number of votes, views, or shares which are associated with each poem and consequently can be linked to specific users. Most of these measures can be viewed from different perspectives. For example, in the case of voting for a poem, one can vote for a poem or one's poem can receive a vote. In this study such difference are represented as different dimensions for VB.

4.13.7 Semantic Categorisation

OB is usually presented as a sequence of page views or events. Although from a technical perspective these are each a distinct request to the server, they can be categorised semantically (Cooley et al., 1999). Although such categorisation may be subjective, it can usually introduce more meaningful performance metrics for the investigated visitors. For example in an eCommerce website Moe (2003) categorises the page views into buying, browsing, searching and knowledge-building. With such categorisations, it can be easier to introduce a trend among the users, and contemplate their needs, or influence their OB.

In this experiment the user activities are described based on events. For example, reading a poem, voting, or even receiving an error during registration have each been considered as an event. However, these events have been categorised semantically. For example, the events have been categorised to Page View and Non-Page View. In Non-Page View events, the semantic meaning of the events is considered for further analysis. For instance, although an error in registration and an error in sending poem are associated with different events, they can both be considered as a failed attempt and be categorised and investigated together. Also, further semantic considerations have been used specifically for this experiment. For example, received votes for each poem have been investigated based on whether the vote is submitted by the poem owner or another registered user. Also, this experiment is closely linked to a poetry competition, and the poems could have been analysed based on their content. While this is beyond the scope of this study, some associated quantitative measures such as number of words or characters in each poem are included in the analyses.

4.14 Data Preparation Procedure

In the previous sections some of the primary steps for data preparation were explained. In this section the implementation of these steps for the collected data from the experiment is described. The final set of data is exported from the server as a MySQL file, and after preparation, it is used to test research hypotheses. The original MySQL file contains the collected data from 28/2/2013 to 15/10/2014, and during this time 252,556 events have been logged in **`eventlogs`** table. Figure 7 shows the total number of events which are logged in each week during the data collection period, and before starting the data preparation procedure.

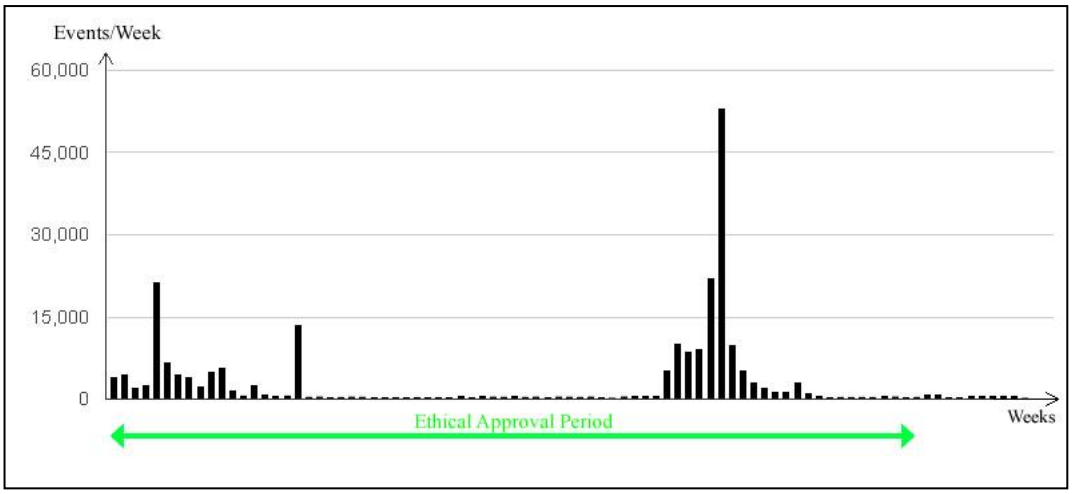

Figure 7 - Total number of events logged in each week from 28/2/2013 to 15/10/2014

In the next sections, the procedure which is used to convert this raw data to what is used for hypotheses testing is described.

4.14.1 Defining Experiment Stages

The raw data contains records which have been collected in more than 19 months. However, the obtained Ethical Approval only covers 17 months of this period (highlighted in Figure 7). Additionally, as explained earlier (see section 4.1), the data is associated with a poetry competition in two sequencing years (2013 and 2014). On this basis, the experiment is divided into two stages. The duration for each of the stages is the same (23 weeks), and it starts from the first day of competition in each year.

After choosing the appropriate time periods for each stage of the *experiment*, any *PHP sessions* outside these periods are identified and flagged. As explained earlier (see section 4.12.2), any records stored after age verification is associated with an identifier

(**viewerID**). This link is used to flag any records in the **eventlogs** table which should not be included in "Stage 1" or "Stage 2" of the experiment (see section 4.8).

4.14.2 Special Users

The developed platform was used primarily by the users for submitting poems, and browsing the submitted poems. However, special modules were developed for managing the generated content, and monitoring the website's performance. Also, the platform was used by the first judging panel (Dr Peter Fenn and Dr John McAuliffe) to rate every single approved poem. Such functionalities are only available to very few users, and these users have been flagged in the database as administrators. As part of data preparation any records associated with the administrators were identified and flagged.

4.14.3 Non-Participants

At the start of their first visit, every website visitor is presented with a notification to confirm their age, and agree with the privacy policy of the website (Figure 8). They cannot go any further unless they tick the first two boxes and submit the form. However, there is another box which can be ticked as sign of agreement to participate in this research.

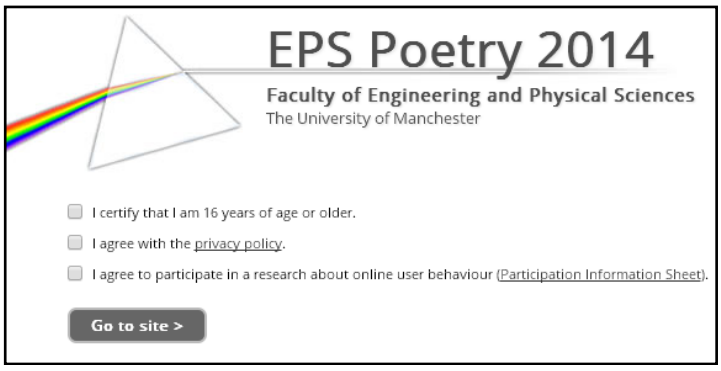

Figure 8 - Age verification, privacy policy acceptance, research participation

In both stages of the experiment around two thirds of the users have agreed to participate in this research. As part of the data preparation procedure, any records which is stored in **eventlogs** table before this confirmation, or that is associated with non-participants (who did not agree to participate in this research) are identified and flagged.

4.14.4 Multiple Registrations

In few instances, the same session is associated with more than one registered users. Although possible, it is unlikely for two students to use the same browser for

registration. On the other hand, this could result from an attempt to rig the votes. These sessions have been identified in the database, and any record associated with them has been flagged. The removal of such a record during data preparation procedure has also created a one to one relationship between sessions and registered users. This has been used later in the process to improve identification of the individuals associated with each visit (see section 4.14.5).

4.14.5 Visit Identification

The visits are initially defined based on PHP sessions. On this basis each visit receives a unique identifier based on sessions (**`viewerID`**). In reality the sessions can last for multiple visits. For example, if a user leaves the website without closing the browser, and come back to it after couple of hours the PHP session will continue. To overcome this technical limitation, a cut-off point of 30 minutes has been introduced. The following definition is used for identifying single visits in this experiment:

> ***In this experiment, a "single visit" is defined as a sequence of requests within a single "PHP session", in which no sequencing requests are more than 30 minutes apart.***

All records within the **`eventlogs`** table which are not flagged in previous steps have been grouped based on their PHP session identifier, and using the cut-off point of 30 minutes, the visits are identified, and each have been given a unique identifier (**`visitID`**).

4.14.6 Robotic and Semi-Robotic Behaviour

After identifying single visits, number of requests within each visit has been analysed to identify robotic behaviour. It is important to emphasise that visits are only based on records from the participants (i.e. have ticked a box to take part in this experiment), yet some of them can have been generated by software. For example, by investigating the identified single visits, in two instances it is very likely for the requests to have been initiated from an automated software. In both of these instances, more than 6,500 requests have been made in around one minute. Furthermore, there are few other visits which have an unexpectedly high number of requests. However, in further investigation of such records, it is quite obvious that these have been generated by continuously refreshing specific pages. To understand this behaviour, it needs to be noted that in the second stage of the experiment, the competition rules suggest a higher chance of short

listing for the poems with a higher number of views. Although, the number of views was not displayed on the poem page, the users could sort the poems based on number of views, and see the number of views in the listed results. After investigation of visits with high number of requests, any visit with more than 300 requests was flagged as containing robotic or semi-robotic behaviour.

4.14.7 The Subject Group

After removing the previously explained flagged records, a list of all participants with existing visits have been produced, and have been divided based on defined stages of the experiment. However, there are few participants which have records in both stages. This can technically be explained as continuous existence of ICs for around one year (which is the expiry duration on the original IC). These records have been removed to enable the division of participants into "Stage 1" and "Stage 2" of the experiment (see section 4.8). Additionally, after combining the visits, there are few participants which are associated with more than one registered user. These participants have also been removed to enable creating a many to one relationship between participants and registered users. The remaining participants are the investigated Subject Group for this study. Each of these participants is identified with a unique identifier (**`participantID`**), and during the investigated period there is at least one single visit to the website for every one of them. Furthermore, each participant is linked to either "Stage 1" or "Stage 2" of the experiment.

4.14.8 Final Dataset

After removing unidentifiable (before age confirmation) and flagged records (explained in previous sections), the data is stored in **`spss_eventlogs`** table. This table includes every record from **`eventlogs`** table, which is within time periods of "Stage 1" or "Stage 2" of the experiment, and has not been flagged during previously explained steps. Figure 9 shows the number of events per week after data preparation process, and highlights the 23 weeks which are used to identify each stage of the experiment.

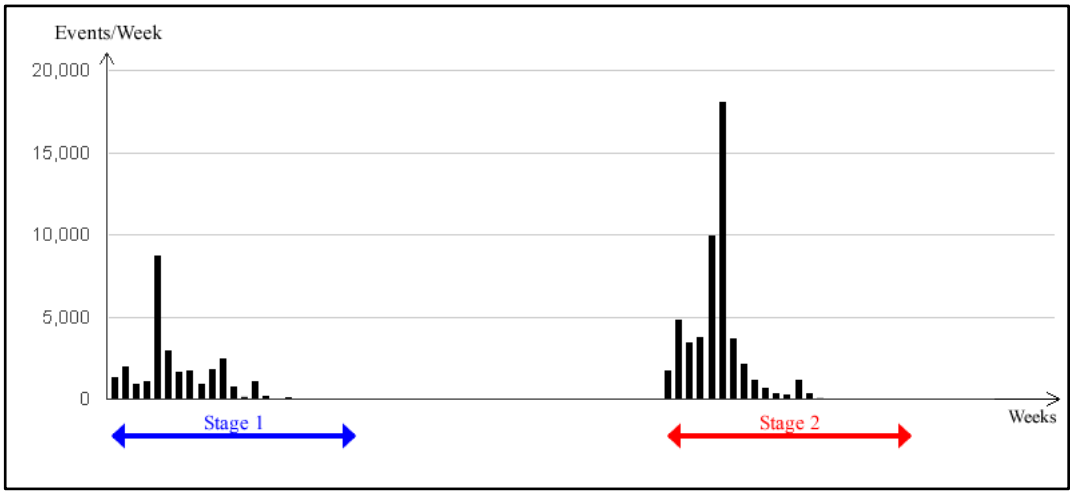

Figure 9 - Total number of events per week after data preparation process

The comparison between Figure 7 (before preparation) and Figure 9 (after preparation) can visualise the result of data preparation procedure.

4.15 Summary

Based on the selected paradigm and approach (see section 3.6), collecting and analysing quantitative data from an online experiment was chosen as the method for this investigation. This chapter described some of the technical and non-technical steps which were taken for running the experiment. Initially, the competition which is used as the vehicle for this experiment was described (see section 4.1), and the scope and participants of the experiment were explained (see sections 4.2 and 4.3). This was followed by detailing the process for obtaining Ethical Approval for this research (see section 4.4). Thereafter, development of the website and the bespoke tracking system which is used in this experiment were described (see sections 4.5 and 4.6). Also, OB and VB were specifically defined in this experiment (see sections 4.9 and 4.10). Data collection procedure during the experiment was described by using a hypothetical scenario (see section 4.12.1). In this experiment, empirical data is collected by using a bespoke Tracking System (see section 4.6) which has been specifically designed and developed for this research. This Tracking System, collects quantitative data in real time, and stores them in a database on the server. The structure for this database was described in this chapter, and the database tables which are specifically used by the Tracking System were explained (see section 4.12.4). The data which has been collected by the Tracking System can be considered as raw data, and it requires preparation before statistical analyses.

Some of the steps which have been suggested in the literature for data preparation in similar scenarios have been reviewed (see section 4.13). Finally, data processing procedures which have been implemented in this investigation were described (see section 4.14). In the next chapter, the final dataset (see section 4.14.8) is used to investigate research hypotheses.

5 Results

In previous chapter, data collection (see section 4.12) and data preparation procedures (see section 4.14) have been explained in the experiment. In this chapter the investigated metrics are described and specifically defined for the experiment. These metrics are used to investigate research hypotheses (see section 3.7). There are many metrics which have been suggested and used in the literature and practice for investigating OB (see sections 2.2.4 and 2.2.7). Some of these metrics have been adopted and defined specifically for the experiment which is associated with this research (see sections 5.2 and 5.3). Also, additional metrics (see section 5.2.5) are defined based on the specific aspects of this project (i.e. submitting and reading poems). This is followed by describing and justifying the statistical methods which are used in this research (see section 5.4). Finally, research hypotheses are investigated in regards to Online Environment (see section 5.5), and different dimensions of VB (see sections 5.6, 5.7 and 5.8).

5.1 Primary Definitions

Tracking data is one of the useful information which is available to project managers. Large amounts of such data is usually collected automatically by tracking systems and stored quantitatively. This data can be analysed to provide different metrics associated with the website and its users' OB. Although various metrics can be defined and measured for each website, the project managers usually concentrate on the ones which are related to the project objectives. For example, in eCommerce projects the important metrics can be associated with purchase rates and profitability. However, in Web 2.0 projects other factors such as active participation measures might be the matter of interest for project managers. These matters of interest are usually associated with the Key Performance Indicators (KPI) for the project. Although from charity projects to online banking projects a wide range of KPIs can be defined and measured, in within the online platform which is associated with these projects there are usually a limited number of metrics which can represent users' interaction with the website. Some of these metrics such as number of views for different pages or average duration of page views can be measured for almost every website. However, the same information can have different interpretations, and even opposite relation in regards to project KPIs. For example, in a website which generates revenue based on per display advertising, a lengthier navigational path can be considered good, while in an eCommerce website a

lengthier path can be considered as a negative influence on usability, and hence not good.

There are different metrics which have been used in the literature and practice to assess web based phenomena, and compare different users and projects. In this section some of these metrics are discussed and categorised. In later sections, these metrics are investigated for collected data from the experiment, and the results are presented in form of tables and graphs. In practice, it might not be necessary to use all of these metrics, and the relevant metrics should be chosen by the project managers based on project objectives. The metrics which are used by project managers should reflect the project's KPIs. These indicators can differ from one project to the other; nevertheless they can be linked to quantitative metrics derived from the website such as traffic, visitor loyalty, and profitability. Some of the discussed metrics are rather generic, and can be defined for most of online platforms. When more Web 2.0 specific details are included (e.g. number of poems or votes), specific metrics are created for the sake of this research. However, these metrics can be defined in similar ways for any other Web 2.0 application.

Before explaining the investigated metrics for this experiment, some primary definitions are presented in this section. These definitions are specific to this experiment, and are used when investigated metrics and their measures are described later (see sections 5.2 and 5.3).

5.1.1 Event Log

As explained earlier (see section 4.6), in this research a bespoke Tracking System has been used to collect behavioural data. This data is stored in **`eventlogs`** table of the database (see section 4.12.2) and each record is associated with the type of event, a timestamp, and an identifier for the user who has triggered this event. Each of these records are defined in this research as an Event Log (EL). A list of events which can be triggered have been defined and logged during the experiment. These events include the traditional "page views", but have been extended to include additional events such as "receiving an error", "voting" or "clicking on a share button". In addition to the type of event, the time of the event is also logged and used in this research. The complete list of events which have been logged during this experiment with the frequency of logging after preparation is displayed in Table 2.

Rank	Event Description	Total		Rank	Event Description	Total
1	Poem Page	24,748		**27**	Ceremony Page	175
2	Home Page	14,981		**28**	Edit Poem Page	162
3	List of Poems Page	14,527		**29**	Edit Poem - Success	146
4	Home Page Tabs	4,201		**30**	Home Read More Click	140
5	Special Link Click	3,110		**31**	Ceremony Booking - Fail	132
6	Registration Page	2,229		**32**	Un-Vote Poem	113
7	Login Page	1,814		**33**	Ceremony Booking - Success	68
8	Cookie Login - Success	1,605		**34**	Booking - Success	67
9	Login - Fail	1,480		**35**	Send Comment - Success	66
10	Trying to Vote Not Reg	1,449		**36**	Booking - Fail	57
11	My Account Page	1,264		**37**	Contact - Success	41
12	Submit Poem Page	935		**38**	Search Page	40
13	Winners 2013 Page	890		**39**	Search Results	40
14	Verification Page	878		**40**	Change Password Page	21
15	Registration - Fail	802		**41**	Submit Poem - Fail	14
16	Registration - Success	800		**42**	Send Audio Page	12
17	About Us Page	685		**43**	My Favourites Page	11
18	Vote Poem	662		**44**	Privacy Policy Page	10
19	Login - Success	600		**45**	Send Audio - Success	10
20	Remember Me	524		**46**	Participation Information Sheet	8
21	Verification - Success	471		**47**	Cookie Login - Fail	4
22	Logout	345		**48**	Send Comment - Fail	4
23	Contact Us Page	297		**49**	Password Change - Success	3
24	Submit Poem - Success	236		**50**	Fav Poem	3
25	Verification - Fail	230		**51**	Password Change - Fail	2
26	Booking Page	192		**52**	Edit Poem - Fail	2

Table 2 - Description and frequency of "Event Logs" after data preparation (Total: 81,306)

It is important to note that additional information which is logged alongside the records for events could make a specific event to present a group of sub-events. For example, "Poem Page" is defined as an event and represents any request to view any of the available poems. However, the identifier of the requested poem is also stored which can be treated as a sub-event. The total number of such records before data preparation was 252,556 and it was reduced to 81,306 records after data preparation (see section 4.14). It should also be noted some of these events such as "Special Link Click" include a variety of events which are identified based on additional data stored in **`eventlogs`** table.

5.1.2 Single Visit

As explained earlier (see section 4.14.5), in this investigation Single Visit (SV) is defined as a sequence of consecutive events logged within a single PHP session, with a cut-off point of 30 minutes. In other words, it is a sequence of ELs which are all linked to a single IC on the client device, and no consecutive ones are further than 30 minutes apart. In this experiment, each Single Visit is associated with a unique identifier (**`visitID`**), and there are a total of 12,472 visits after data preparation process.

5.1.3 Recognised Participant

Each time a website visitor has visited the website without having the age verification IC, they have been presented with a form which in addition to age verification and privacy policy acceptance, invites them to take part in this research. If they choose to take part, an IC will be stored on their device, which identifies them in their probable next visits. The time of agreement is stored in the database with a unique identifier (**`participantID`**). However, as explained earlier (see section 4.6.1) due to limitations associated with ICs, these are not necessarily connected to one individual. In other words, if one of the individuals who are participating in this research use multiple devices to access the website, they will have more than one **`participantID`**. With this in mind, there are 5,759 *Recognised Participants* (RPs) in this research.

5.1.4 Registered User

During some parts of the experiment, the students with an official email address from the University of Manchester could register, and benefit from additional Web 2.0 functionalities such as poem submission, commenting or voting. The emails are validated on the server, but due to unavailability of a database of valid university emails, any email ending with "manchester.ac.uk" or "man.ac.uk" has been accepted. Any visitor who has completed the registration process, is referred to as Registered User (RU), and is identified with a unique identifier (**`regID`**). It is important to note that during the second stage of the competition, the registrations from the previous year was dismissed, and the students who had registered in the first stage, needed to register again for active participation. This removes the possibility of using an account for which the email address had been expired. In this experiment, there are a total of 590 Registered Users after data preparation process.

5.1.5 Verified User

After each successful registration, a Verification Code is sent to the provided email address. The registered users can verify their access to that email by entering the code in the website. Any registered user which has successfully finished the verification process is referred to in this research as Verified User (VU). In this experiment, there are a total of 511 VUs after data preparation process.

5.2 Quantifying POB

As explained earlier (see section 2.2), two dimensions are assumed for OB in Web 2.0 platforms. The first dimension is *Passive Online Behaviour* (POB) and it is associated with using the platform without using its Web 2.0 features. Such activities includes visiting (or re-visiting) websites, and browsing their web pages. The associated metrics with such behaviours are categorised in this research and in each category different metrics are defined specifically for the experiment.

5.2.1 Traffic Metrics

As explained earlier (see section 2.1.4), websites' traffic represents the individuals who visit them, and it is one of the important factors for comparing and investigating websites. However, due to diversity of online platforms, in some cases the exact definitions and measures can be different. In this section the relevant metrics are explained, and the calculation method in the experiment is explained.

The number of times a website has been visited during a time period is one of the traffic indicators and can be used to measure or compare the popularity. On this basis, Number of Single Visits (NSV) is defined in this experiment as the total number of SVs which are logged during a specific period of time. Also, this can be narrowed to specific visits during time (e.g. the number of visits from Registered Users within a certain time period). Another important metrics for measuring traffic is the number of unique visitors. However, the *uniqueness* can be interpreted differently. Although it is rather easy to distinguish between different visits to the websites, sometimes it can be more complex to realise which one of these visits have been made by the same person (i.e. a unique visitor). For example, if an individual use their office computer and their personal laptop to access the same website, it would be difficult to link these visits purely based on the tracking data. As explained earlier, due to technical limitations which are associated with use of ICs (see section 4.6.1), and restrictions which are imposed by the Ethics Committees (see section 4.4), there are limitations in tracking the users. In this experiment these limitations have been acknowledged and the term Recognised Participant (RP) is used instead of *unique visitor*. On this basis, Number of Recognised Participants (NRP) is defined based on existence of a specific IC (**`participantID`**) during specific time period. Furthermore, when analysing registered and verified users, a combination of ICs and registration details and login history is used to improve on this metric. Despite the explained differences between

NRP and number of unique visitors, it should be noted that in comparisons within the same experiment, any bias caused by the limitation of ICs exists in both sides of the argument.

While the number of visits to the website is an important metric in relation to websites' traffic, it can be as important to find the number of pages which are viewed over a period of time. In some cases this can be directly linked to the revenue stream of the website. For example, for the websites which use per display adverting, keeping the users on the website, and their continuous surfing of the pages can be as important as attracting them in the first place. In other words, when the websites are paid to display advertisements, if one user views two web pages, they would be equal to two users who each view one web page. It is important to note that due to increasing use of JavaScript in new online platforms, the interaction between the platform and the users does not necessarily involve the navigation among web pages. Hence, new interpretations of traditional metrics might be required by the increasing use of JavaScript in web design. The term *event* has been used in this experiment to expand on some of the existing metric. For example, in the addition to traditional page to page navigation in the website, changing the tabs to consume additional data has been considered as page view. On this basis, the specific ELs are marked as Page View events, and the Number of Page Views (NPV) is defined as the number of ELs linked to events marked as "**`isPage`**" during a specific period of time. After data preparation, around 84% of the investigated ELs are considered as page view (69,586 records), and the other 16% are presenting other events such as "receiving an error" or "submitting a form".

5.2.2 Visit Metrics

There are different measures which can be used to quantify the attributes of each visit (sometimes referred to as depth of visit). The main factors which have been used for such an analysis include number of visited pages and the time spent on the website. Considering the technical and ethical limitations in this research, the following metrics have been used for measuring depth of visit.

The time which is spent by visitors in each visit to the website can represent their interest towards the website, and has been used as a metric in the literature and practice to measure depth of visit (Danaher et al., 2006). In this experiment, a similar concept is defined based on technical limitations, and the requirements from the Ethics Committees. On this basis, Visit Duration (VD) is the amount of time between the first

and last Event Log (EL) recorded for a Single Visit (SV). The following limitations can be recognised:

- For first visits (for which **`participantID`** is not available as an IC), any time before participation agreement is ignored in this definition.
- The time spent on the site after the last request (e.g. last page view) is ignored in this definition (i.e. time spent on the last page is ignored).
- On the basis of this definition, VD cannot be defined for visits with a single page view (or first visits with single page view after participation agreement).

Despite the above limitations, it should be noted that in comparisons within the same experiment, any bias exists in both sides of the argument.

Average Visit Duration (AVD) is defined in this experiment for a set of visits (SVs) as the statistical average of Visit Duration (VD) when single page view visits are dismissed. On this basis this metric can only be defined for SVs with at least one visit for which VD can be defined.

The number of pages which are viewed within the same visit can represent visitors' interest towards the website, and can be used to compare visits. In this experiment, the tracking data which is collected during a visit is further than just the page views. Some of the events have been combined with the traditional definition of page views to calculate the length of visit. For example, changing tabs in the website's homepage uses JavaScript to display hidden text, and has been treated as page view. On this basis, Visit Length (VL) is defined for a Single Visit (SV) as the number of ELs linked to events marked as "**`isPage`**" during that visit. Any page views before participation agreement are ignored in this definition. Average Visit Length (AVL) is defined in this experiment for a set of visits (SVs) as the statistical average of Visit Length (VL) for those visits.

Number of unique page views in a visit can also be important to analyse user behaviour by eliminating duplications in users' navigational path through the website in each of their visits. Due to the navigation structure of a website, during a visit some of the pages might be viewed more than once. For example, a participant in the experiment might start from the home page, then navigate to list of poems, and then read a poem. Then they may go back to the list and choose another poem, click on it, read it and then close the browser. While in this visit 5 pages have been viewed (VL=5), only 4 distinct pages have been viewed (list of poems is viewed twice). In this experiment, the recorded ELs are grouped in each visit based on the combination of **`eventID`** and **`int1`** fields to

produce the number of unique pages which are viewed. In the case of "poem list", any sorting order has been considered as a different page, and in this case the combination of "**`eventID`**", "**`int1`**", "**`int2`**", "**`int3`**" has been used for grouping. On this basis, Visit Compressed Length (VCL) is defined for a Single Visit (SV) as the total number of unique pages viewed during that visit. Finally, Average Visit Compressed Length (AVCL) is defined in this experiment for a set of Single Visits (SVs) as the statistical average of Visit Compressed Length (VCL) for those visits.

5.2.3 Navigational Metrics

The visit metrics which were discussed in the previous section ignore the order of events during a visit, and focus on existence and frequencies. The order of events together and the navigational path are investigated in this section and various measures are introduced to compare visits from navigational perspective.

One of the common reports for a website is the popular Entry Pages and Exit Pages. In a typical visit, the Entry Page (sometimes referred to as Landing Page) is the Home Page for the website. On this basis the home page usually contains different sections associated with and linked to different parts of the website. However, this is not always the case. For example, in this experiment when someone shares their poem on Facebook, the Landing Page for the referring visitors would be that specific poem. So, the stickiness of the "Poem Page" could be as important as that of the "Home Page".

In this experiment due to the expectations of the Ethics Committees, the Landing Page (LP) is defined for first visits as the page which has been viewed just after participation agreement. This is due to the fact that in the first visit ICs are not used before visitors consent. However, in secondary visits, the use of ICs has already been agreed, and the LP is the first viewed page of the visit. In this experiment, Landing Page Rate (LPR) for a specific page or set of pages is defined as the proportion of Single Visits (SVs) within a set of SVs for which those pages are the Landing Page.

The last page visited by the user in a specific visit can be also important. By analysing the top exit pages, the managers can improve the stickiness for these pages to improve depth of visit for the users. The Exit Page (EP) is defined in this experiment for a Single Visit (SV) as the last page viewed during that visit. In this experiment, Exit Page Rate (EPR) for a specific page or set of pages is defined as the proportion of Single Visits (SVs) within a set of SVs for which those pages are the Exit Page.

One of the common metrics for websites is associated with Single Page Visits. The bounced visits are defined as those visits for which the entry page is also the exit page. In other words, it represents the users who enter the website but leave without any further interaction. On this basis, in this experiment, Bounce Rate (BR) is defined for a set of Single Visits (SVs) as the proportion of bounced visits (VL<2) among them. As explained earlier any page views before participation agreement have been ignored due to the expectations of the Ethics Committees. Bounce page is defined in this experiment as a page which is the only page which is viewed during that visit. On this basis, proportion of visits (SVs) within a set of SVs for which a specific page or set of pages are the bounce page is defined as Bounce Page Rate (BPR) for those pages.

Visit Path (VP) is defined for a Single Visit (SV) as a sequence of page views which are linked to that visit. The pages can be categorised based on their type, and in this case VP represents a sequence of page types (i.e. referring to viewing any poem rather than a specific poem). In this experiment, Visit Path Rate (VPR) is defined as proportion of visits with a specific Visit Path (VP) among a set of Single Visits (SVs). This metric is used to identify popular (or frequent) VPs.

The type of content which is provided by the website can influence Visit Duration (VD). For example, it is understandable if the visitors spend less time on the pages containing still images in comparison to the pages which contain lengthy videos. It is also important to emphasise that amount of time spent on a page can represent both engagement and lack of usability. While it is desirable for the users to spend time consuming the content, the time spent on navigation or registration might not be appreciated. In this experiment, Page View Duration (PVD) is defined for any page view during the visit except the Exit Pages (more accurately the ones which no event is logged after them). For such page views, the amount of time between the page request and the next request to the server is defined as PVD for this page view. In this experiment, Average Page View Duration (APVD) is defined for a specified set of pages as the statistical average of the amount of time spent on those pages within a set of visits (SVs). It should be noted that in this calculation the page views for which PVD is not defined (exit pages) have been ignored. Also, for a specific visitor or group of visitors, APVD can be calculated for all pages, and represent overall Average Page View Duration for those visitors.

5.2.4 Loyalty Metrics

One of the objectives for visitor identification is to recognise returning visitors. Depending on the project, this information can be used differently, and in some cases this can be a vital part of the concept. For example, Facebook usually sorts the content based on individuals' interests and previous engagements. This puts an enormous value on user identification to the extent that before login there is only one primary item available on the home page; the registration form.

In this experiment the content is available to the public, and internet ICs have been used for identifying returning visitors. In case of Registered Users, additional techniques have been used to enhance user identification. One of the factors which can indicate loyalty among the users is the rate of returned visitors. This can also be combined by traffic metrics to create meaningful and related measures. In this experiment based on implemented user identification techniques, different metrics have been used to measure and compare visitors' loyalty.

Visit Count is the total number of visits which are linked to an individual. However, based on the limitations associated with ICs, two different metrics have been used in this study. Recognised Participant Visit Count (RPVC) is the total number of visits associated with a unique participation IC (identified with **`particpantID`**) during a specific time period. When login and registration information is available, it has been merged with information available from ICs, and on this basis Registered User Visit Count (RUVC) is defined for a Registered User (RU) as the total number of visits during a specific time, and based on merger of ICs from the same user. This metric is only defined for Registered Users, and although it reduces the limitations caused by ICs, it cannot remove them entirely. For example, if an individual use three device to access the website, but have logged into their account only from two of these devices, the visits from these two devices can be merged to produce a more accurate number for this Registered User. However, the visits made from the third device cannot be linked to them.

For returning visitors, the time between visits can be associated with their loyalty and the website's stickiness. When visits are combined based on ICs, Average Time Between Recognised Participant Visits (ATBRPV) is defined as the statistical average time between their visits. This should be noted that this is only defined for Recognised Participants (RPs) with more than one visit during the investigated time period. When

visits are identified and linked based on ICs and additional registration information, Average Time Between Registered User Visits (ATBRUV) is defined as the statistical average time between their visits. This should be noted that this is only defined for Registered Users (RUs) with more than one visit during the investigated time period.

Rate of Identified Returns (RIR) can be calculated when the returning visitors are identified, and their visits are combined. Irrespective of the type of return identification, RIR is defined as the proportion of visitors with more than one visit. In case of merely using ICs for return identification, RIR shows the proportion of Recognised Participants with more than one visit. When the registration information is used in addition to ICs, more visits can be identified and linked to the same individual, however this is not available for all visitors.

5.2.5 Project Specific Metrics

Any website or type of website can have specific metrics. For example, as important as it can be for an eCommerce website to have a high traffic and loyal visitors, it would not be useful unless if it consequently brings sales or other financial gains. Hence, the number of visits which are converted to sales can be a crucial metric in an eCommerce website. Other factors such as the profit margins and cost of marketing can also play a significant role in combination to the general web metrics. On the other hand, in a Web 2.0 platform, collaboration and engagement metrics can be matter of interest, and can be defined respectively.

One of the justifications for use of the developed web application in association with "EPS Poetry Competition" was to enable public users to read the submitted poems. Hence, in this experiment specific metrics have been defined to measure and compare visits to "Poem Page".

In this experiment, Poem Page View (PPV) is defined as the number of times the poem page has been viewed in an SV. Average Poem Page View (APPV) is defined in this experiment for a set of visits (SVs) as the statistical average of PPV for those visits. Number of Poem Page Views (NPPV) is defined in this experiment as the total number of times the poem page has been viewed during a time period. Additionally, Unique Poem View (UPV) is defined as the number of distinct poems viewed in an SV. For a set of SVs, Average Unique Poem View (AUPV) is defined in this experiment as the statistical average of UPV for those visits. In this experiment, Extended Unique Poem View (EUPV) is defined as the number of distinct poems viewed in a set of SVs. For

example, after identification of return visits for each individual this can show the number of distinct poems read by them during all their visits. It should be noted that while going to a poem page is regarded as reading that poem, this is not necessarily the case. For example, the users might go to the poem page for a specific poem, and do not read it at all, or only read parts of it. This cannot be verified in this experiment, and technically what is regarded as reading poems is requesting them from server by users. Finally, in this experiment, Average Extended Unique Poem View (AEUPV) is defined as the statistical average of EUPV for a set of SVs.

5.3 Quantifying AOB

In a Web 2.0 platform, users can choose to use the provided Web 2.0 features and get actively involved with the platform. However, even without active participation they can get involved by surfing the website and consuming its content. The metrics which have been discussed so far can similarly be used for any other online platform. However, in the experiment associated with this research, Registered Users (RUs) can get actively involved and influence the output from the underlying software. As explained earlier (see section 2.2.6), based on what has been suggested by Benkler (2002) three dimensions for AOB have been considered and corresponding metrics have been defined in this experiment. These are "Content", "Relevance/Accreditation" and "Value-Added Distribution". Additionally, users' interactions with SNSs are described in more details (see section 5.3.4) and corresponding metrics are defined.

5.3.1 Content

Content generation is one of the fundamental aspects of AOB in Web 2.0 platforms. Without content the other possible forms of participation will be redundant. In this experiment the generated content can be in form of "poem", "comment" and "audio link". Statistical measures are used to quantify the amount and quality of the content. In this experiment, registration is required for any form of content generation. Hence, these measures are defined for a Registered User (RU), or a set of Registered Users (RUs). No content is generated by unregistered participants, and when required in the analyses these measures are used as zero.

In this experiment "**`hasPoem`**" and "**`numPoems`**" have been used to quantify existence and total number of submitted poems. Similarly, "**`hasComment`**" and "**`numComments`**" measure the generated content in form of comments. Finally,

"**hasAudio**" and "**numAudios**" are defined to quantify the generation of content in form of audio link.

5.3.2 Relevance/Accreditation

Usually, different interactions between Web 2.0 platforms and their users are used to sort the content. For example, in Facebook different factors including the number of "Like"s are used to sort the content in News Feed. To measure active participation in this experiment, a voting process is available to the users, and a cash reward (£50 in the first stage and £100 in the second stage) was used as an incentive. In the first stage of the experiment, email verification was not required for voting. Additionally, the number of votes were hidden in the first stage and not hidden in the second stage of experiment. Also, as a more sentimental incentive, in the second stage the poems which received the highest number of votes in each school were displayed in the Home Page. In this experiment, "**hasVote**" and "**numVotes**" have been used to quantify existence and total number of votes which have been received from each RU. The number of Facebook Likes which are received by each poem has also been used as a measure of accreditation, and "**hasFbLike**" and "**numFbLikes**" are used to represent accreditation on SNSs.

5.3.3 Value-Added Distribution

The final form of Active Participation is sharing the content. There are different methods which could be used to share the content which is available in this experiment. For example, the visitors can email the link to their friends, or share it on SNSs such as Facebook and Twitter. The actual data representing the number of shares can be very difficult to obtain, and it is not available in this experiment. However, a related metric has been defined based on the number of times the share buttons were clicked. On this basis, the number of times the share buttons associated with Facebook, Twitter and GooglePlus were clicked is calculated, and "**hasShare**" and "**numShares**" are defined for each visitor to represent their attempt for distributing the content in SNSs.

5.3.4 Social Media Metrics

As explained earlier (see section 2.1.6), SNSs can be seen as a communication tool for their users. In essence, users get notified about the content which is created and/or sorted by other users. Furthermore, they can continue the communication process by notifying people in their own network. Different techniques and implementations can be imagined for such process, some of which has been utilised by existing SNSs. For

example, in Facebook users receive the sorted content in their News Feed, and they can notify their friends by clicking the “Like” or “Share” button. The generated and shared content can be in different forms including text, image, video and link. When users share a link, the actual content exists outside that platform. Should the recipients choose to click on the link, they will be redirected to the source of content. The traffic which can be attracted from SNSs could be a motive for web based projects to use facilitating features. One of the common techniques for connecting to SNSs is use of “Share Buttons”.

In this experiment, three major SNSs (Facebook, Twitter and GooglePlus) were chosen and the connection was investigated. To do so, the connection between the website and SNSs is categorised to Outgoing and Incoming. “Outgoing Connections” are associated with the amount of “Share”s in chosen SNSs. On the other hand, “Incoming Connections” are associated with the traffic which is referred to the website from the investigated SNSs. The share buttons were used in association with each poem, and visitors could share each poem by clicking on these buttons.

In this experiment, Number of Outgoing Visits (NOV) is defined as the total number of visits within specific time duration which are associated with click on any of the custom “Share Buttons”. On this basis, Rate of Outgoing Visits (ROV) is defined for a set of SVs as the proportion of them with an Outgoing Connection. It is important to note that clicking on the provided share buttons is not the only way to share the poems, and visitors can alternatively copy and paste the URL in each of the investigated SNSs. However, the defined metric is only based on the data which could be collected from the visitors. Furthermore, there is a possibility for the users to click on the button without finishing the share process in the associated SNS. Once again, this could not be verified by this research, and it is considered as a limitation.

When the poems are shared using the provided “Share Buttons”, the shared link is associated with an identifier. So, when users of the associated SNSs click on the link and are redirected to the website, they can be identified. On this basis, Number of Incoming Visits (NIV) is defined in this experiment as the total number of visits within specific time period which are associated with the “link” which was generated by the custom “Share Buttons”. On this basis, Rate of Incoming Visits (RIV) is defined for a set of SVs as the proportion of them with a logged Incoming Connection. Once again it should be emphasised that Incoming Connections are the ones which could have been tracked through the tracking system, and there can be more traffic from the investigated

SNSs or other networks which have not been identified. This is considered as a limitation for this research, however the same technique has been used throughout the experiment, and the results are mainly used to compare, and not to measure.

5.4 Statistical Method

As explained earlier (see section 3.8), eight hypotheses were proposed in this research in regards to OB and its association with Online Environment and VB. In the following sections these hypotheses will be tested. To investigate these hypotheses, the associated null hypotheses are described in each section, and their probability is statistically investigated.

To use parametric statistical tests including T-Test, the assessment of normality assumption needs to be taken in to account. Shapiro-Wilk test provided by SPSS software has been used to investigate whether the data is normally distributed (Shapiro et al., 1968; Thode, 2002; Ghasemi and Zahediasl, 2012). In such tests, the score in the sample is compared with the set of scores which are normally distributed with the same mean and SD. The null hypothesis is the normal distribution of sample, and a p-value of less than 0.05 rejects the null hypothesis. For the large sample size, usually the p-value is less than 0.05 which means the null hypothesis will be rejected and the sample will be considered as not normally distributed (Öztuna et al., 2006). Although it might be only slight deviation from normal distribution which might not affect credibility parametric test results (Öztuna et al., 2006; Field, 2013).

To compare independent samples, T-Test compares the differences in mean of scores in each group, and it should be used when data is normally distrusted. However, in non-parametric cases, the Mann-Whitney U-Test can be used, which compares the differences in the rank position of the scores in two groups (Field, 2013). To assess the association between continuous variables, Spearman's Rank Correlation Coefficien" is used for non-parametric variables and Pearson Correlation Coefficient is used for parametric variables. The correlation coefficient is between -1 to +1, while ±1 represent perfect association and 0 represents no linear relationship between two variables (Field, 2013).

In this research, after collecting tracking data from the experiment (see section 4.12), the data has been prepared (see section 4.14) to be statistically analysed by SPSS version 20 (SPSS Inc., Chicago, IL). Shapiro-Wilk test shows that the collected data is mostly nonparametric; hence Mann-Whitney U-Test is used to compare two

independent samples. To assess the association between continuous variables, Spearman's Rank Correlation Coefficient is used. For all tests, a value of "$p<0.05$" is used to suggest statistical significance. Although the data was tested for normality and it is not normally distributed, considering the large sample sizes the parametric tests (including "Independent Samples T-Test" and "Pearson Correlation Coefficient") are also employed and the results are reported for guidance. However, parametric tests have not been used without considering non-parametric tests to reject any of the null hypotheses. When multiple cases are involved, data is presented by the number of cases (N), mean, standard deviation (SD), median, and interquartile range (IQR). When reporting the effects of Online Environment, the rate of change is also presented for both mean and median.

5.5 Effects of Online Environment

In behaviourist psychology, environment has been suggested as one of the most influential factors in forming the behaviour (see section 3.3). When human behaviour as a whole is considered, the environment can be identified as everything outside the human and human mind. The number of elements included in such a broad categorisation can be enormous, and very complex to identify. In online context, users are real humans who are communicating through the Internet with a machine (server). Hence, the associated environment can include anything inside and outside the browser. For example, while the users can be influenced by the technical features of the website, temperature of the room can also influence their behaviour. However, in this experiment, Online Environment is referred to what is presented through the web browser to the users, and any other environmental factors have been ignored.

To investigate the effects of Online Environment, the experiment is divided into two different stages (see section 4.8), each associated with one year of the poetry competition. Some of the aspects of the Online Environments which are provided for the users in each stage of the experiment are different. The difference between the platforms in each stage is further than visual difference, and includes logical differences, which are explained later (see section 5.5.2).

5.5.1 Investigated Null Hypotheses

To investigate association between OB and Online Environment, the following Null Hypotheses are considered, and each have been statistically investigated in the experiment:

- $H1_0$. In different Online Environments, POB is not significantly different for similar group of users.
- $H2_0$. In different Online Environments, AOB is not significantly different for similar group of users.

To test these Null Hypotheses, POB and AOB are compared for research participants in each stage of the experiment. The similarity of users is only assumed based on the fact that the experiment is associated with the same competition, and within the same university faculty (EPS). Also, in both stages of the experiment, the traffic to the website was initiated by sending invitation emails, to same Email Lists (EPS-UG, EPS-PG, EPS-STAFF). These lists include the emails for undergraduate, postgraduate, and staff within EPS. This should be noted that considering the fact that the experiment stages are around one year apart, the actual people who are included in these lists are not necessarily the same. Despite possible difference, the subject group in each stage of the experiment is considered as "similar", and any possible differences are considered as limitations for this study.

5.5.2 Online Environments

For the purpose of this experiment, and based on the deadline enforced by the obtained Ethical Approval, a 23 weeks period has been used as the duration of each stage of the experiment. Each stage starts with the arrival of students to the website after receiving the first invitation email for the EPS Poetry Competition, and it lasts for 23 weeks since then. Some of the metrics which have been used for comparing the experiment stages are defined for a time period, and unless otherwise is stated it is the 23 weeks in that stage of the experiment. For example, although the data is also available for the period between these two stages, when the traffic is compared for Stage 1 and Stage 2, these metrics are calculated for the same amount of time, and any traffic in between is disregarded in the analyses.

The people who receive the invitation email in each of the stages are similar, but not identical. It includes anyone who is included in the Undergraduate, Postgraduate, or Staff Email Lists within the Faculty of Engineering and Physical Sciences during the competition period. The total number of the individuals included in these email lists, and the accurate number of people who received the invitation email is not available, but it has been assumed that it covers a similar group, and any difference resulted from variation of invitees has been ignored and the resulting bias is considered as a limitation.

After first implementation of the platform in 2013 (Stage 1), it was re-designed in 2014 and for the poetry competition which is associated with the Stage 2 of the experiment. The initial design was mainly in black and white, and colours were not used for text or images (Figure 10). One of the major visual differences was the addition of colours to the website's design. This was extended further by asking the competitors to choose a colour for their poems, and this colour was used in association with each poem. This included the colour which was used for the title of the each poem in the "List of Poems", and the background colour in the associated poem page was also reflected by the poems' chosen colour. Figure 10 displays and compares some of the main pages in each stage of the experiment.

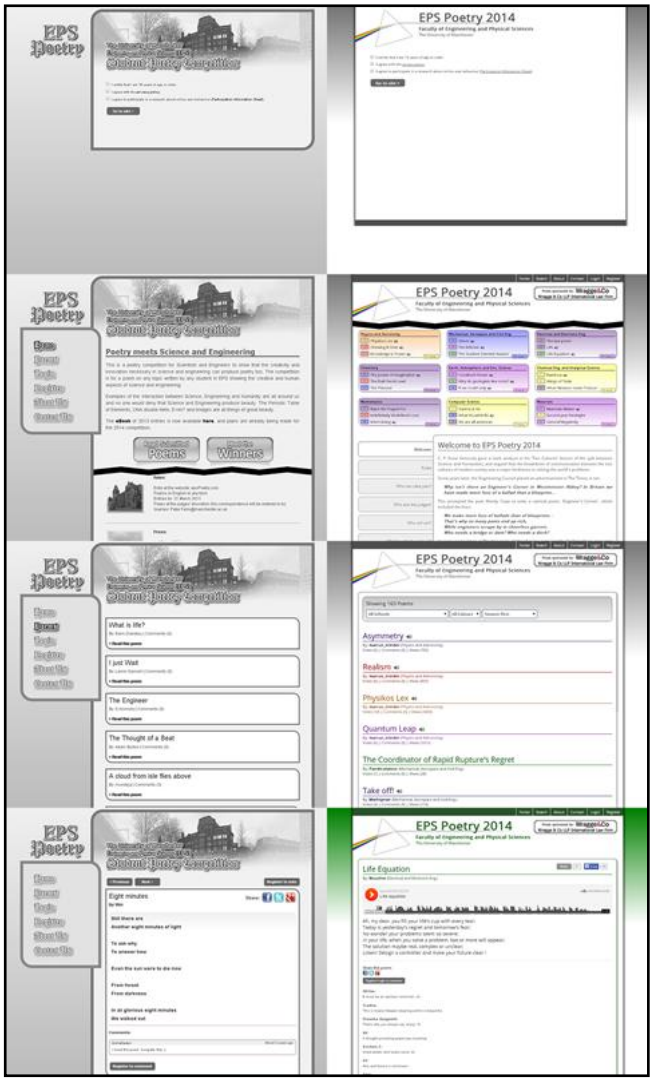

Figure 10 - The visual differences between the main pages in Stage 1 (left) and Stage 2 (right).

Despite some visual difference for the platform in Stage 1 and Stage 2 of the experiment, the overall structure of the website is similar, and after age verification, it consists of 3 main pages; Home Page, List of Poems, and Poem Page.

In addition to the different design, there were also few logical differences in the two stages of experiment. The main change was recognition of popularity of the poems in the short-listing process. This was achieved by introduction of a point system based on number of views and votes for each poem.

In Stage 1, users were asked to vote for their favourite poems, and it was announced that the most popular poem will win a cash prize (£50). In this stage of the experiment, votes were accepted from any Registered User, even if they had not verified their email address. The number of votes for each poem was hidden at this stage, and only the most popular poem was announced at the end of competition. In Stage 2, the users' votes were given a higher recognition. This was by displaying the number of votes for each poem, and giving one point to each vote which was received by each poem. Although unlike Stage 1, email verification was required for voting in Stage 2, the number of people who voted increased from 166 (59 verified) in Stage 1 to 412 (all verified) in Stage 2. In other words, number of people with verified email address who voted in Stage 2 is around 7 times that of Stage 1 (598% increase).

As a new feature in Stage 2, poems were categorised based on the associated school. The EPS faculty has 9 schools, which were presented as boxes in the Home Page (Figure 10). As soon the poems were added to the website, the top 3 poems (based on number of votes) in each school were displayed and linked in the Home Page. Also, the colour from the top poem in each school was used as the background colour of that school's box on the Home Page. In other words, poems with more votes had a chance to be displayed in the Home Page, and ultimately change the colour of the box representing their school.

In Stage 2 the number of views for each poem was also recognised further in comparison to Stage 1. For example, the number of views for each poem was displayed in the "Poem List Page" and a "Sort Functionality" was introduced which could be used to sort the poems based on their number of views. Also, it was stated in the rules that up to 5 points will be awarded to the poems based on number of unique views.

In Stage 1, the users were invited to send an audio version of their poem (for example, in mp3 format) to be linked to their poem. However, only one audio version was received, and even that was later removed from the website. The student who had sent the audio requested the link to be removed and explained that they only sent the audio because they thought it was mandatory. In Stage 2 of the experiment, this was radically

changed. The final component in the pointing system was associated with audio version for each poem. The rules stated that each poem can receive an additional 3 points if an audio version is also submitted. Also, the process was made easier by asking for a SoundCloud Link instead of the mp3 file. The link to the audio version could be submitted alongside the poem, and this would embed a SoundCloud Player in "Poem Page" above the associated poem. Although submitting the audio version was optional, and the suggested number of points was relatively small (equivalent to 3 votes), a total of 58 audio links was received from 49 different competitors. In other words, an easier way of submission and "3 points" persuaded more than a third of the competitors to add audio content to the website. In addition to its effect on final shortlist, the pointing system was also used to choose readers' choice, and a cash prize (£100) was designated to the poem which can achieve the highest number of points.

5.5.3 POB Comparison

The traffic metrics (see section 5.2.1) are calculated for research participants in each stage of the experiment during the investigated 23 weeks.

	Stage 1	**Stage 2**	**Change**
NSV	4,066	8,406	+107%
NRP	1,809	3,950	+118%
NPV	23,698	44,562	+88%

Table 3 - Traffic Comparison between the experiment stages

As displayed in Table 3, all traffic metrics show an increase (ranging from 88% to 118%). It is important to note this traffic is representing both competitors and any other individuals who accessed the website during the 23 week period of each stage. The visit metrics are compared based on data available for all visits, and the resulted are presented in Table 4.

	Stage 1 (N=4,066)				**Stage 2 (N=8,406)**				**Change (%)**		**P-Value**	
	Mean	**SD**	**Median**	**IQR**	**Mean**	**SD**	**Median**	**IQR**	**Mean**	**Median**	**T-Test**	**U-Test**
VD	193.9	507.0	22.0	136.0	224.5	597.7	24.0	158.0	15.78	9.09	0.005	0.219
VL	5.8	10.8	2.0	5.0	5.3	10.0	2.0	5.0	-9.04	0.00	0.007	0.001
VCL	3.9	6.5	2.0	4.0	3.1	3.6	2.0	3.0	-19.60	0.00	0.000	0.000

Table 4 - Visit Comparison

Bounced visits are linked with single page visits, and lack a key part of information for the visit which is its duration. This is due to the fact that entry and exit page are the same for these visits, and visit duration cannot be calculated with this experiments tracking system. On this basis, the previous analysis has been repeated for after excluding the bounced visits, and the results are displayed in Table 5. The result is

similar in both comparisons, and it suggests an increase in visit duration, and a decrease in the total number of viewed pages and number of unique viewed pages in each visit. It should be noted that in theses analyses visits from the same visitors have not been combined.

	Stage 1 (N=2,512)				Stage 2 (N=4,968)				Change (%)		P-Value	
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	Median	T-Test	U-Test
VD	312.1	614.9	92.0	252.0	376.2	739.5	111.0	309.0	20.52	20.65	0.000	0.000
VL	8.8	12.8	5.0	7.0	8.3	12.2	5.0	6.0	-6.08	0.00	0.083	0.105
VCL	5.7	7.7	4.0	4.0	4.6	4.0	3.0	4.0	-18.92	-25.00	0.000	0.000

Table 5 - Visit Comparison (Excluding Bounced Visits)

As explained earlier (see section 4.6.1), ICs have been used in this experiment to identify returning visitors. On this basis, the visits which share the same participation IC on the client side have been combined (i.e. merger based on **`participantID`**), and the visit metrics have been calculated and compared (Table 6).

	Stage 1 (N=1,809)				Stage 2 (N=3,950)				Change (%)		P-Value	
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	Median	T-Test	U-Test
AVD	220.1	411.9	73.0	220.1	183.4	356.1	60.0	191.0	-16.69	-17.81	0.001	0.000
AVL	6.8	12.4	4.0	6.0	5.2	8.7	2.7	5.0	-24.44	-32.57	0.000	0.000
AVCL	4.7	7.9	3.0	3.3	3.3	3.4	2.0	3.0	-30.29	-33.33	0.000	0.000

Table 6 - Visits Comparison (Recognised Participants)

The results show a significant difference between the experiment stages for all measured metrics. According to this analysis the duration and length of visits are reduced in Stage 2. For a group of participants who have gone through registration and verification process, additional information is available and can be used together with the information obtained from the ICs. This improves the identification of returning visitors especially when the visits are made from multiple devices. On this basis, the visits are further combined for Registered Users and the visit metrics are reported for them in Table 7.

	Stage 1 (N=232)				Stage 2 (N=358)				Change (%)		P-Value	
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	Median	T-Test	U-Test
AVD	367.7	364.7	257.5	358.5	314.9	327.7	191.8	293.2	-14.37	-25.53	0.074	0.007
AVL	7.1	5.2	6.0	4.6	7.1	5.7	5.5	4.5	-0.86	-8.33	0.893	0.165
AVCL	4.4	2.7	4.0	2.0	4.4	2.5	4.0	2.0	-1.48	-0.83	0.765	0.614

Table 7 - Visits Comparison (Registered Users)

The only significant differences which can be observed after combining the visits exist between average duration of visits from the Registered Users, for which the average shows 14% decrease.

So far in this section different aspects of visits have been considered and compared for the two stages of the experiment. However, none of these aspects is associated with the order of viewed pages during a visit. Therefore, the navigational metrics are investigated to compare order of pages in each stage. The landing page is the first page which is viewed in a single visit. Table 8 shows the top three landing pages in each stage of the experiment. In each case number and proportion of the visits with this landing page is reported.

	Stage 1			Stage 2		
Rank	**Landing Page**	**Total**	**LPR (%)**	**Landing Page**	**Total**	**LPR (%)**
1	Home Page	2,125	52	Home Page	3,690	44
2	List of Poems Page	720	18	Poem Page	3,395	40
3	Poem Page	578	14	List of Poems Page	574	7

Table 8 - Popular "Landing Pages" in each stage of the experiment

The highest change in LRP can be observed for "Poem Page" (increase from 14% to 40%). In other words, there is an increase in proportion of visitors who are coming to the website following a link to a particular poem. Similar analyses have been done for the popular exit pages, and the results are reported in Table 9.

	Stage 1			Stage 2		
Rank	**Exit Page**	**Total**	**EPR (%)**	**Exit Page**	**Total**	**EPR (%)**
1	Home Page	1,135	28	Poem Page	3,619	43
2	Poem Page	1,032	25	Home Page	2,050	24
3	List of Poems Page	1,013	25	List of Poems Page	1,015	12

Table 9 - Popular "Exit Pages" in each stage of the experiment

Once again the biggest change is associated with "Poem Page", and in Stage 2 it has become the Exit Page for 43% of the visits. In other words, in Stage 2 bigger proportions of visitors start or end their visit on "Poem Page". It does not mean that these visits start and end on the same page, but this is investigated by comparing Bounce Rate (BR). This is the proportion of visits with single page view. The results are calculated for each stage of the experiment, and are reported in Table 10.

Stage 1			Stage 2		
Visits	**Bounced**	**BR (%)**	**Visits**	**Bounced**	**BR (%)**
4,066	1,554	38.2	8,406	3,438	40.9

Table 10 - Bounce Rate Comparison

The results show a small increase in BR. However, this difference is investigated further by comparing popular bounce pages. Popular bounce pages in each stage of the experiment have been analysed and the top three popular bounce pages are reported in Table 11. It should be noted that BPR is calculated based on "total number of bounced

visits" in each stage of the experiment (not "total number of visits"). The results show a significant increase in BPR for "Poem Page". In other words, in Stage 2 more than half of the bounced visits belong to the visitors who come to the website, view a poem, and leave the website without any further interaction.

	Stage 1			Stage 2		
Rank	**Bounce Page**	**Total**	**BPR (%)**	**Bounce Page**	**Total**	**BPR (%)**
1	Home Page	636	41	Poem Page	1,935	56
2	List of Poems Page	490	32	Home Page	1,016	30
3	Poem Page	221	14	List of Poems Page	220	6

Table 11 - Popular Bounce Pages in each stage of the experiment

To further investigate the users' navigational behaviour, in each visit the order of page views has been taken into account. In this analysis, different poems have not been distinguished and "Poem Page" is used for all poems. The top 10 popular Visit Paths during each stage of the experiment are reported in Table 12. The results show the popular paths have changed both in order of popularity and proportion. The biggest change in VPR is linked to the path containing a single poem view. Such visits only contributed to around 5% of visits in Stage 1, but nearly a quarter of visits in Stage 2 only consist of a single poem view.

	Stage 1			Stage 2		
Rank	**Visit Path**	**Total**	**VPR (%)**	**Visit Path**	**Total**	**VPR (%)**
1	Home	635	16	Poem	1,935	23
2	List	490	12	Home	1,017	12
3	Poem	221	5	Poem → Poem	251	3
4	Home → List	134	3	List	220	3
5	Home → List → Poem	102	3	Home → List	151	2
6	Booking Page	62	2	Home → Poem	134	2
7	My Account Page	48	1	Poem → Registration Page	121	1
8	Poem → Registration Page	44	1	Home → Home	107	1
9	Login Page → Verification Page	43	1	Verification Page	73	1
10	Home → Home	34	1	Ceremony Page	71	1

Table 12 - Popular "Visit Paths" in each stage of the experiment

As the final navigational metric, average page view duration (APVD) is calculated for each stage by excluding the exit pages (for which the PVD is not available). The results are displayed in Table 13.

	Stage 1	Stage 2	Change (%)
APVD	37.5	47.2	+25.9

Table 13 - Average Page View Duration Comparison (Excluding Exit Pages)

The results show around 10 seconds (26%) increase in average time spent on each page. APVD is also calculated for popular pages, and the results are reported in Table 14.

	Stage 1	Stage 2	Change (%)
Home Page	55	50	-9.1
List of Poems Page	32	46	43.8
Poem Page	32	45	40.6

Table 14 - Average Page View Duration for specific pages in each stage of the experiment

While the amount of time spent on "Poem Page" and "List of Poems" have both increased more than 40% in Stage 2, the amount of time spent on the "Home Page" shows around 9% decrease.

The loyalty metrics are defined in this experiment by concentrating on identification of returned visitors. The identification is generally based on ICs, but it has been enhanced for Registered Users by using their login history in addition to ICs. However, this information is not available for other visitors. Table 15 displays the results in each stage of the experiment and based on different identification methods which were described.

	Stage 1	Stage 2	Change
Number of Returned Visitors (based on ICs)	473	1,069	+126%
Number of Returned Visitors (based on ICs and login history)	169	267	+58%
RIR (based on ICs)	26.1	27.1	+4%
RIR (based on ICs and login history)	74.8	74.6	-0

Table 15 - Loyalty Comparison

The results show a significant increase in the number of returned visitors; however by considering the number of visits in each stage, Rate of Identified Returns (RIR) has not changed significantly. In these analyses, the number of visits by each visitor has not been taken into consideration, and the visitors have only been distinguished based on having or not having return visits. Further metrics have been used for investigating the number of visits. To reflect on the type of information used for return identification, different metrics have been used. RPVC and ATBRPV are defined when only ICs are used, and when registration information and login history is combined with ICs these are replace by RUVC and ATBRUV. It should also be noted that ATBRPV and ATBRUV are only defined for identified returning visits (Table 16).

The loyalty metrics are significantly different in case of Verified Users and users who have poem. For Verified Users (VU) and Returned Verified Users (VU*) number of visits and time between visits has reduced. However, for users who have poems (HP), number of visits has increased more than 33%.

		Stage 1					Stage 2					Change (%)		P-Value	
Group	Metric	N	Mean	SD	Med	IQR	N	Mean	SD	Med	IQR	Mean	Med	T-Test	U-Test
RP	RPVC	1,809	2.2	9.2	1.0	1.0	3,950	2.1	5.4	1.0	1.0	-5.39	0.00	0.603	0.639
RP*	RPVC	473	5.8	17.5	3.0	3.0	1,069	5.2	9.7	2.0	2.0	-10.55	-33.33	0.479	0.130
RP*	ATBRPV	473	5.5	9.2	1.7	6.2	1,069	5.5	10	1.6	5.2	-0.40	-6.42	0.967	0.437
RU	RUVC	232	9.0	26.9	3.0	7.0	358	8.6	17.6	3.0	6.0	-5.42	0.00	0.806	0.569
RU*	RUVC	169	12.0	31.0	5.0	9.0	267	11.1	19.8	4.0	9.0	-7.59	-20.00	0.733	0.142
RU*	ATBRUV	169	6.1	7.2	3.7	7.4	267	4.4	7.7	1.9	4.1	-27.56	-48.23	0.022	0.000
VU	RUVC	153	12.9	32.5	5.0	9.0	358	8.6	17.6	3.0	6.0	-33.64	-40.00	0.121	0.000
VU*	RUVC	147	13.4	33.1	6.0	10.0	267	11.1	19.8	4.0	9.0	-16.80	-33.33	0.452	0.005
VU*	ATBRUV	147	5.9	6.6	3.8	7.3	267	4.4	7.7	1.9	4.1	-25.51	-49.88	0.038	0.000
HP	RUVC	120	15.3	36.2	7.5	9.0	107	20.4	25.9	12.0	23.0	33.37	60.00	0.219	0.001
HP*	RUVC	116	15.8	36.8	8.0	9.0	100	21.8	26.2	13.0	24.0	37.81	62.50	0.166	0.000
HP*	ATBRUV	116	6.2	6.5	4.4	6.9	100	4.5	5.8	2.9	5.1	-26.90	-34.82	0.048	0.018

Table 16 - Loyalty Comparison (* only returned visitors)

The website associated with this research has been available almost 24/7 during the experiment, and users could access the website at any hour of the day. Figure 11 shows the proportion of visits in each hour of day and each stage of the experiment.

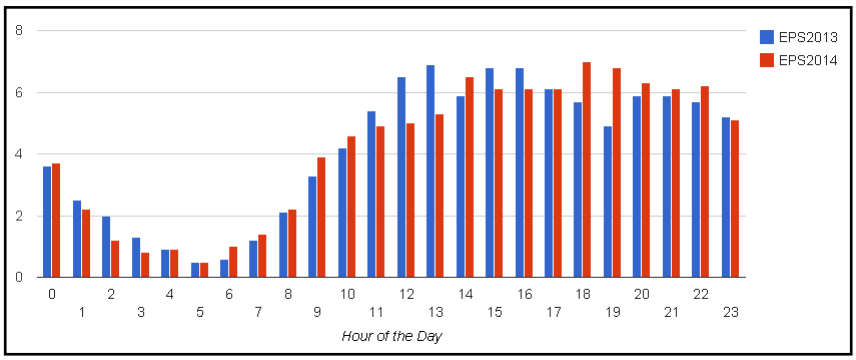

Figure 11 - the proportion of visits in each hour of day and each stage of the experiment

The results show a similar pattern in both stages, and there are fewer visits in early hours of the day (1am to 8am) in both stages of the experiment.

As discussed earlier (see section 5.2.5), specific metrics have been defined for this project. These are inspired by the project's purpose outside this research. The first "EPS Poetry Competition" took place in 2012, and the competitors sent their poems by email. One of the justifications for use of a Web 2.0 platform in association with this competition was for everyone to be able to read these poems. The metrics are defined to investigate the total number of poems and the number of unique poems which are read within each visit or by a specific individual or group of individuals. Initially, the total number of poem page views (NPPV) is calculated and compared for each stage of the experiment (Table 17).

	Stage 1	Stage 2	Change
NPPV	9,492	15,256	+61%

Table 17 - Poem View Comparison

The result of shows around 61% increase in number of poem page views in Stage 2. Furthermore, the number of poems (PPV) and unique poems (UPV) which are viewed in each visit are compared for each stage of the experiment, and the results are shown in Table 18.

	Stage 1 (N=4,066)				Stage 2 (N=8,406)				Change (%)		P-Value	
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	Median	T-Test	U-Test
PPV	2.3	8.2	0.0	2.0	1.8	5.2	1.0	2.0	-22.24	N/A	0.000	0.000
UPV	1.9	6.1	0.0	2.0	1.3	2.2	1.0	1.0	-34.31	N/A	0.000	0.000

Table 18 - Poem View Comparison

The results show a significant difference among the visits in each stage. It is important to emphasise that although the sample size is quite large, it is not normally distributed. In this case while the mean has decreased, the median has increased for both metrics. In other words, the difference in median illustrates that in Stage 1 more than half of visits did not include a poem view, but this was not the case in Stage 2. The rate of visits with no poem views was reduced from 55% in Stage 1 to 34% in Stage 2. Also, in these comparisons the returned visitors have not been taken into account, so unique poems are limited to single views. To investigate the difference further, the visits are combined for the retuning visitors based on ICs, and the Project Specific Metrics are compared in Table 19.

	Stage 1 (N=1,809)				Stage 2 (N=3,950)				Change (%)		P-Value	
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	Median	T-Test	U-Test
APPV	3.2	10.6	1.0	3.0	2.1	5.1	1.0	1.0	-34.48	0.00	0.000	0.000
AUPV	2.6	7.7	1.0	2.5	1.5	2.1	1.0	0.4	-42.02	0.00	0.000	0.039
EUPV	4.0	9.9	1.0	4.0	2.2	3.6	1.0	1.0	-44.22	0.00	0.000	0.165
AEUPV	2.6	7.7	1.0	2.3	1.4	2.1	1.0	0.5	-43.97	0.00	0.000	0.151

Table 19 - Poem View Comparison (Visits are combined based on ICs)

The results show a significant difference and reduction in average poems which have been viewed in each visit.

5.5.4 AOB Comparison

Different factors have been considered within each stage of the experiment to quantitatively compare users' involvement and active participation. These measures are calculated and presented in Table 20. In this table the data is presented in two set of columns. The first three columns are associated with the users who have agreed to take part in this research. Further analysis has been done on this group in other sections of this research. However, the next three columns represent the actual data (i.e. participants and non-participants) as a total and without linking to the individuals or their behaviour.

	Research Participants			Actual Data		
	Stage 1	Stage 2	Change (%)	Stage 1	Stage 2	Change (%)
Registered Users (RUs)	232	358	54	354	824	133
Verified Users (VUs)	153	358	134	210	514	145
Users with Poem (hasPoem)	120	107	-11	166	141	-15
Total Poems (numPoems)	125	127	2	171	163	-5
Users with Comments (hasComment)	13	40	208	16	51	219
Total Comments (numComments)	17	53	212	20	65	225
Users with Audio (has Audio)	0	34	N/A	0	49	N/A
Total Audio Links (numAudios)	0	43	N/A	0	58	N/A
Users who Voted (hasVote)	97	290	199	166	412	148
Verified Users who Voted (hasVote + isVerified)	44	290	559	59	412	598
Total Votes (numVotes)	110	483	339	181	849	369
Votes from Verified (numVotes + isVerified)	53	483	811	69	849	1130

Table 20 - Active Participation Comparison

In most of the investigated metrics, the results show an increase in active participation of the users in Stage 2. However, two of very fundamental active participation factors show up to 15% decrease. These factors are associated with the number of competitors in each stage and the total number of submitted poems. On the other hand, the number of people associated with other forms of participation such is commenting and voting is significantly higher in Stage 2. The most prominent change in participation is associated with the audio link. While in the first stage no one accepted to associate an audio version of their reading with the poem, in the second stage more than a third of users were happy to do so.

The social media connections are measured in this experiment by utilising custom "Share Buttons" and identifying incoming and outgoing traffic associated with these buttons (Table 21).

	Stage 1	Stage 2	Change (%)
NIV	250	630	+152
NOV	19	53	+179
RIV (%)	6.15	7.49	
ROV (%)	0.47	0.63	

Table 21 - Social Media Comparison (Outgoing and Incoming Visits)

The results indicate a considerable increase both in number of incoming and outgoing visits in Stage 2 of the experiment.

5.5.5 Evaluating Hypotheses

Following a behaviourist approach, two hypotheses (H1 and H2) were stated earlier (see section 3.7.1) in regards to association between Online Environment and different dimensions of OB. To test these hypotheses, two null hypotheses ($H1_0$ and $H2_0$) were proposed (see section 5.5.1). To test the null hypotheses, POB and AOB were compared for similar group of users in two different Online Environments. Based on presented

results (see sections 5.5.3 and 5.5.4), it can be argued that both $H1_0$ and $H2_0$ are rejected, and data from this experiment supports H1 and H2. This will be discussed in more details later (see section 6.1.1).

5.6 Effects of Inclusion

The experiment for this research was designed as an exclusive Web 2.0 platform. For registration, an official university email was required, and the email access was verified by emailing the users a Verification Code. Although Active Participation (e.g. sending poems or comments) was only possible for Registered Users, the website was available to the public for Passive Participation (e.g. browsing the website). In following sections the differences in OB are investigated based on Inclusion dimension of VB.

5.6.1 Investigated Null Hypotheses

To investigate association between OB and Inclusion dimension of VB, the following Null Hypotheses are considered, and each have been statistically investigated in the experiment:

- $H3_0$. Inclusion is not significantly associated with POB.
- $H4_0$. Inclusion is not significantly associated with AOB.

To test these Null Hypotheses, research participants are categorised based on what is considered as Inclusion in this experiment, and different elements of their POB and AOB are compared.

5.6.2 Investigated Cases

In this experiment, the users can be divided into three groups; unregistered, registered (i.e. completed registration form), and verified (i.e. registered, and verified their email address). To investigate the effects of Inclusion, the OB for the members of each group is compared. There is a logical difference between the two stages of the experiment, and unlike Stage 1, email verification is mandatory in Stage 2 of the experiment. Consequently, all registered users who have logged-in during Stage 2 are also verified. On this basis the presented results focus on the difference among unregistered and registered (also verified) users in Stage 2 of the experiment. Table 22 shows number of visits (NSV) and page views (NPV) for each group. The ICs and login history is used for this categorisation, and LVU represents Visits and OB which are "Linked to Verified Users". On the other hand, UVU represents Visits and OB which are "Unlinked to Verified Users". It should be noted that due to limitation of ICs, UVU can

also be associated with possible OB from verified users which have not been identified. This is recognised as a limitation for this research.

	UVU	LVU
NSV	5,344	3,062
NPV	25,284	19,278

Table 22 - Number of visits and page views for unregistered and registered users

The verified users in Stage 2 are linked to around 36% of the visits, and 43% of page views. In this section their OB is compared with unregistered users, to investigate the effects of Inclusion in a Web 2.0 platform.

5.6.3 POB Comparison

As explained in previous section, more than a third of the visits in Stage 2 are linked to registered users. The visit metrics for these visits are compared with the other visits and the results are displayed in Table 23.

	UVU (N=5,344)				LVU (N=3,062)				P-Value	
	Mean	**SD**	**Median**	**IQR**	**Mean**	**SD**	**Median**	**IQR**	**T-Test**	**U-Test**
VD	164.9	451.1	17.0	125.0	328.4	780.1	43.0	244.0	0.000	0.000
VL	4.7	9.3	2.0	4.0	6.3	11.2	3.0	6.0	0.000	0.000
VCL	2.9	3.6	1.0	3.0	3.5	3.5	2.0	4.0	0.000	0.000

Table 23 - Visit Comparison for unregistered and registered users

As shown in the results, there is a significant difference between the two groups for duration (VD), number of pages (VL) and number of unique pages (VCL) viewed in each visit. The average duration of visits is 2.7 minutes for unregistered users (UVU) while it is around 5.5 minutes for registered users (LVU). Number of page views and unique page views also show an increase for registered users.

The landing page is the first page which is viewed in each visit. Table 24 shows the top three landing pages in Stage 2, and the results are compared for registered and unregistered users. In each case total number and proportion of the visits with this landing page is reported.

	UVU (Total Visits=5,344)			LVU (Total Visits=3,062)		
Rank	**Landing Page**	**Total**	**LPR (%)**	**Landing Page**	**Total**	**LPR (%)**
1	Poem Page	2,872	54	Home Page	1,695	55
2	Home Page	1,995	37	Poem Page	523	17
3	List of Poems Page	274	5	List of Poems Page	300	10

Table 24 - Popular Landing Page Comparison for unregistered and registered users

The results are different for registered and unregistered users. While more than half of the visits form unregistered users start from the poem page, the registered users are

more likely to start their visit from the home page. In fact, only 17% of the registered users' visits start from the poem page. Similar analyses have been done for the popular exit pages, and the results are reported in Table 25.

	UVU (Total Visits=5,344)			LVU (Total Visits=3,062)		
Rank	**Exit Page**	**Total**	**EPR (%)**	**Exit Page**	**Total**	**EPR (%)**
1	Poem Page	2,878	54	Home Page	1,042	34
2	Home Page	1,008	19	Poem Page	741	24
3	List of Poems Page	436	8	List of Poems Page	579	19

Table 25 - Popular Exit Page Comparison for unregistered and registered users

Similar to previous results, the biggest difference for exit page is associated with poem page. In more than half the visits from unregistered users, poem page is the last page of the visit, while this is less than a quarter of visits for registered users. To further investigate users' OB, for each visit the order of pages which have been viewed have been taken into account. In this analysis, different poems have not been distinguished and "Poem Page" is used for all poems. The top 10 popular paths for registered and unregistered users are reported in Table 26.

	UVU			LVU		
Rank	**Visit Path**	**Total**	**VPR (%)**	**Visit Path**	**Total**	**VPR (%)**
1	Poem	1759	32.9	Home	531	17.3
2	Home	486	9.1	Poem	176	5.7
3	Poem → Poem	234	4.4	List	129	4.2
4	Poem → Registration Page	114	2.1	Home → List	101	3.3
5	List	91	1.7	Home → Home	58	1.9
6	Home → Poem	86	1.6	Home → Poem	48	1.6
7	Home → List	50	0.9	Verification Page	36	1.2
8	Home → Home Page Tabs	49	0.9	Ceremony Page	32	1
9	Home → Home	49	0.9	Home → List → Home	29	0.9
10	Home → List → Poem	42	0.8	Verification Page → Submit Poem Page	27	0.9

Table 26 - Popular Visit Path Comparison for unregistered and registered users

The popular path comparison reiterates the results from landing page and exit page comparisons. Among the visits from unregistered users, around one third consists of a single poem view, however this is less than 6% for registered users. In other hand, registered users are more likely to have single page visits which only contain "Home Page". Overall, 46% of visits from unregistered users contain a single page view, while this is around 32% for registered users.

The Bounce Rate (BR) shows the proportion of visits with a single page view, and results are calculated and compared in Table 27 for registered and unregistered users.

UVU			LVU		
Visits	**Bounced**	**BR (%)**	**Visits**	**Bounced**	**BR (%)**
5,344	2,469	46.2	3,062	969	31.6

Table 27 - Bounce Rate Comparison for unregistered and registered users

Popular bounce pages for registered and unregistered users are reported in Table 28. It should be noted that BPR is calculated based on total number of bounced visits in each group, not the total number of visits.

	UVU (Total Bounced Visits=2,469)			LVU (Total Bounced Visits=969)		
Rank	**Bounce Page**	**Total**	**BPR (%)**	**Bounce Page**	**Total**	**BPR (%)**
1	Poem Page	1,759	71	Home Page	531	55
2	Home Page	485	20	Poem Page	176	18
3	List of Poems Page	91	4	List of Poems Page	129	13

Table 28 - Popular Bounce Page Comparison for unregistered and registered users

To further investigate the differences between navigational behaviour of registered and unregistered users, the page view duration (PVD) is compared for each group. In this analysis every page view in Stage 2 (excluding the "Exit Pages" for which PVD is not defined) is compared, and the results are displayed in Table 29.

	UVU (N=19,873)				LVU (N=16,331)				P-Value	
	Mean	**SD**	**Median**	**IQR**	**Mean**	**SD**	**Median**	**IQR**	**T-Test**	**U-Test**
PVD	41.1	143.6	9.0	21.0	54.5	188.8	7.0	18.0	0.000	0.000

Table 29 - Page View Duration (seconds) Comparison for unregistered and registered users

Average time spent on popular pages is also calculated and compared for registered and unregistered users (Table 30).

	UVU	LVU
Home Page	38	64
List of Poems Page	38	52
Poem Page	43	48

Table 30 - Page View Duration (seconds) Comparison for unregistered and registered users

The results show a significant difference between registered and unregistered users, and in average registered users spend around 15 seconds longer on each page. In case of the home page in particular, registered users spend around 70% longer than unregistered users.

To compare loyalty among the registered and unregistered users, the number of visits (RPVC) and the average number of days between visits (ATBRPV) are compared (Table 31). The results are based on identification of returned participants by ICs, and ATBRPV is only calculated for the participants with more than one visit (RP*).

		UVU					LVU					P-Value	
Group	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
RP	**RPVC**	3,395	1.6	2.9	1.0	0.0	555	5.5	11.9	2.0	3.0	0.000	0.000
RP*	**RPVC**	722	3.7	5.8	2.0	1.0	347	8.2	14.4	3.0	5.0	0.000	0.000
	ATBRPV	722	5.5	9.8	1.4	5.8	347	5.4	10.5	1.9	4.2	0.938	0.038

Table 31 - Loyalty Comparison for unregistered and registered users

While the number of visits is significantly different between the groups (much higher for registered users), the average time between visits is close (around 5.5 days) for both groups.

As a project specific metric, poem view is compared for unregistered (UVU) and registered (LVU) users, and the results are shown in Table 32.

	UVU (N= 5,344)				LVU (N=3,062)				P-Value	
	Mean	**SD**	**Median**	**IQR**	**Mean**	**SD**	**Median**	**IQR**	**T-Test**	**U-Test**
PPV	2.0	5.3	1.0	1.0	1.6	4.9	0.0	2.0	0.001	0.000
UPV	1.4	2.3	1.0	0.0	1.0	1.9	0.0	1.0	0.000	0.000

Table 32 - Poem View Comparison for unregistered and registered users

The results suggest a significant difference between two groups both in regards to the total number of poem views (PPV) and unique poems which are viewed in each visit (UPV). The results show a decrease in both of these variables when the registered users are involved. However, this is when Single Visits are examined, and the fact that registered users are more likely to revisit the website is not reflected in the results. To overcome this limitation and further investigate the effects of Inclusion on project specific metrics, the visits are grouped for both registered and unregistered users. To do so, the poem views and unique poem views have been considered in all visits made by each visitor (identified by ICs), and the results are compared for registered and unregistered users (Table 33).

	UVU (N=3,395)				LVU (N=555)				P-Value	
	Mean	**SD**	**Median**	**IQR**	**Mean**	**SD**	**Median**	**IQR**	**T-Test**	**U-Test**
APPV	2.1	5.3	1.0	1.0	2.0	2.9	1.4	1.9	0.325	0.141
AUPV	1.5	2.2	1.0	0.0	1.4	1.6	1.0	1.2	0.036	0.088
EUPV	1.9	3.0	1.0	1.0	4.0	5.7	2.0	4.0	0.000	0.000
AEUPV	1.5	2.2	1.0	0.0	1.2	1.6	0.9	1.1	0.000	0.000

Table 33 - Poem View Comparison for unregistered and registered users

According to the results, the investigated groups are significantly different in the total unique poems they view during all their visits. According to results, registered users in average view 4 distinct poems during all their visits in Stage 2, however this is just below 2 distinct poems for unregistered users.

5.6.4 AOB Comparison

All of the Web 2.0 features which were provided in the experiment were only available to RUs. In other words, AOB was only provided for the ones who achieved Inclusion dimension of VB. However, sharing the available content is available to any Internet users. Hence, this has been as the only metric to investigate the association between Inclusion and AOB. As explained earlier (see section 5.3.4), the social media connections are measured in this experiment by utilising custom "Share Buttons" and identifying incoming and outgoing traffic associated with these buttons. This functionality (sharing poems) is available to any visitor, and registration is not required. On this basis, the rate of incoming and outgoing visits are calculated and compared for registered and unregistered users (Table 34).

	UVU	**LVU**
NIV	577	53
NOV	28	25
RIV (%)	10.80	1.73
ROV (%)	0.52	0.82

Table 34 - Social Media Comparison for unregistered and registered users

The results show that more than 10% of visits from unregistered users are referred to from SNSs, while this is less than 2% for Registered Users. On the other hand, the rate of outgoing visits is slightly higher for Registered Users, indicating the extra tendency towards sharing the available content from registered users.

5.6.5 Evaluating Hypotheses

Inclusion is presented in this research as the first dimension for VB (see section 2.4.6.1). Two hypotheses (H3 and H4) were stated earlier (see section 3.7.2) in regards to association between Inclusion and different dimensions of OB. To test these hypotheses, two null hypotheses ($H3_0$ and $H4_0$) were proposed (see section 5.6.1). To test the null hypotheses, POB and AOB were compared for users with different levels of Inclusion. Based on presented results (see sections 5.6.3 and 5.6.4), it can be argued that both $H3_0$ and $H4_0$ are rejected, and data from this experiment supports H3 and H4. This will be discussed in more details later (see section 6.1.2).

5.7 Effects of Involvement

In previous section, the effects of Inclusion in the experiment were investigated. As described earlier, in this experiment, membership was only available for individuals with an official university email. After registration, the users were able to become an

Active Participant, and use Web 2.0 features of the website such as sending content, or voting. However, not all Registered Users were actively involved with the platform.

5.7.1 Investigated Null Hypotheses

To investigate association between OB and Involvement dimension of VB, the following Null Hypotheses are considered, and each have been statistically investigated in the experiment:

- $H5_0$. Involvement is not significantly associated with POB.
- $H6_0$. Involvement is not significantly associated with AOB.

To test these Null Hypotheses, research participants are categorised based on what is considered as Involvement in this experiment, and different elements of their POB and AOB are compared.

5.7.2 Investigated Groups

In Stage 1 of the experiment, email verification was not mandatory, and Registered Users could vote for their favourite poems without verifying their email address. However, only after email verification they could send content in form of poem and comments. Figure 12 shows the Venn diagram for Registered Users and their involvement with the community during Stage 1. It should be noted that the numbers reported on this diagram, are only associated with the users who agreed to participate in this research, and the actual numbers are higher.

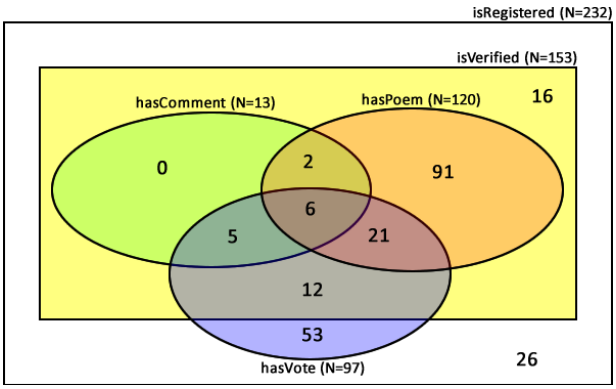

Figure 12 - Venn diagram for Registered Users and their involvement during Stage 1

In Stage 2 of the experiment, email verification became mandatory to increase the credibility of the votes by avoiding any votes from users with invalid emails. Furthermore, in this stage sending audio version of poems became a new type of involvement. Also, by changing the competition rules users' involvement with the

community increased. Figure 13 shows the Venn diagram for Verified Users and their involvement with the community during Stage 2.

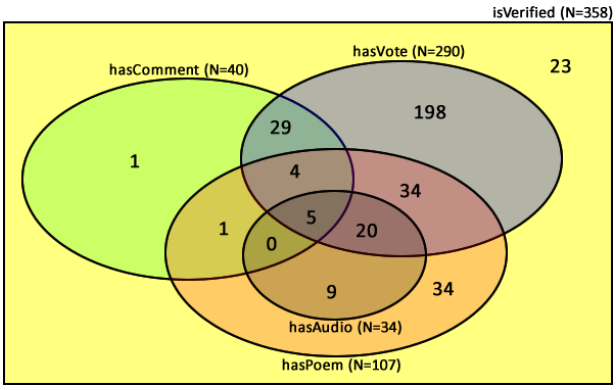

Figure 13 - Venn diagram for Verified Users and their involvement during Stage 2

Ideally, to investigate the effects of Involvement, users with any type of Involvement (or combination of Involvements) should be analysed and compared. However, as displayed in the Venn diagrams (Figures 12 and 13), the number of cases in some groups are very few or unavailable. Also, due to lack of mandatory email verification in Stage 1 and the cash prize for the poem with most votes, some of the users could have been created with invalid email addresses just for the sake of casting a vote. Hence, to investigate the effects of Involvement, the Verified Users in Stage 2 have been investigated, and their behaviour is compared based on their Involvement with the community.

5.7.3 POB Comparison

There are 8,406 visits during Stage 2 from individuals who have agreed to be part of this research. These visits are investigated for their duration (VD), number of pages (VL), and the number of unique pages (VCL). The results are compared for the users based on their Involvement with the community, and the results are displayed in Table 35. The results show a significant difference among the visits from people with or without Involvement. In all cases the results are higher for people who are involved with the community. The highest rate of difference in visit duration can be observed for users who have poem. In average each of their visits takes 6.3 minutes, while this is around 2.8 minutes for the other users. The results also show that the longest average time for visit is associated with people who are involved with the community by sending audio version for their poems. In average they spend around 7.2 minutes on each visit.

		Not Involved					Involved					P-Value	
Involvement	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	**VD**	6,219	169.4	446.3	21.0	136.0	2,187	381.1	879.6	35.0	289.0	0.000	0.000
	VL	6,219	4.9	9.1	2.0	4.0	2,187	6.6	12.3	3.0	6.0	0.000	0.000
	VCL	6,219	3.0	3.6	1.0	3.0	2,187	3.5	3.5	2.0	3.0	0.000	0.000
Comment	**VD**	7,933	215.0	576.1	23.0	151.0	473	383.4	869.2	51.0	318.0	0.000	0.000
	VL	7,933	5.2	10.1	2.0	5.0	473	6.2	9.5	3.0	6.0	0.036	0.000
	VCL	7,933	3.1	3.6	2.0	3.0	473	3.4	3.3	2.0	4.0	0.047	0.000
Audio	**VD**	7,579	201.8	539.2	23.0	147.0	827	432.1	958.9	35.0	384.0	0.000	0.000
	VL	7,579	5.1	9.9	2.0	5.0	827	6.9	11.5	3.0	6.0	0.000	0.000
	VCL	7,579	3.1	3.6	2.0	3.0	827	3.5	3.4	2.0	4.0	0.005	0.000
Vote	**VD**	5,830	174.4	465.5	18.0	132.0	2,576	337.9	810.4	43.0	244.0	0.000	0.000
	VL	5,830	4.8	9.2	2.0	4.0	2,576	6.4	11.6	3.0	6.0	0.000	0.000
	VCL	5,830	3.0	3.6	1.0	3.0	2,576	3.4	3.5	2.0	3.0	0.000	0.000
Any	**VD**	5,413	165.2	450.1	17.0	125.0	2,993	331.6	786.8	43.0	248.0	0.000	0.000
	VL	5,413	4.7	9.3	2.0	4.0	2,993	6.3	11.3	3.0	6.0	0.000	0.000
	VCL	5,413	3.0	3.6	1.0	3.0	2,993	3.5	3.5	2.0	4.0	0.000	0.000

Table 35 - Comparing Visits in Stage 2 based on existence and type of involvement

In the above analyses, visits from Unregistered Users were also included. However, for Registered Users, login history is used in addition to ICs to link the visits from same individuals. The visit metrics are calculated and reported in Table 36 for Registered Users. Once again existence and type of involvement has been used to categorise and compare the users.

		Not Involved					Involved					P-Value	
Involvement	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	**AVD**	251	258.7	311.9	165.0	209.3	107	446.5	327.5	375.1	404.2	0.000	0.000
	AVL	251	6.9	5.8	5.0	4.5	107	7.4	5.4	6.4	5.2	0.386	0.021
	AVCL	251	4.5	2.7	4.0	2.4	107	4.2	1.8	3.9	1.9	0.266	0.807
Comment	**AVD**	318	310.2	332.0	183.3	288.4	40	352.0	292.2	260.0	302.0	0.405	0.042
	AVL	318	7.1	5.8	5.5	4.7	40	6.7	5.1	5.5	4.7	0.612	0.740
	AVCL	318	4.4	2.5	4.0	2.0	40	4.0	2.1	3.5	2.3	0.304	0.532
Audio	**AVD**	324	295.4	321.0	179.0	257.0	34	500.4	338.0	416.4	390.4	0.002	0.000
	AVL	324	7.1	5.9	5.5	4.6	34	6.7	3.5	6.2	4.1	0.614	0.315
	AVCL	324	4.4	2.5	4.0	2.3	34	3.7	1.3	3.5	1.8	0.006	0.278
Vote	**AVD**	68	351.0	330.5	227.9	399.5	290	306.4	327.1	183.5	270.6	0.318	0.330
	AVL	68	6.4	4.9	5.0	5.3	290	7.2	5.8	5.5	4.5	0.218	0.121
	AVCL	68	4.0	2.3	3.7	2.5	290	4.5	2.5	4.0	2.0	0.141	0.219
Any	**AVD**	23	269.2	363.5	130.0	354.7	335	318.0	325.5	194.0	296.1	0.536	0.018
	AVL	23	4.2	4.0	3.1	4.3	335	7.2	5.7	5.5	4.5	0.002	0.000
	AVCL	23	3.3	2.6	2.5	3.0	335	4.4	2.4	4.0	2.0	0.045	0.002

Table 36 - Comparing Visit Metrics for registered users based on existence and type of involvement

Although significant difference can still be observed for some of the metrics such as visit duration, in some cases (such as involvement by voting) no significant difference exists. The most significant difference and the highest rate of change can be observed for users who are involved with sending poem and audio.

To investigate the navigational differences in relation to users' Involvement, Stage 2 visits are investigated, and the bounce rate is compared for groups with and without different types of Involvement (Table 37).

	Not Involved			Involved		
Involvement	**Visits**	**Bounced**	**BR (%)**	**Visits**	**Bounced**	**BR (%)**
Poem	6,219	2,792	44.9	2,187	646	29.5
Comment	7,933	3,304	41.6	473	134	28.3
Audio	7,579	3,197	42.2	827	241	29.1
Vote	5,830	2,612	44.8	2,576	826	32.1
Any	5,413	2,502	46.2	2,993	936	31.3

Table 37 - Bounce Rate Comparison based on existence and type of involvement

The results suggest up to 15% lower bounce rate for the users who are involved with the community. In other words, people who are involved with the community are less likely to have visits with a single page view. Additionally, for the visits which are not bounced (i.e. more than one page is viewed during the visit), the average page view duration (APVD) is compared in Stage 2 visits based on Involvement (Table 38).

	Not Involved					Involved					P-Value	
Involvement	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	3,427	59.0	149.0	20.0	30.6	1,541	83.8	184.5	18.0	62.0	0.000	0.174
Comment	4,629	64.9	157.0	19.9	35.8	339	90.5	209.8	17.8	53.6	0.005	0.609
Audio	4,382	63.5	158.5	19.7	33.3	586	90.6	179.4	20.3	80.7	0.000	0.427
Vote	3,218	62.0	153.4	20.7	33.6	1,750	75.3	174.5	17.0	43.2	0.005	0.000
Any	2,911	62.8	155.4	21.0	33.8	2,057	72.2	169.1	17.0	41.0	0.044	0.000

Table 38 - Comparing average page view duration based on existence and type of involvement

The results show a significant difference for people who have any form of Involvement in comparison to the ones who are not actively involved. In specific Involvement types such as having poem or audio, although U-Test does not show a significant difference, T-Test shows a very significant difference ($p<0.001$ for involvement with poem or audio), and the rate of mean difference is more than 40%. For example, while people without poem spend around 59 seconds on each page, for people with poem this is around 84 seconds.

To investigate the effects of users' Involvement on their loyalty, total number of visits (RUVC) and the average time (days) between visits (ATBRUV) are compared for Registered Users in Stage 2. This should be noted that only cases with more than one visit have been included in this analysis (Table 39). The results show a significant difference between number of visits for people who are involved with poem, audio and votes in comparison to the ones who are not involved. The biggest rate of mean difference exists between users with and without poem. In this case the average number of visits from the involved users is more than 4 times that of uninvolved users.

		Not Involved					Involved					P-Value	
Involvement	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	**RUVC**	167	4.7	10.4	3.0	2.0	100	21.8	26.2	13.0	24.0	0.000	0.000
	ATBRUV	167	4.3	8.7	1.1	3.6	100	4.5	5.8	2.9	5.1	0.808	0.000
Comment	**RUVC**	234	10.7	18.7	4.0	9.0	33	14.1	26.5	4.0	7.0	0.479	0.576
	ATBRUV	234	4.5	8.0	1.9	4.1	33	3.7	4.8	1.7	3.2	0.455	1.000
Audio	**RUVC**	234	9.2	17.5	3.0	5.0	33	25.0	28.2	16.0	22.0	0.000	0.000
	ATBRUV	234	4.5	8.0	1.8	4.3	33	4.0	5.5	2.9	3.1	0.659	0.264
Vote	**RUVC**	50	9.4	9.8	7.0	9.0	217	11.5	21.4	4.0	7.0	0.281	0.042
	ATBRUV	50	5.2	7.8	2.9	6.2	217	4.2	7.7	1.8	3.5	0.430	0.164
Any	**RUVC**	11	5.2	4.0	5.0	8.0	256	11.4	20.1	4.0	9.0	0.001	0.496
	ATBRUV	11	4.5	10.5	1.0	7.3	256	4.4	7.6	1.9	4.1	0.967	0.126

Table 39 - Comparing Loyalty based on existence and type of involvement

The final metrics which are compared in association with POB are the project specific metrics (see section 5.2.5). The number of poems (PPV) and unique poems (UPV) which are viewed in each visit are compared based on existence and type of Involvement (Table 40).

		Not Involved					Involved					P-Value	
Involvement	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	**PPV**	6,219	2.0	5.2	1.0	2.0	2,187	1.4	5.3	0.0	1.0	0.000	0.000
	UPV	6,219	1.4	2.3	1.0	1.0	2,187	0.9	1.8	0.0	1.0	0.000	0.000
Comment	**PPV**	7,933	1.8	5.2	1.0	2.0	473	1.5	4.6	0.0	2.0	0.154	0.000
	UPV	7,933	1.3	2.2	1.0	1.0	473	1.0	1.8	0.0	1.0	0.001	0.000
Audio	**PPV**	7,579	1.8	5.2	1.0	2.0	827	1.6	4.8	0.0	2.0	0.223	0.000
	UPV	7,579	1.3	2.2	1.0	1.0	827	0.9	1.6	0.0	1.0	0.000	0.000
Vote	**PPV**	5,830	1.9	5.2	1.0	2.0	2,576	1.6	5.3	0.0	2.0	0.038	0.000
	UPV	5,830	1.4	2.3	1.0	1.0	2,576	1.0	1.9	0.0	1.0	0.000	0.000
Any	**PPV**	5,413	2.0	5.3	1.0	1.0	2,993	1.6	5.0	0.0	2.0	0.001	0.000
	UPV	5,413	1.4	2.3	1.0	0.0	2,993	1.0	1.9	0.0	1.0	0.000	0.000

Table 40 - Comparing Poem View based on existence and type of involvement

The results show a significant difference between involved and uninvolved users. This significant difference in all cases represents less number of poem views and unique poem views per visit for involved users. However, this does not take into the account multiple visits by the same users. Hence, the visits from registered users are combined and the total unique poems viewed by each user during all visits (EUPV) are compared based on their involvement with the community (Table 41).

		Not Involved					Involved					P-Value	
Involvement	**Metric**	**N**	**Mean**	**SD**	**Med**	**IQR**	**N**	**Mean**	**SD**	**Med**	**IQR**	**T-Test**	**U-Test**
Poem	**EUPV**	251	3.2	4.4	2.0	3.0	107	11.9	9.4	10.0	15.0	0.000	0.000
Comment	**EUPV**	318	5.6	7.0	2.0	7.0	40	7.4	10.2	4.0	8.0	0.298	0.207
Audio	**EUPV**	324	5.0	6.9	2.0	5.0	34	13.3	8.2	14.5	15.0	0.000	0.000
Vote	**EUPV**	68	6.0	6.5	4.5	10.0	290	5.8	7.6	2.0	6.0	0.861	0.615
Any	**EUPV**	23	3.0	4.8	0.0	6.0	335	6.0	7.6	3.0	7.0	0.008	0.000

Table 41 - Comparing total unique poem views based on existence and type of involvement

The results show a significant difference when users are involved by sending poem or audio. The highest rate of difference can be observed between people with or without poem. In average people who have poem, view around 12 distinct poems in their visits, while for other users this is 3.2 distinct poems.

5.7.4 AOB Comparison

To compare users' interaction with social media networks based on their Involvement, rate of incoming visits (coming from SNSs) and outgoing visits (associated with sharing in SNSs) are calculated for users based on their Involvement (Table 42).

	Not Involved					Involved				
Involvement	**Visits**	**NIV**	**NOV**	**RIV (%)**	**ROV (%)**	**Visits**	**NIV**	**NOV**	**RIV (%)**	**ROV (%)**
Poem	6,219	614	30	9.87	0.48	2,187	16	23	0.73	1.05
Comment	7,579	624	52	8.23	0.69	827	6	1	0.73	0.12
Audio	7,933	620	45	7.82	0.57	473	10	8	2.11	1.69
Vote	5,830	580	35	9.95	0.60	2,576	50	18	1.94	0.70
Any	5,413	578	29	10.68	0.54	2,993	52	24	1.74	0.80

Table 42 - Social Media Comparison based on existence and type of involvement

The results suggest a higher rate of incoming visits (RIV) for users who are not involved. On the other hand, rate of outgoing visits (ROV) is generally higher for involved users. Users who are involved with audio have the highest rate of outgoing visits, showing their higher tendency towards sharing the content.

5.7.5 Evaluating Hypotheses

Involvement is presented in this research as the second dimension for VB (see section 2.4.6.2). Two hypotheses (H5 and H6) were stated earlier (see section 3.7.3) in regards to association between Involvement and different dimensions of OB. To test these hypotheses, two null hypotheses ($H5_0$ and $H6_0$) were proposed (see section 5.7.1). To test the null hypotheses, POB and AOB were compared for users with different levels of Involvement. Based on presented results (see sections 5.7.3 and 5.7.4), it can be argued that both $H5_0$ and $H6_0$ are rejected, and data from this experiment supports H5 and H6. This will be discussed in more details later (see section 6.1.3).

5.8 Effects of Influence

In the previous section the effects of Involvement was discussed. The results showed significant differences among most of the investigated factors in regards to OB for involved and uninvolved users. Involvement was analysed by looking at the existence of content (poem, comment, audio) or existence of content evaluation (vote). So far, Involvement has been analysed by looking at nominal variables (e.g. having or not having poem), and the users' Involvement has not been compared in relation to their Involvement's quality and quantity. In this section this will be attempted by focusing on scale (or continuous) and ordinal variables. Additionally, as explained earlier (see section 2.4.6.3) the users' Influence is defined in this research based on the influence

they can have on other users. For example, when a student submits a poem, and it receives 20 votes from other students, one can argue the first student has affected the OB for the other 20 individuals. In this research this is regarded as Influence dimension of VB, and its association with OB is investigated in this section.

5.8.1 Investigated Null Hypotheses

To investigate association between OB and Influence dimension of VB, the following Null Hypotheses are considered, and each have been statistically investigated in the experiment:

- $H7_0$. Influence is not significantly associated with POB.
- $H8_0$. Influence is not significantly associated with AOB.

To test these Null Hypotheses, research participants are categorised based on what is considered as Influence in this experiment, and different elements of their POB and AOB are compared.

5.8.2 Measuring Influence

Within the over taken experiment in this research, the main content which is provided by users is their poem. In this section the behavioural factors for people who have submitted poem will be further investigated based on their influence on the community. Different factors have been used to distinguish users' Influence on the community.

While submitting poem is required for taking part in EPS Poetry Competition, any other form of involvement such as sending audio and comments are optional. Level of Involvement (invLevel) is defined in this experiment for each user as the total number of types of Involvement which is associated with them. The maximum level of Involvement is 4, which is associated with users who have poem, audio, comment and vote.

The quality of submitted poems (as the principle generated content) in measured in this experiment based on the ranks which are received by each poem from two independent judges. Dr Peter Fenn (from School of Mechanical, Aerospace and Civil Engineering) and Dr John McAuliffe (from School of Arts, Languages and Cultures) ranked each poem on a scale of 1 to 5. The total rank received by each poem from both judges is used in this experiment as a measure of quality for the poems. For users with more than one poem, the statistical average is used (hence it is called ppRank or Per Poem Rank).

In this experiment there was no limit on the number of submitted poems from each student, and some have chosen to submit more than one. Also, there was no limit on size of poems, and poems with different sizes have been received. Number of poem (numPoems), and the average number words in each poem (ppPoemWords) are used for each user to quantitatively measure amount of involvement in generating primary content.

The total number of votes received by each poem (excluding any vote from the owner) has been used to measure perceived quality and popularity of each poem. It is presented as "ppRecExtVotes", and it is defined in this experiment for any user who has poem. Also, the total number of comments received by each poem (excluding comments from the owner) has been used to measure external interest, and is presented as "ppRecExtComments", and it is defined in this experiment for any user who has poem.

The total number of shares received by each poem has also been used to measure quality and popularity of each poem. It should be noted that number of clicks on provided "Share Buttons" towards three major SNSs (Facebook, Twitter, GooglePlus) is used for calculating this metric. Registration is not required for sharing, and anyone can click on the Share Buttons. However, to complete the share process, they will need to login to associated SNS (e.g. Facebook) and complete the sharing process. The completion of process cannot be verified, and in this experiment "ppRecShares" is defined for any user who has poem, based on the number of clicks on their poem's "Share Buttons".

For each poem, the number of times it has been viewed and also the number of distinct visits in which that poem has been viewed are calculated, and used to measure popularity. On this basis, "ppRecViews" and "ppRecVisits" are defined for each user who has poem to measure the popularity of their poem.

Facebook users generally indicated their interest towards content by their "Likes". In this experiment number of "Likes" received by each poem has been used as a measure of quality and popularity. It should be noted that Facebook Likes are associated with each URL, and in this experiment, different URLs have been used for each poem for tracking purposes. Also, users can share the poem by copying the actual URL without the tracking identifier in the URL. Consequently, up to four URLs can be found in association with the number of Facebook Likes. These have been used in two forms; the visible number which is displayed in the poem page next to the "Like Button", and the

total number of "Likes" associated with all possible URLs. These measures have been calculated per poem for each user, and are presented as "ppFbButtons" and "ppFbLikes" respectively.

5.8.3 Investigated Group

There are 107 users in Stage 2 of the experiment who have submitted poem, and 2,187 visits are linked to these users. For those users any correlation between OB (e.g. visit duration, number of visits) and the variables associated with Influence (such as number of times their poem has been viewed, or number of votes it has received) is investigated.

5.8.4 POB Comparison

The visits from users who have poem are investigated, and the correlations with different aspects of their Influence on the community with Visit Metrics are reported in Table 43.

	VD		**VL**		**VCL**	
	Spearman	Pearson	**Spearman**	Pearson	**Spearman**	Pearson
invLevel	**0.02**	0.07**	**0.01**	0.02	**-0.02**	-0.02
ppRank	**0.01**	0.06**	**-0.02**	-0.03	**-0.07****	-0.07**
numPoems	**0.11****	0.09**	**0.08****	0.01	**0.07****	-0.00
ppPoemWords	**0.05***	0.02	**0.03**	-0.02	**0.03**	0.03
ppRecExtVotes	**-0.01**	0.12**	**-0.08****	-0.01	**-0.17****	-0.11**
ppRecExtComments	**-0.00**	0.12**	**-0.06****	-0.00	**-0.15****	-0.11**
ppRecShares	**0.02**	0.03	**-0.03**	0.02	**-0.10****	-0.10**
ppRecVisits	**-0.03**	0.12**	**-0.08****	-0.03	**-0.17****	-0.12**
ppRecViews	**-0.04**	0.13**	**-0.08****	-0.01	**-0.17****	-0.13**
ppFbLikes	**-0.03**	0.11**	**-0.09****	-0.00	**-0.17****	-0.12**
ppFbButtons	**-0.02**	0.12**	**-0.07****	-0.01	**-0.16****	-0.13**

Table 43 - Visit Comparison for visits from users with poem is Stage 2 (N=2,187)

While significant correlations can be observed among some of the investigated factors (in many cases $p<0.001$), the amount of correlation is not significant ($-0.18<r<0.18$). However, when the average visit duration (AVD) and the average number of unique pages per visit are calculated for individual users, the amount correlations are relatively higher (Table 44).

While the independently measured quality of poems (ppRank) and the level of involvement (invLevel) are not correlated with Visit Metrics, the perceived quality of the content and attracted interest are correlated with visit duration and length. The results show users who their poem have received more interest are more likely to have lengthier visits with less number of unique pages.

	AVD		AVL		AVCL	
	Spearman	Pearson	**Spearman**	Pearson	**Spearman**	Pearson
invLevel	**0.17**	0.16	**-0.01**	-0.06	**-0.16**	-0.18
ppRank	**-0.12**	-0.05	**-0.08**	-0.13	**-0.08**	-0.06
numPoems	**0.17**	0.27**	**0.11**	0.03	**0.04**	-0.04
ppPoemWords	**0.14**	0.02	**0.07**	-0.03	**0.12**	0.13
ppRecExtVotes	**0.04**	0.16	**-0.16**	-0.14	**-0.33****	-0.27**
ppRecExtComments	**0.13**	0.20*	**-0.08**	-0.05	**-0.34****	-0.19*
ppRecShares	**0.19***	0.11	**-0.08**	-0.05	**-0.22***	-0.17
ppRecVisits	**0.06**	0.14	**-0.12**	-0.12	**-0.33****	-0.29**
ppRecViews	**0.11**	0.20*	**-0.07**	-0.05	**-0.31****	-0.26**
ppFbLikes	**0.23***	0.26**	**-0.09**	-0.09	**-0.28****	-0.23*
ppFbButtons	**0.24***	0.25*	**-0.03**	-0.08	**-0.25****	-0.20*

Table 44 - Visit Comparison for users with poem in Stage 2 (N=107)

To investigate the effects of Influence on navigational behaviour of users, average page view duration (APVD) and bounce rate (BR) are investigated for users in Stage 2 who have poem (Table 45). The results show a positive correlation between investigated factors and users' Influence. In other words, people who have been more influential in respect to attracted interest generally spend more time on each page view, and also more likely to be associated with single page view (or bounced) visits.

	APVD		BR	
	Spearman	Pearson	**Spearman**	Pearson
invLevel	**0.26****	0.15	**0.16**	0.04
ppRank	**-0.07**	0.00	**-0.04**	-0.05
numPoems	**0.13**	0.18	**-0.15**	-0.02
ppPoemWords	**0.11**	0.03	**0.03**	-0.05
ppRecExtVotes	**0.19***	0.20*	**0.29****	0.19
ppRecExtComments	**0.19**	0.22*	**0.18**	0.10
ppRecShares	**0.24***	0.08	**0.11**	0.09
ppRecVisits	**0.11**	0.22*	**0.32****	0.24*
ppRecViews	**0.12**	0.24*	**0.33****	0.22*
ppFbLikes	**0.27****	0.32**	**0.29****	0.23*
ppFbButtons	**0.30****	0.28**	**0.26****	0.18

Table 45 - Navigation Comparison for users with poem in Stage 2 (N=107)

To investigate the effects of Influence on users' loyalty, in Stage 2 all returning users (more than 1 visit) have been considered. The total number of visits (RUVC), and the average time between visits (ATBRUV) have been analysed for these users (Table 46). The results show a significant correlation between number of visits and majority of investigated factors. The highest correlation can be observed for level of Involvement ($r=0.47$, $p<0.001$). In other words users who are differently involved (poem, audio, vote, comment) are more likely to visit the website more frequently. This is also true for users whose poem have received more interest both on epsPoetry.com (vote, comment, view) and on social media (Facebook Likes).

	RUVC		ATBRUV	
	Spearman	Pearson	**Spearman**	Pearson
invLevel	**0.47****	0.30**	**-0.14**	-0.13
ppRank	**-0.10**	0.04	**-0.05**	-0.01
numPoems	**-0.01**	0.01	**0.01**	-0.02
ppPoemWords	**0.03**	-0.04	**-0.03**	0.01
ppRecExtVotes	**0.33****	0.62**	**-0.10**	-0.09
ppRecExtComments	**0.29****	0.69**	**-0.20***	-0.13
ppRecShares	**0.29****	0.36**	**-0.13**	-0.15
ppRecVisits	**0.43****	0.69**	**-0.09**	-0.16
ppRecViews	**0.45****	0.74**	**-0.12**	-0.17
ppFbLikes	**0.41****	0.65**	**-0.23***	-0.20*
ppFbButtons	**0.42****	0.68**	**-0.29****	-0.20*

Table 46 - Loyalty Comparison for returning users with poem in Stage 2 (N=100)

While significant correlation can be observed between number of visits (RUVC) and users' Influence, the average time between visits (ATBRUV) shows less correlation. Nevertheless, the correlation trend seems to be negative. In other words, people with more influence have closer visits (less time between visits). This is particularly significant for the number of "Facebook Likes" which displayed next to the associated poem ($r=-0.29$, $p<0.001$).

It was previously observed that there is a significant difference between the number of poems read by people who have taken part in the competition and the others (see section 5.7). In this section the number of poems and unique poems read by these users are investigated for any correlation with their Influence (Table 47).

	APPV		AUPV		EUPV		AEUPV	
	Spearman	Pearson	**Spearman**	Pearson	**Spearman**	Pearson	**Spearman**	Pearson
invLevel	**0.16**	0.03	**0.05**	-0.09	**0.37****	0.35**	**-0.05**	-0.17
ppRank	**-0.15**	-0.14	**-0.12**	-0.14	**-0.07**	-0.05	**-0.09**	-0.13
numPoems	**-0.02**	-0.03	**0.01**	-0.03	**-0.01**	0.06	**0.01**	-0.02
ppPoemWords	**0.21***	0.03	**0.22***	0.17	**0.17**	0.07	**0.17**	0.10
ppRecExtVotes	**-0.09**	-0.05	**-0.21***	-0.18	**0.15**	0.24*	**-0.31****	-0.22*
ppRecExtComments	**-0.03**	0.00	**-0.18**	-0.11	**0.18**	0.26**	**-0.23***	-0.15
ppRecShares	**-0.11**	-0.02	**-0.21***	-0.19*	**0.09**	0.10	**-0.26****	-0.21*
ppRecVisits	**-0.02**	-0.08	**-0.12**	-0.19	**0.23***	0.28**	**-0.25***	-0.24*
ppRecViews	**0.03**	0.00	**-0.14**	-0.19	**0.24***	0.27**	**-0.27****	-0.24*
ppFbLikes	**-0.05**	-0.01	**-0.18**	-0.12	**0.22***	0.26**	**-0.27****	-0.18
ppFbButtons	**0.00**	0.01	**-0.18**	-0.12	**0.28****	0.30**	**-0.25****	-0.16

Table 47 - Poem View Comparison for users with poem in Stage 2 (N=107)

Poem's length (number of words in the poem) is the most correlating factor with the average number of poems (APPV) and unique poems (AUPV) which are viewed in each visit. The result suggests people with longer poems read more poems in their visit. This should be noted when poems viewed in all visits are considered (EUPV), involvement level (invLevel) shows a significant correlation ($r=0.37$, $p<0.001$). In other words, people who are more actively involved with the community also read more poems. This

is also true for people who received more interest on Facebook. For example number of "Likes" shown on the poem page (ppFbButtons) is correlated with the total poems read, while the number of received votes is not significantly correlated with it.

5.8.5 AOB Comparison

In this experiment different features were provided to facilitate users' Active Participation. In this section any correlation between use of these features and users' Influence is investigated, and the results are presented in Table 48.

	numVotes		**numComments**		**numShares**	
	Spearman	Pearson	**Spearman**	Pearson	**Spearman**	Pearson
invLevel	**0.75****	0.48**	**0.45****	0.39**	**0.06**	0.03
ppRank	**0.15**	0.13	**0.10**	0.06	**-0.02**	-0.07
numPoems	**0.22***	0.65**	**0.12**	0.00	**-0.04**	-0.01
ppPoemWords	**0.07**	-0.06	**0.06**	0.03	**-0.01**	-0.02
ppRecExtVotes	**0.49****	0.44**	**0.24****	0.42**	**0.30****	0.13
ppRecExtComments	**0.23***	0.14	**0.33****	0.17	**0.21***	0.00
ppRecShares	**0.33****	0.21*	**0.08**	0.01	**0.72****	0.61**
ppRecVisits	**0.45****	0.24*	**0.14**	0.24*	**0.27****	0.01
ppRecViews	**0.46****	0.20*	**0.15**	0.19	**0.30****	0.02
ppFbLikes	**0.47****	0.27**	**0.17**	0.20*	**0.47****	0.15
ppFbButtons	**0.46****	0.31**	**0.17**	0.20*	**0.31****	0.17

Table 48 - Active Participation Comparison for users with poem in Stage 2 (N=107)

The results suggest a significant correlation between the number of votes which have been casted by each user and the attention their poem has received. In other words, users whose poems receive more votes, shares, or likes are more likely to also cast more votes. Similar results are found for correlation between number of shares by each users and the attention their poem has received. However, in regards to number of comments, the correlation with Influence is less significant.

5.8.6 Evaluating Hypotheses

Influence is presented in this research as the third and final dimension for VB (see section 2.4.6.3). Two hypotheses (H7 and H8) were stated earlier (see section 3.7.4) in regards to association between Influence and different dimensions of OB. To test these hypotheses, two null hypotheses ($H7_0$ and $H8_0$) were proposed (see section 5.8.1). To test the null hypotheses, POB and AOB were compared for users with different levels of Involvement. Based on presented results (see sections 5.8.4 and 5.8.5), it can be argued that both $H7_0$ and $H8_0$ are rejected, and data from this experiment supports H7 and H8. This will be discussed in more details later (see section 6.1.4).

5.9 Summary

In this chapter, initially some primary definitions for the experiment were stated (see section 5.1). To quantify and measure POB and AOB, a series of metrics in association with OB were mathematically defined for this experiment (see sections 5.2 and 5.3). Thereafter, statistical analyses were employed to provide emperical evidence in regards to the research hypotheses (see section 3.7). Different comparisons were made between independent samples, and associations were investigated between ordinal and scale variables (see sections 5.5, 5.6, 5.7, and 5.8).

In the next chapter, the findings are discussed, and the discovered associations between VB and OB are described. This will be followed by presenting VB as a hierarchy which consists of Inclusion, Involvement, and Influence (see section 6.2).

6 Discussion

The research journey started before this PhD, and with one question: "Why people collaborate on Wikipedia?" However, during the first year, and by reviewing the existing literature, it was realised how big a question this can be. Furthermore, there are many other successful Web 2.0 platforms which use the same resource as Wikipedia, but utilise different implementations of Web 2.0 features (see section 2.1). The freely available efforts of internet users are obtainable by many Web 2.0 projects, yet like any other resource this needs to be managed to optimise the outcome (Ardichvili et al., 2003 Tedjamulia et al., 2005; Hsu and Lin, 2008; van Mierlo, 2014). Hence, this investigation moved towards Influencing Factors for OB. Among the factors which are suggested in the literature, SB was highlighted by many scholars as a significant and important Influencing Factor for OB in Web 2.0 platforms (Roberts, 1998; Teo et al., 2003; Lin, 2007; Lin, 2008; Kim and Zhang, 2011; Zhao et al., 2012). SB is presented and measured in the literature as a qualitative phenomenon (e.g. Bollen and Hoyle, 1990; Hagerty and Patusky, 1995). While this factor can be used by Project Managers to influence OB, one can argue there are some practical problems with SB and its suggested measures in Web 2.0 projects (see section 2.4.5). Firstly, investigation of SB would have to involve interaction with users, and interpreting their perceptions. This like any other qualitative investigation would be subjective, and considering the variety of languages and cultures which can be assumed for global users of Web 2.0 platforms, one may argue that scientific investigation of such audience might not be an option for some of the low budget Web 2.0 projects. Furthermore, another practical problem exists if project managers choose to investigate SB, and make practical decisions about the platform. This problem occurs in the early stages of project, from when the concept is forming to when the platform is being developed by programmers. During these stages the platform does not exist, and hence there are no users. Without users there is no SB to be investigated by project managers. Such shortcomings present a gap in the literature, and this research program attempted to fill this gap (see section 2.4.5).

Aims and objectives for this research (see section 1.2) are based on the identified gap in the literature (see section 2.4.5). In chapter 3, research paradigms were reviewed (see section 3.1), and postpositivism (as defined by Guba and Lincoln, 1994) was identified as an appropriate paradigm for guiding this investigation. Psychological approaches were also reviewed (see section 3.2) and based on the selected paradigm and nature of

investigation, behaviourism (Watson, 1913; Boakes, 1984) was chosen as an appropriate approach in this investigation. While some of the critiques towards behaviourism were highlighted (e.g. Stich, 1998; Chomsky, 1959), its use in this research was justified (see section 3.4). Following this approach, OB was investigated in this research without considering cognitive factors. It should be emphasised that this research does not dismiss the effects of cognitive factors on OB, but investigation of such factors and their effects on OB is beyond the scope of this research program.

Based on existing Web 2.0 features, and the way these features are adopted by users, VB is defined in this research to discriminate and compare users in Web 2.0 projects (see section 2.4.6). There are three dimensions which are proposed in this study for VB, namely Inclusion, Involvement, and Influence. It should be noted that no claims are made in this research about any possible association between SB as defined in the literature, and VB as defined in this research. However, the dimensions which are proposed for VB are inspired by and sometimes associated with the dimensions which are suggested by some scholars for SB. Hence, one might expect some kind of association or similar effects, and this can be investigated in the future.

While many Web 2.0 platforms provide users with the possibility of affecting the available content and many other forms of active participation, many users ignore such possibilities and their interactions with the platform only consist of using the available content. On this basis, two different dimensions are suggested in the literature (Benkler, 2002; Ardichvili et al., 2003 Tedjamulia et al., 2005; Hsu and Lin, 2008; van Mierlo, 2014; Nonnecke and Preece, 2000; Chen, 2004; Preece et al., 2004; Nonnecke et al., 2006; Chen and Hung, 2010), and assumed in this research for Online Behaviour (OB), namely Passive Online Behaviour (POB) and Active Online Behaviour (AOB).

Following postpositivism as the paradigm and behaviourism as the approach, running an experiment was chosen as the appropriate method for this investigation (Saunders et al., 2007). An experiment was designed in association with a Student Poetry Competition in the University of Manchester (see section 4.1). On this basis, a Web 2.0 platform was developed (see section 4.5), and publicised (see section 4.7) to attract traffic. Also, Ethical Approval (see section 4.4) was obtained from the University's Ethics Committees which allowed using a bespoke Tracking System (see section 4.6) to collect behavioural data from visitors who agreed to take part in this investigation (see Appendix 1). The accumulated data is used as the empirical data for this research, and it has been used to test the research hypotheses (see section 3.7).

6.1 Findings

In this section the statistical results which were presented in previous chapter (see sections 5.6, 5.7, 5.8, and 5.9), and were used to reject null hypotheses are described. A summary of results are presented in the following sections, and each section is associated with two of the research hypotheses (see section 3.7). The first section concentrates on Online Environment, and the three subsequent sections each concentrate on one of the proposed dimensions for VB.

6.1.1 Online Environment

Behaviourism assumes a significant role for environment in shaping human behaviour (Watson, 1913; Skinner, 1974; Boakes, 1984). This was investigated in the experiment by comparing POB and AOB for users in two different stages of the experiment. To test the hypotheses, different Online Environments were presented to users in each stage of the experiment (see section 4.8). The environmental differences included both visual differences (e.g. layout, colour, font, etc.) and logical differences (e.g. rules, features, algorithms, etc.). The results showed a considerable increase in Traffic Metrics (see section 5.2.1) in Stage 2 in comparison to Stage 1 of the experiment (see section 5.5.3). For example, in Stage 2, the number of visits more than doubled, and the number of viewed pages showed around 88% increase. Results also showed a significant difference between visit duration and the number of viewed pages in each stage of the experiment (see section 5.5.3). The findings also highlighted a difference in navigational behaviour of the visitors in each stage (see section 5.5.3). This includes change in popular landing pages, exist pages, bounce pages and even the average amount of time which users spend on each page. Loyalty Metrics have also been compared in each stage of the experiment, and the results did not show any significant difference between users in each stage. However, when Verified Users or Users with Poem are considered, the Loyalty Metrics are significantly different in each stage of the experiment (see section 5.5.3). Also, the number of poems which were read by visitors was considered as a Project Specific Metric, and the results showed a significant difference between the two stages of the experiment (see section 5.5.3). The above metrics are all associated with POB, and hence the findings rejects similarity of POB in different Online Environments. Hence, hypothesis H1 (see section 3.7.1) is supported and findings present an association between Online Environment and POB. Furthermore, the results show that AOB has also changed considerably in Stage 2 of the experiment. In this stage, the competition rules were modified, and a pointing system was introduced to

increase engagement, and some of the anticipated metrics have been significantly influenced. For example, the number of votes which are received from Verified Users in Stage 2 is nearly 7 times that of Stage 1. In this research any interaction with SNSs is also considered as AOB (see section 2.2.6), and the results show a significant difference in such activities in each stage of the experiment. Such results rejects the similarity of AOB in different Online Environments, and the data supports hypothesis H2 (see section 3.7.1). This suggests that there is an association between Online Environment and AOB.

6.1.2 Inclusion

In many Web 2.0 platforms, registration is the first step for active participation. In this research, the Inclusion dimension of VB is associated with any connections between users and the platform which enables them to perform AOB in that platform (see section 2.4.6.1).

To investigate the effects of Inclusion, Verified Users in Stage 2 of the experiment were compared with other users in this stage (see section 5.6). The results showed a significant difference in Visit Metrics, and illustrated that Verified Users spend longer on the website and view more pages (see section 5.6.3). They were also significantly different in their navigational habits and the average time they spent on each page (see section 5.6.3). Loyalty was also significantly different, and Verified Users in average visited the website three times more than the others. There was also a significant difference between the amount of poems which were read by each group, as the verified users read more poems in each of their visits (see section 5.6.3). These findings reject the null hypothesis $H3_0$ (see section 5.6.1), and hence data from this experiment supports hypothesis H3 (see section 5.6.5). This suggests an association between the Inclusion dimension of VB and POB. On the other hand, investigating the association between Inclusion and AOB can be problematic considering in many cases Inclusion is prerequisite for AOB. However, while in this experiment, Inclusion is associated with registration, there is one aspect of AOB which is available to all users even if they are not registered. This aspect of AOB is associated with sharing (see section 2.2.6.3) the available content in SNSs. In Stage 2 of the experiment, this activity has been compared for Verified Users and others. The results show that the proportion of visits with sharing activity is higher for Verified Users, indicating the extra tendency towards sharing the available content. On this basis, data from the experiment supports hypothsis H4 (see

section 5.6.5). Consequently, this research suggests an association between the Inclusion dimension of VB and AOB.

6.1.3 Involvement

In many Web 2.0 platforms, registration is required for using the provided Web 2.0 features (e.g. sending comments). When users pass such steps, they can use the provided features and get actively involved with the platform. Although, there is a diverse range of such features, one can argue they all can be represented as connections between users and the platform. In this research, such connections represent the Involvement dimension of VB (see section 2.4.6). The second proposed dimension for VB is Invlovement (see section 2.4.6.2).

To investigate the effects of Involvement, the users in Stage 2 of the experiment were categorised and compared based on existence and type of their Involvement (see section 5.7). Visit duration and number of viewed pages were significantly different for users with or without Involvement (see section 5.7.3). The most significant difference could be observed for people who were involved by sending poems or audio versions (see section 5.7.3). In case of page view duration, non-parametric tests (U-Test) only showed significant difference for people who are involved with voting. However, the parametric test (T-Test) which has been used for guidance, showed a significant difference, and suggests a higher page view duration for people who are involved with the community with any form of Involvement (poem, comment, audio, vote). The Loyalty Metrics also showed a significant difference when users were involved with the community (in particular users with poem or audio). The number of poems and unique poems which are read in each visit are also significantly different for people who are involved with the community. As explained earlier (see section 5.7.5), the null hypothesis $H5_0$ of similarity was rejected, and hypothesis H5 was supported. These findings suggest an association between the Involvement dimension of VB and POB. Furthermore, AOB is also found to be associated with Involvement. For example, the findings show individuals for whom Involvement can be assumed, are more likely to share the available content in SNSs (see section 5.7.4). In this experiment, this can be observed for individuals with any type of Involvement (e.g. poem, comment, audio, vote). Hence, hypothesis H6 is supported, and this research suggests an association between the Involvement dimension of VB and AOB (see section 5.7.5).

6.1.4 Influence

One of the defining attributes of Web 2.0 features is associated with their users' ability to influence the output from underlying software. For example, when a user sends a comment on a blog post, this comment is included in future responses from the server, and can be seen by other users. If another user replies to this comment, it can argued that the first user has influenced the other user's OB. Although, there is a diverse range of Web 2.0 features to facilitate such influence on others, one can argue they all can be represented as connections between the influencing users and the associated platform. In this research, such connections represent the Influence dimension of VB.

The associations between users' Influence and their OB were investigated in this research by introducing specific metrics which can be used in this experiment for discriminating users based on their Influence (see section 5.8.2). It should be noted that many of the suggested metrics are specific to this experiment, however similar metrics can be defined for other Web 2.0 platforms based on their specific features. In this experiment, the quality of the poems was measured based on ranks received from independent panel of judges, and this did not show any significant association with users' OB. However, the perceived quality of the content (measured based on "Votes" and "Facebook Likes" associated with each poem) showed a significant association with users' OB. For example the users whose poem attracted more positive feedback (e.g. Votes, Likes) had longer visits with less number of page views per visit. They also tend to have more single page visits and in average spend longer on each of their page views. They were also significantly different in case of their loyalty with more number of visits and less time between each of their visits (see section 5.8.4). The total number of unique poems which are read by these users was also significantly different, and they tend to read more poems. Similar results were found in regards to other forms of Influence (see section 5.8.2). Consequently, the null hypothesis $H7_0$ (similarity of POB) was rejected. So, the data supports hypothesis H7 (see section 5.8.6). These findings suggest an association between the Influence dimension of VB and POB. Furthermore, the association between Influence and AOB is investigated in this experiment, and the results show significant correlations between users' Influence (e.g. the number of votes they have received) and some of their AOB metrics (e.g. the number of votes they have cast). Similar correlations have been discovered between other representations of Influence and AOB in this research. Consequently, the null hypothesis of similarity

($H8_0$) was rejected, and the data supports hypothesis H8 (see section 5.8.6). This suggests an association between the Influence dimension of VB and AOB.

6.2 Virtual Belonging Hierarchy

In this study, the VB was proposed as a quantitative factor to describe the logical connections which can be assumed between a Web 2.0 platform and its users (see section 2.4.6). Three different dimensions were described for such connections, namely Inclusion, Involvement, and Influence. To examine any association between these dimensions and OB, empirical data was used from an experiment (see chapter 4), and "lack of association" was statistically rejected in this experiment. The findings suggest significant associations between all dimensions of VB and both dimensions of OB.

While each dimension of VB seems to be associated with OB, it should also be emphasised that these dimensions are also associated with each other. In other words, one's Inclusion is required for Involvement, and both Inclusion and Involvement are required for Influence. On this basis, in Web 2.0 projects this research presents VB as a hierarchy of connections between the associated Web 2.0 platforms and their users (Figure 14).

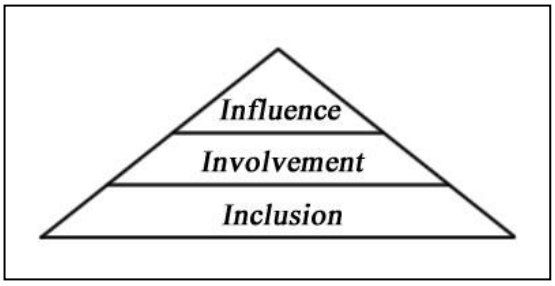

Figure 14 - Virtual Belonging Hierarchy in Web 2.0 Projects

Once again it should be emphasised that due to diversity of Web 2.0 platforms, and their conceptual and functional differences, such dimensions can have different representations in each platform. For example, in Facebook Inclusion can be assumed for anyone who has gone through the registration process and has successfully created an account with Facebook. Thereafter, Involvement can be assumed for them if they use available Web 2.0 features for active participation (e.g. sharing a photo, or commenting on someone else's photo). Finally, Influence can be assumed for them if they influence other users' OB (e.g. when their photo receives comments or likes from other users). While *VB Hierarchy* is easily observable for Facebook and similar platforms, it can be more complex to observe and describe in some other forms of Web 2.0 platforms. For

example, unlike many other Web 2.0 platforms, in case of Wikipedia registration is not required for active participation. For instance, although free and relatively simple registration is available to all users, many of Wikipedia articles can be edited by users even without registration. Hence, one can argue that for anyone who is aware of this rule of Wikipedia, Inclusion can be assumed. This is despite the fact that they might never actually edit an article in Wikipedia. However, if they choose to edit an article or add a new article, Involvement can be assumed for them. Thereafter, the article that they have added (or the section that they edited) can be edited by other users. When a Wikipedia article is edited, one can argue Influence can be assumed for anyone who had previously worked on that article.

6.3 Contributions

This research is one of the few studies which dismiss qualitative data (which is obtained from actual users), and concentrate on a quantitative representation of OB and its association with VB as a quantitative representation of users' connections to Web 2.0 platforms. Also, this research has proposed and undertaken an original event-based method for tracking OB. Finally, the findings provide a practical guidance for project managers, particularly in early stages of Web 2.0 projects. Therefore, this research program contributes to knowledge, research and practice.

This study has identified a gap in the literature in regards to Web 2.0 projects (see section 2.4.5). As an attempt to fill this gap, it has proposed VB as a new factor to explain the logical connections between Web 2.0 platforms and their users (see section 2.4.6). The proposed model has also been tested with an experiment (see chapter 4). This research program as a whole makes the following contributions to the knowledge, research, and practice in the field of Web 2.0 projects.

This study contributes to the knowledge in regard to some of the definitions in the field of Web 2.0 projects. As explained earlier (see section 2.1), previous researchers have provided various definitions for Web 2.0 platforms (e.g. Ahuja and Galvin, 2003; Lin, 2007; Maloney-Krichmar and Preece, 2002; Stockdale and Borovicka, 2006). This study reviews the existing terminologies and definitions, and highlights some of the limitations which are associated with these definitions. On this basis, this study proposes a specific definition for Web 2.0 platforms which extends the existing boundaries in defining Web 2.0 platforms (see section 2.1.3). Furthermore, OB is defined in such platforms by this research from a rather technical perspective (see

section 2.2.1), to remove some of the existing limitations with regard to definitions for OB in the literature. This study also contributes to knowledge by reviewing some of the existing measures for POB and AOB (see sections 2.2.4 and 2.2.7). Additionally, this research highlights the evolution of Web 2.0 as a concept (see section 2.1.2), and points out some of the deficiencies which are caused by the evolving nature of the Web 2.0 technologies. Furthermore, this study contributes to literature by proposing a variety of measures which are mathematically defined for investigating different aspects of OB (see sections 5.2 and 5.3).

This study also contributes to research by proposing and undertaking an event based method for collecting behavioural data in Web 2.0 platforms (see section 4.6.2). Additionally, this study contributes to research by suggesting VB as a quantitative factor which can be used for investigating the connections between Web 2.0 platforms and their users (see section 2.4.6).

This research also contributes to practice by introducing VB (see section 2.4.6) which can be used as an objective and quantitative measure to compare users in Web 2.0 platforms, and to predict their OB. Lack of dependency on human interaction, makes VB an appropriate factor to be used by project managers, particularly in scenarios where linguistic and cultural diversity among users can add to the complexity and cost of qualitative investigations. Furthermore, in some scenarios qualitative data might not be available at all. For example, in the early stages of Web 2.0 projects, in which the concept is forming and the platform is being developed, there are no users, and hence there cannot be any interaction for qualitative investigation. Finally, VB Hierarchy (see section 6.2) can provide a new perspective for project managers with regard to the connections between Web 2.0 platforms and their users. Such a perspective can provide practical directions for project managers in choosing and implementing Web 2.0 features.

6.4 Limitations

This investigation has been authorised by the University of Manchester's Ethics Committees, and hence it is bound by the obtained Ethical Approval (see section 4.4). This approval enforces some practical obligations which have influenced the Web 2.0 platform which is used for the experiment. For example, according to this approval use of ICs and behavioural tracking is only possible after users' consent. Also, according to this approval, the collected behavioural data can only be used for this study if the users

have given prior consent to take part in this research. Furthermore, while the Tracking System has been collecting data for more than three years, the Ethical Approval only allows use of data which has been collected from March 2013 until July 2014. Such limitations can influence the behaviour, or limit the observation, and consequently it could have had an impact on the results. Another limitation which was resulted from the Ethical Approval is associated with emails. The approval dictates that data must be stored within the University and hence the server for the experiment had to be part of the University's network. On the other hand, there are some technical limitations for use of email protocols within the network. Consequently, during the experiment all verification emails had to be sent manually and the unavoidable delay could have affected the outcome.

To investigate the effects of Online Environment on OB, two stages of the experiment have been compared. These stages are associated with a Poetry Competition in two consecutive academic years within the University of Manchester, and hence the investigated subject groups are different. However, when comparing their OB, these subject groups have been assumed to be similar, and any impacts from differences among the subject groups have been ignored.

This investigation is closely linked with a Web 2.0 platform, and a Tracking System which have been used for the experiment. Consequently, some technical limitations are associated with this research, and could have had an impact on the results. First and foremost, ICs have been used by the Tracking System for identification, and this technique can be associated with some limitations (see section 4.6.1). For example, users with multiple devices or multiple browsers, or the ones who disable or remove ICs, or use incognito browsing cannot be identified purely based on ICs. When additional data (such as registration or login history) is available it has been used for further identification. However, such information is not available for all users. Hence, this research is bound with any limitations which are associated with use of ICs. Furthermore, other technical factors such as browser caching, network issues, and server downtime could have affected data collection by the Tracking System. Any inadequacy of the Tracking System in identifying and tracking the users is considered as limitation for this research, and could have affected the results.

Sharing content is identified by this research as AOB, and it has been investigated in the experiment. It should be noted that usually sharing activity for one platform takes place on another platform (e.g. News from BBC can be shared on other platforms such as

Facebook or Twitter). To investigate AOB, custom “Share Buttons” were provided for each poem in the experiment, and clicking on these buttons is considered as sharing. However, the act of sharing is not completed by clicking on the button, and after click the users need to complete the process on the associated SNS (e.g. Facebook). However, the completion of the act cannot be verified by this research, and hence it could have affected the results. This is considered as a limitation for this research.

Finally, the undertaken statistical analyses can only show “association”, and the results can only be considered in the context of prediction rather than causation. In other words, this research cannot claim that the significant differences in OB, which have been observed and reported for the experiment, are caused or influenced by users’ VB.

7 Conclusion

In previous chapters the objectives were set, methods were chosen, data was collected and analysed, and the findings were presented and discussed. In this chapter, initially aim and objectives are reviewed and the achievement of each is described. This will be followed by practical recommendations for Web 2.0 managers, and suggestions for future researchers.

7.1 Reviewing Aim and Objectives

The aim of this research (see section 1.2) was to investigate OB in Web 2.0 platforms, and to present a quantitative model in association with OB which can be used to discriminate and compare users in any Web 2.0 platform without human interaction. The following objectives were identified to achieve this aim:

- To improve understanding of OB (particularly active and passive participation in Web 2.0 platforms) by reviewing the associated literature
- To improve understanding of influencing factors for OB (particularly factors which are associated with active participation in Web 2.0 platforms) by reviewing the associated literature
- To revisit behaviourism (after a century from its introduction) by proposing and applying a quantitative method for investigating human behaviour in an online experiment based on a behaviourist approach
- To propose and test a quantitative model which can be used to discriminate and compare users in any Web 2.0 platform by investigating and categorising suggested Web 2.0 features in the literature
- To suggest practical guidelines for Web 2.0 managers (particularly in the early stages of development in which feedback from users might not be an option) by proposing a quantitative model in association with OB which can be used in any Web 2.0 platform

The first objective was achieved by reviewing the literature in regards to OB (see section 2.2), identifying POB and AOB as its dimensions (see sections 2.2.3 and 2.2.5), and presenting a definition for it (see section 2.2.1). The second objective was achieved

by reviewing the literature in regards to influencing factors for POB and AOB (see section 2.3), and categorising them into cognitive (e.g. trust, commitment, etc.), and functional (e.g. usability, accessibility, etc.). Among the cognitive factors, SB was highlighted by many previous researchers (see section 2.4.4) as an influential and significant factor in Web 2.0 projects. Consequently, SB was investigated further, and a gap was identified in the literature (see section 2.4.5). As an attempt to fill the identified gap, VB was introduced based on a new perspective towards Web 2.0 features, and users' logical connections to Web 2.0 platforms (see section 2.4.6). In this research three dimensions have been identified for VB, namely Inclusion, Involvement, and Influence (see sections 2.4.6.1, 2.4.6.2 and 2.4.6.3). The third objective was achieved by reviewing some of the major psychological approaches (see section 3.2), and focusing on behaviourism (see section 3.3) as the chosen approach. This was followed by reviewing some of its critiques, and thereafter suitability of behaviourism for investigating OB in this research was justified (see sections 3.3.2 and 3.4). Furthermore, a study design was proposed based on behaviourist approach (see section 3.6). The proposed method was applied in an experiment (see chapter 4), and in this experiment the following 8 hypotheses were investigated:

- H1. Online Environment is significantly associated with POB.
- H2. Online Environment is significantly associated with AOB.
- H3. Inclusion dimension of VB is significantly associated with POB.
- H4. Inclusion dimension of VB is significantly associated with AOB.
- H5. Involvement dimension of VB is significantly associated with POB.
- H6. Involvement dimension of VB is significantly associated with AOB.
- H7. Influence dimension of VB is significantly associated with POB.
- H8. Influence dimension of VB is significantly associated with AOB.

As explained earlier (see sections 5.5, 5.6, 5.7 and 5.8), the data from experiments rejects all null hypotheses. Hence, based on this experiment this study supports all above hypotheses. The fourth objective of this research was achieved by presenting VB as a hierarchy (see section 6.2). This model can be used to discriminate and compare users in any Web 2.0 platform by categorising Web 2.0 features, and user connections to the platform (see section 2.4.6). The fifth and final objective for this research was achieved by presenting VB as a predictive factor which can be examined objectively and without human interaction. This can provide guidance for project manager

especially in the early stages of Web 2.0 projects in which human interaction might not be an option. Finally, with proposed model for VB (see section 6.2), and the proposed event-based tracking method (see section 4.6.2), the web manager would be able to investigate OB in real time and without human interaction.

7.2 Practical Recommendations

Following the research which was undertaken, some practical recommendations are listed in this section which can be considered by Web 2.0 managers. These recommendations are not directly derived from the research hypotheses, but can be seen as practical interpretation of VB hierarchy. The recommendations are based on positive association between VB and OB in Web 2.0 platforms. Additionally, the hierarchical aspect of VB and the sequential nature of its levels have been considered.

The first set of recommendations is associated with "Inclusion" dimension of VB. Given its positive association with OB, it is suggested to avoid complex registration forms. Managers should keep away from asking for additional information unless there is a justification for it. For example, registration can be simply done by choosing a unique username and a password. It can become more complex by asking for an email address and to go further and ask for it to be verified. Furthermore, sometimes personal information (e.g. date of birth, gender, home address) are asked during registration. While such information can have marketing value, complex registration forms can reduce the number of people who register. In recent years another method has been used by some platforms which should be considered as an optional alternative to registration. In this method users login to the platform by their Facebook or Google account. Hence, the registration process is dismissed, and in some cases users' information can be obtained from their Facebook or Google account.

The second set of recommendations is associated with "Involvement" dimension of VB. First and foremost, considering the positive effect of "Involvement" on OB, managers should think about providing different methods to engage their users. This dimension can be achieved by a variety of Web 2.0 features (e.g. commenting, voting, posting, tagging), but the managers should choose wisely based on associated context. Also, considering the fact that many users might not be registered, managers should consider providing some forms of "Involvement" for unregistered users. In Wikipedia project, this is achieved by using IP Address (Wikipedia, 2016a). Other platforms such as BBC use a combination of ICs and IP Address to enable their users to "love" available

content without registration. Web managers can also consider providing different sets of Web 2.0 features with different levels of accessibility. For example, "voting" can be based on IP Address and available to everyone, while more sensitive activities such as "commenting" can only be available to registered users. Finally, other users' activities can be used as a measure of popularity. For example, managers should consider promoting the content which receives the highest number of clicks. In many cases the most popular items can be different when considering number of votes and number of clicks. Hence, this can reflect on the home page in separate sections (e.g. most popular and most viewed).

The third and final set of recommendations is associated with "Influence" dimension of VB. The web managers should consider the positive effect of "Influence" on OB (particularly AOB), and facilitate different features to report and highlight users' influence on others. For example, it would be advantageous to include number of views and number of votes for each posted item. One can also suggest providing "sorting facilities" based on number of votes and views to highlight its importance. Furthermore, there are currently different plugins which are provided by popular SNSs (e.g. Facebook, Google Plus, Twitter) which can be used to promote influence. In this case, the number of Facebook Likes or Re-tweets can be displayed next to each item. Final recommendation is to use prominent design for displaying the influence. For example, the font-size which is used by YouTube to show number views is the same size as videos' title, and much bigger than its description.

It should be noted that all above recommendations are subjective, and based on the context, managers should customise them for each online platform.

7.3 Future Work

The described limitations in previous section can provide some directions for future research. Furthermore, this research suggests VB as a predictive factor for OB in Web 2.0 platforms. As described earlier (see section 2.4.6), this factor was built upon the elements which are suggested in the literature for SB. However, the association between VB and SB (or any other cognitive factors) has not been examined or claimed by this research. Future studies might investigate such associations, and the results may be able to discover a link between "emotional feelings" and "logical connections" in Web 2.0 platforms. Moreover, VB has been proposed by this study as a factor which can be used for investigating OB. However, its association with OB has only been examined in one

particular experiment. The proposed model of VB can be tested in other experiments, and with other types of Web 2.0 platforms. The results then could be used to explain the psychological reasons for the observed associations between VB and OB.

Finally, the proposed model of "VB Hierarchy" can be enhanced by future studies. In this case, one would suggest two considerations. Firstly, adding a representation for "Passive Participation prior to Inclusion", and lastly, dividing "Influence" into "Influence on POB" and "Influence on AOB".

References

Ahuja, M. K., & Carley, K. M. (1998). Network structure in virtual organizations. *Journal of Computer-Mediated Communication*, *3*(4).

Ahuja, M. K., & Galvin, J. E. (2003). Socialization in virtual groups. *Journal of Management*, *29*(2), 161-185.

Aldhaban, F. (2012, July). Exploring the adoption of Smartphone technology: Literature review. In *Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET'12:* (pp. 2758-2770). IEEE.

Alexa, (2016). *Alexa - Top Sites in United Kingdom*. [online] Available at: http://www.alexa.com/topsites/countries/GB [Accessed 9 Feb. 2016].

Anant, S. S. (1966). Need to belong. *Canada's Mental Health*, *14*(2), 21-27.

Anderson, J. R. (1990). *Cognitive psychology and its implications* . WH Freeman/Times Books/Henry Holt & Co.

Anderson, P. (2007). All That Glisters Is Not Gold'—Web 2.0 And The Librarian. *Journal of Librarianship and Information Science*, *39*(4), 195-198.

Ardichvili, A., Page, V., & Wentling, T. (2003). Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of knowledge management*, *7*(1), 64-77.

Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., & Wang, X. (2006, April). Talk to me: foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 959-968). ACM.

Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, 0047287512475217.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*.

Balasubramanian, S., & Mahajan, V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce*, *5*(3), 103-138.

Barua, A., Chellappa, R., & Whinston, A. B. (1995). Creating a collaboratory in cyberspace: theoretical foundation and an implementation. *Journal of Organizational Computing and Electronic Commerce*, *5*(4), 417-442.

Bay, E., Hagerty, B. M., Williams, R. A., Kirsch, N., & Gillespie, B. (2002). Chronic stress, sense of belonging, and depression among survivors of traumatic brain injury. *Journal of Nursing Scholarship*, *34*(3), 221-226.

BBC (2007). *Apple's 'magical' iPhone unveiled*. [online] Available at: http://news.bbc.co.uk/1/hi/technology/6246063.stm [Accessed 4 Feb. 2016].

BBC News (2014). *Sony Pictures computer system hacked.* [online] Available at: http://www.bbc.co.uk/news/technology-30189029 [Accessed 12 Dec. 2014].

BBC News (2015). *TalkTalk hack affected 157,000 customers.* [online] Available at: http://www.bbc.co.uk/news/business-34743185 [Accessed 16 Feb. 2016].

BBC News (2016). *How Pokemon Go took over the web*. [online] Available at: http://www.bbc.co.uk/news/technology-36824737 [Accessed 24 Jul. 2016].

Beer, D. D. (2008). Social network (ing) sites… revisiting the story so far: A response to danah boyd & Nicole Ellison. *Journal of Computer-Mediated Communication*, *13*(2), 516-529.

Benkler, Y. (2002). Coase's Penguin, or, Linux and the Nature of the Firm, *Yale Law Journal*, 4(3), pp. 1-73.

Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*,*35*(3), 283-312.

Berners-Lee, T. (2014). *Tim Berners-Lee on Twitter*. [online] Available at: https://twitter.com/timberners_lee/status/511988109211627520 [Accessed 9 Feb. 2016].

Birdsall, W. F. (2007). Web 2.0 as a social movement. *Webology*, *4*(2), 5-11.

Bishop, J. (2007). Increasing participation in online communities: A framework for human–computer interaction. *Computers in human behavior*,*23*(4), 1881-1893.

Blackwell, C., Birnholtz, J., & Abbott, C. (2014). Seeing and being seen: Co-situation and impression formation using Grindr, a location-aware gay dating app. *new media & society*, 1461444814521595.

Boakes, R. (1984). *From Darwin to behaviourism: Psychology and the minds of animals*. CUP Archive.

Bollen, K. A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social forces*, *69*(2), 479-504.

Bornstein, M. H., & Lamb, M. E. (1999). *Developmental psychology: An advanced textbook* . Lawrence Erlbaum Associates Publishers.

Boyd, D. (2006). Friends, friendsters, and myspace top 8: Writing community into being on social network sites.

Boyd, D., Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

BPS, (2014). British Psychological Society. [online] Available at: http://www.bps.org.uk/ [Accessed 19 Jun. 2014].

Buchner, A. G., & Mulvenna, M. D. (1998). Discovering internet marketing intelligence through online analytical web usage mining. *ACM Sigmod Record*,*27*(4), 54-61.

Buchner, A. G., Mulvenna, M. D., Anand, S. S., & Hughes, J. G. (1999, July). An internet-enabled knowledge discovery process. In *Proceedings of the 9th international database conference, Hong Kong* (Vol. 1999, pp. 13-27).

Burby, J., Brown, A., & WAA Standards Committee. (2007). Web analytics definitions. *Washington DC: Web Analytics Association*.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures.*Information systems research*, *12*(4), 346-362.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN systems*, *27*(6), 1065-1073.

Chai, S., & Kim, M. (2012). A socio-technical approach to knowledge contribution behavior: An empirical investigation of social networking sites users. *International Journal of Information Management*, *32*(2), 118-126.

Chang, Y. F., Chen, C. S., & Zhou, H. (2009). Smart phone for mobile commerce. *Computer Standards & Interfaces*, *31*(4), 740-747.

Chen, C. J., & Hung, S. W. (2010). To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information & Management*, *47*(4), 226-236.

Chen, F. C. (2004). Passive forum behaviors (lurking): A community perspective. In *Proceedings of the 6th international conference on Learning sciences* (pp. 128-135). International Society of the Learning Sciences.

Cheung, C. M., & Lee, M. K. (2009). Understanding the sustainability of a virtual community: model development and empirical test. *Journal of Information Science*.

Chiang, I., Huang, C. and Huang, C. W. (2010). Traffic metrics and Web 2.0-ness, *Online Information Review*, 34(1), 115-126.

Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science*. Authors Cooperative.

Chiu, C. M., Hsu, M. H., & Wang, E. T. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision support systems*, *42*(3), 1872-1888.

Choenarom, C., Williams, R. A., & Hagerty, B. M. (2005). The role of sense of belonging and social support on stress and depression in individuals with depression. *Archives of psychiatric nursing*, *19*(1), 18-29.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, 35(1), 26-58.

Chung, J. Y., & Buhalis, D. (2008). Information needs in online social networks. *Information Technology & Tourism*, *10*(4), 267-281.

Church, K., & de Oliveira, R. (2013, August). What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. In*Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services* (pp. 352-361). ACM.

Clarà, M., & Barberà, E. (2013). Learning online: massive open online courses (MOOCs), connectivism, and cultural psychology. *Distance Education*, *34*(1), 129-136.

Clifton, B. (2012). *Advanced web metrics with Google Analytics*. John Wiley & Sons.

Conrady, R. (2007). Travel technology in the era of Web 2.0. In *Trends and issues in global tourism 2007* (pp. 165-184). Springer Berlin Heidelberg.

Constant, D., Kiesler, S., & Sproull, L. (1994). What's mine is ours, or is it? A study of attitudes about information sharing. *Information systems research*,*5*(4), 400-421.

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and information systems*, *1*(1), 5-32.

Cothrel, J., & Williams, R. L. (1999). On-line communities: helping them form and grow. *Journal of knowledge management*, *3*(1), 54-60.

Creswell, J. W. (2007). *Designing and Conducting Mixed Methods Research*, Thousand Oaks, CA, Sage.

Damsgaard, J. (2002). Managing an Internet portal. *Communications of the Association for Information Systems*, *9*(1), 26.

Danaher, P. J., Mullarkey, G. W., & Essegaier, S. (2006). Factors affecting web site visit duration: a cross-domain analysis. *Journal of Marketing Research*, *43*(2), 182-194.

Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), pp. 319-339.

Dergiades, T., Milas, C., & Panagiotidis, T. (2014). Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers*, gpu046.

Dictionary.com. (n.d.). belongingness. *Dictionary.com Unabridged*. [online] TheFreeDictionary.com. Available at: http://www.thefreedictionary.com/belongingness [Accessed 9 Mar. 2016].

DiNucci, D. (1999). Design & New Media: Fragmented Future-Web development faces a process of mitosis, mutation, and natural selection. *PRINT-NEW YORK*, *53*, 32-35.

Duffy, A. M. (2012). *Fellow Travellers: What do users trust on recommender websites? A case study of TripAdvisor. com* (Doctoral dissertation).

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014. *Pew Research Center*, *19*.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication, 11*(2), 415-441.

Emerson, S. L. (1983). Usenet - A Bulletin Board for Unix Users. *Byte, 8*(10), 219.

ENIAC. (1961). *The ENIAC Story*. [online] Available at: http://ftp.arl.mil/~mike/comphist/eniac-story.html [Accessed 25 Jan. 2016].

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Filieri, R. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management, 51*, 174-185.

Finholt, T., & Sproull, L. S. (1990). Electronic groups at work. *Organization Science, 1*(1), 41-64.

Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating a critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest, 13*(1), 3-66.

Fitzgerald, B., & Howcroft, D. (1998). Towards dissolution of the IS research debate: from polarization to polarity. *Journal of Information Technology, 13*, 313-326.

Fricker, R. D. (2008). Sampling methods for web and e-mail surveys. *N. Fielding*, 195-216.

Gage, N. L. (1989). The paradigm wars and their aftermath a "historical" sketch of research on teaching since 1989. *Educational researcher, 18*(7), 4-10.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2), 486-489.

Gomadam, K., Ranabahu, A., Nagarajan, M., Sheth, A. P., & Verma, K. (2008, September). A faceted classification based approach to search and rank web apis. In *Web Services, 2008. ICWS'08. IEEE International Conference on* (pp. 177-184). IEEE.

Gonçalves, B., & Ramasco, J. J. (2008). Human dynamics revealed through Web analytics. *Physical Review E, 78*(2), 026123.

Google (n.d.) Google Analytics Official Website – Web Analytics & Reporting – Google Analytics. [online] at: http://www.google.co.uk/analytics, Accessed on 20/11/2012

Google. (2016). *Algorithms Inside Search*. [online] Available at: https://www.google.com/insidesearch/howsearchworks/algorithms.html [Accessed 28 Mar. 2016].

Greenhow, C., Robelia, B., & Hughes, J. E. (2009). Learning, teaching, and scholarship in a digital age Web 2.0 and classroom research: What path should we take now?. *Educational researcher*, *38*(4), 246-259.

Gu, F., Widén-Wulff, G. (2011). Scholarly communication and possible changes in the context of social media: A Finnish case study. *The electronic library*, *29*(6), 762-776.

Guadagno, R. E., Okdie, B. M., & Kruse, S. A. (2012). Dating deception: Gender, online dating, and exaggerated self-presentation. *Computers in Human Behavior*, *28*(2), 642-647.

Guba, E. G., & Lincoln, Y. S. (1994) Competing paradigms in qualitative research. In N. K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Hagerty, B. M., & Patusky, K. (1995). Developing a measure of sense of belonging. *Nursing Research*, 44(1), 9-13.

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., & Bouwsema, M. (1993). An emerging theory of human relatedness. *Image: The Journal of Nursing Scholarship*, *25*(4), 291-296.

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., Bouwsema, M. and Collier, P. (1992) Sense of belonging: A vital mental health concept. *Archives of Psychiatric Nursing*, *6*(3), 172-177.

Hagerty, B. M., Williams, R. A., Coyne, J. C., & Early, M. R. (1996). Sense of belonging and indicators of social and psychological functioning. *Archives of psychiatric nursing*, *10*(4), 235-244.

Haythornthwaite, C. (2005). Social networks and Internet connectivity effects.*Information, Community & Society*, *8*(2), 125-147.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464-472.

Holmberg, K., Huvila, I., Kronqvist-Berg, M., & Widén-Wulff, G. (2009). What is library 2.0?. *Journal of Documentation*, *65*(4), 668-681.

Hsu, C. L., & Lin, J. C. C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation.*Information & management*, *45*(1), 65-74.

Hurtado, S., & Carter, D. F. (1997). Effects of college transition and perceptions of the campus racial climate on Latino college students' sense of belonging. *Sociology of education*, 324-345.

ICO (2011). *Cookies Regulations and the New EU Cookie Law*. [online] Available at: http://www.alexa.com/topsites/countries/GB [Accessed 12 Sep. 2012].

Igbaria, M., Shayo, C., & Olfman, L. (1998). Virtual societies: their prospects and dilemmas. *Psychology and the Internet. Academic Press, San Diego, CA*.

Ingram, A. L. (1999). Using web server logs in evaluating instructional web sites. *Journal of Educational Technology Systems*, *28*(2), 137-157.

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of documentation*, *54*(2), 236-243.

InternetLiveStats, (2016). *Total number of websites in real time - Internet Live Stats*. [online] Available at: http://www.internetlivestats.com/watch/websites/ [Accessed 9 Feb. 2016].

Itzhaky, H. (1995). Can social work intervention increase organizational effectiveness? *International Social Work*, *38*(3), 277-286.

Jarvis, M. (2004). *Psychodynamic psychology: Classical theory and contemporary research*. Cengage Learning EMEA.

John, N. A. (2012). Sharing and Web 2.0: The emergence of a keyword. *new media & society*, 1461444812450684.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*,*1*(2), 112-133.

Jones, Q. (1997). Virtual-Communities, Virtual Settlements & Cyber-Archaeology: A Theoretical Outline. *Journal of Computer-Mediated Communication*, *3*(3).

Kankanhalli, A., Tan, B. C., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: an empirical investigation. *MIS quarterly*, 113-143.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

Kershenbaum, A. (1993). *Telecommunications network design algorithms*. McGraw-Hill, Inc..

Kim, A. J. (2000). *Community building on the web: Secret strategies for successful online communities*. Addison-Wesley Longman Publishing Co., Inc.

Kim, D. J. and Zhang, D. (2011). The Effects of Sense of Presence, Sense of Belonging, and Cognitive Absorption on Satisfaction and User Loyalty toward an Immersive 3D Virtual World. *Exploring the Grand Challenges for Next Generation E-Business*, 30-43.

Kim, H., Kim, G. J., Park, H. W., & Rice, R. E. (2007). Configurations of relationships in different media: FtF, email, instant messenger, mobile phone, and SMS. *Journal of Computer-Mediated Communication*, *12*(4), 1183-1207.

Kim, J. W., Choi, J., Qualls, W., & Han, K. (2008). It takes a marketplace community to raise brand commitment: the role of online communities.*Journal of Marketing Management*, *24*(3-4), 409-431.

Koh, J., & Kim, Y. G. (2004). Knowledge sharing in virtual communities: an e-business perspective. *Expert systems with applications*, *26*(2), 155-166.

Koh, J., Kim, Y., G., Butler, B. and Bock, G.W. (2007). Encouraging participation in virtual communities. *Communications of the ACM*, 50 (2), pp.69-73.

Kohavi, R., Mason, L., Parekh, R., & Zheng, Z. (2004). Lessons and challenges from mining retail e-commerce data. *Machine Learning*, *57*(1-2), 83-113.

Krauss, S. E. (2005). Research paradigms and meaning making: A primer.*The qualitative report*, *10*(4), 758-770.

Kwan, I. S., Fong, J., & Wong, H. K. (2005). An e-customer behavior model with online analytical mining for internet marketing planning. *Decision Support Systems*, *41*(1), 189-204.

Lampe, C., Ellison, N., & Steinfield, C. (2006, November). A Face (book) in the crowd: Social searching vs. social browsing. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 167-170). ACM.

Landeweerd, M., Spil, T., & Klein, R. (2013). The Success of Google Search, the Failure of Google Health and the Future of Google Plus. In *Grand Successes and Failures in IT. Public and Private Sectors* (pp. 221-239). Springer Berlin Heidelberg.

Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press.

Lee, F. S., Vogel, D., & Limayem, M. (2003). Virtual community informatics: A review and research agenda. *JITTA: Journal of Information Technology Theory and Application*, *5*(1), 47.

Lee, J., Podlaseck, M., Schonberg, E., & Hoch, R. (2001). Visualization and analysis of clickstream data of online stores for understanding web merchandising. In *Applications of Data Mining to Electronic Commerce* (pp. 59-84). Springer US.

Levett-Jones, T., & Lathlean, J. (2008). Belongingness: A prerequisite for nursing students' clinical learning. *Nurse Education in Practice*, *8*(2), 103-111.

Li, X. (2003). A review of the development and application of the Web impact factor. *Online Information Review*, *27*(6), 407-417.

Liao, L. F. (2008). Knowledge-sharing in R&D departments: A social power and social exchange theory perspective. *The International Journal of Human Resource Management*, *19*(10), 1881-1895.

Limayem, M., Khalifa, M., & Frini, A. (2000). What makes consumers buy from Internet? A longitudinal study of online shopping. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *30*(4), 421-432.

Lin, H. F. (2007). The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Research*, *17*(2), 119-138.

Lin, H. F. (2008). Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Information & Management*, *45*(8), 522-527.

Lin, H. F., & Lee, G. G. (2006). Determinants of success for online communities: an empirical study. *Behaviour & Information Technology*, *25*(6), 479-488.

Liu, H., & Kešelj, V. (2007). Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data & Knowledge Engineering*, *61*(2), 304-330.

Liu, Y., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2011, November). Analyzing facebook privacy settings: user expectations vs. reality. In*Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (pp. 61-70). ACM.

London Stock Exchange, (2016). *TalkTalk Telecom Group Plc*. [online] Available at: http://www.londonstockexchange.com/exchange/prices-and-markets/stocks/summary/company-summary/GB00B4YCDF59GBGBXSTMM.html [Accessed 16 Feb. 2016].

Lu, Y., & Yang, D. (2011). Information exchange in virtual communities under extreme disaster conditions. *Decision Support Systems*, *50*(2), 529-538.

Luckerson, V. (2015). Here's How Your Facebook News Feed Actually Works. *Time. Time*, *9*.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in educational research*, *16*(2), 193-205.

Maloney-Krichmar, D., & Preece, J. (2002). The meaning of an online health community in the lives of its members: Roles, relationships and group dynamics. In *Technology and Society, 2002.(ISTAS'02). 2002 International Symposium on* (pp. 20-27). Ieee.

Manago, A. M., Graham, M. B., Greenfield, P. M., & Salimkhan, G. (2008). Self-presentation and gender on MySpace. *Journal of Applied Developmental Psychology*, *29*(6), 446-458.

Markov, Z., & Larose, D. T. (2007). *Data mining the Web: uncovering patterns in Web content, structure, and usage.* John Wiley & Sons.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*(2), 224.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review*,*50*(4), 370.

McCartney, S. (1999). *ENIAC: The triumphs and tragedies of the world's first computer*. Walker & Company.

McLaren, S. (2006). The interrelations between sexual orientation, sense of belonging and dysphoria among Australian women. *Women & Health*, *43*(3), 123-137.

McLaren, S., Gomez, R., Bailey, M., & Van Der Horst, R. K. (2007). The association of depression and sense of belonging with suicidal ideation among older adults: Applicability of resiliency models. *Suicide and Life-Threatening Behavior*, *37*(1), 89-102.

McLoughlin, C., & Lee, M. J. (2007, December). Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era. In *ICT: Providing choices for learners and learning. Proceedings ascilite Singapore 2007* (pp. 664-675).

McMillan, D. W., Chavis, D. M. (1986). Sense of community: A definition and theory', *Journal of community psychology*. *14*(1), 6-23.

Mellet, K., Beauvisage, T., Beuscart, J. S., & Trespeuch, M. (2014). A "Democratization" of Markets? Online Consumer Reviews in the Restaurant Industry. *Valuation Studies*, *2*(1), 5-41.

Michelin (2012). *The Michelin Guide Dictionary*. [online] Available at: http://www.michelintravel.com/_uploads/Red-Guide-ABCs.pdf [Accessed 7 Aug. 2012]

Middlebrook, S. T., & Muller, J. (2000). Thoughts on Bots: The emerging law of electronic agents. *The Business Lawyer*, *56*(1), 341-373.

Miller, P. (2005). Web 2.0: Building the new library. *Ariadne*, *45*(30), 10.

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, *13*(1), 29-39.

Moodle (2016). *Course set up*. [online] Available at: https://docs.moodle.org/31/en/Teacher_quick_guide [Accessed 20 Jul. 2016]

Network Working Group. (1981). *RFC 801 - NCP/TCP Transition Plan*. [online] Available at: https://tools.ietf.org/html/rfc801 [Accessed 25 Jan. 2016].

Niehaves, B. (2007). On episemological diversity in design science: New vistas for a design-oriented IS research?. *ICIS 2007 Proceedings*, 133.

Nonnecke, B., & Preece, J. (2000, April). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 73-80). ACM.

Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, *6*(1), 7-20.

Nuttman-Shwartz, O., & Dekel, R. (2009). Ways of coping and sense of belonging in the face of a continuous threat. *Journal of traumatic stress*, 22(6), 667-670.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, (1), 17-37.

Open Diary (n.d.) Open Diary. http://www.opendiary.com/entrylist.asp?authorcode=C100003&mode=date [Accessed 12 Jun. 2012].

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research*, *2*(1), 1-28.

Öztuna, D., Elhan, A. H., & Tüccar, E. (2006). Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish Journal of Medical Sciences*, 36(3), 171-176.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web.

Papacharissi, Z. and Rubin, A. (2000). Predictors of internet use. *Journal of Broadcasting and Electronic Media*, 44 (2), pp. 175-96.

Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research*, *14*(4), 284-293.

Pizzato, L., Rej, T., Akehurst, J., Koprinska, I., Yacef, K., & Kay, J. (2013). Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction*,*23*(5), 447-488.

Popper, K. (1959) *The logic of scientific discovery*.

Popper, K. (1963) *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge.

Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation for future research. *Journal of Computer-Mediated Communication,10*(1).

Porter, L., & Golan, G. J. (2006). From subservient chickens to brawny men: A comparison of viral advertising to television advertising. *Journal of Interactive Advertising*, *6*(2), 30-38.

Preece, J. (1998). Empathic communities: reaching out across the Web.*Interactions*, *5*(2), 32-43.

Preece, J., Maloney-Krichmar, D., & Abras, C. (2003). History of online communities. *Encyclopedia of community*, *3*, 1023-1027.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior*, *20*(2), 201-223.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. MIT press.

Ridings, C. M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, *11*(3), 271-295.

Ridings, C., Gefen, D., & Arinze, B. (2006). Psychological barriers: Lurker and poster motivation and behavior in online communities. *Communications of the Association for Information Systems*, *18*(1), 16.

Roberts, T. L. (1998). Are newsgroups virtual communities?. in Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 360-367). ACM Press/Addison-Wesley Publishing Co.

Rosenzweig, M. R., Leiman, A. L., & Breedlove, S. M. (1996). *Biological psychology*. Sinauer Associates.

Royo-Vela, M., & Casamassima, P. (2011). The influence of belonging to virtual brand communities on consumers' affective commitment, satisfaction and word-of-mouth advertising: The ZARA case. *Online Information Review,35*(4), 517-542.

Ryan, J. (2010). *A History of the Internet and the Digital Future*. Reaktion Books.

Sangwan, S. (2005, January). Virtual community success: A uses and gratifications perspective. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 193c-193c). Ieee.

Saunders, M., Lewis, P & Thornhill, A. (2007) *Research Methods for Business Students*. Pearson Education Limited, Harlow.

Schenker, J. L. (2008). Michelin Guide goes Web 2.0: the famous red-covered hotel and restaurant guides turn interactive, community-centered, and free, and the effort is boosting book sales, *Business Week*, No.17.

Schneider, K. J., Pierson, J. F., & Bugental, J. F. (Eds.). (2014). *The handbook of humanistic psychology: Theory, research, and practice*. Sage Publications.

Schollmeier, R. (2001, August). [16] A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications. In *p2p*(p. 0101). IEEE.

Schultz, D. E., & Block, M. P. (2015). US online shopping: Facts, fiction, hopes and dreams. *Journal of Retailing and Consumer Services*, 23, 99-106.

Sedgwick, M., & Yonge, O. (2008). We're it','we're a team','we're family'means a sense of belonging. *Rural and Remote Health, 8*, 1021.

Semaw, S., Renne, P., Harris, J. W., Feibel, C. S., Bernor, R. L., Fesseha, N., & Mowbray, K. (1997). 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature, 385*, 333-336.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324), 1343-1372.

Shen, X., Li, D., & Shen, C. (2006). Evaluating China's university library Web sites using correspondence analysis. *Journal of the American Society for Information Science and Technology*, *57*(4), 493-500.

Shu, W., & Chuang, Y. H. (2011). The perceived benefits of six-degree-separation social networks. *Internet Research, 21*(1), 26-45.

Skinner, B. F. (1974), *About Behaviorism*, New York: Vintage.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism*, 1-10.

Smith, L. D. (1986). *Behaviorism and logical positivism: A reassessment of the alliance*. Stanford University Press.

Smith, M. L. (2006). Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and organization,16*(3), 191-211.

Smithers, R. (2011). *Terms and conditions: not reading the small print can mean big problems*. [online] the Guardian. Available at: http://www.theguardian.com/money/2011/may/11/terms-conditions-small-print-big-problems [Accessed 28 Feb. 2016].

Snow, C. P. (2012). *The two cultures*. Cambridge University Press.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, *1*(2), 12-23.

Staats, W. W. (1996). *Behavior and personality: Psychological behaviorism*. Springer Publishing Company.

Stephens, M., & Collins, M. (2007). Web 2.0, Library 2.0, and the hyperlinked library. *Serials Review*, *33*(4), 253-256.

Stich, S. P. (1984). Is behaviorism vacuous? *Behavioral and Brain Sciences*, 7(04), 647-649.

Stich, S. P. (1998). *Deconstructing the mind*. Oxford University Press on Demand.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. Random House.

Tan, P. N., & Kumar, V. (2004). Discovery of web robot sessions based on their navigational patterns. In *Intelligent Technologies for Information Analysis* (pp. 193-222). Springer Berlin Heidelberg.

Tanasa, D., & Trousse, B. (2004). Advanced data preprocessing for intersites web usage mining. *Intelligent Systems, IEEE*, *19*(2), 59-65.

Tedjamulia, S. J., Dean, D. L., Olsen, D. R., & Albrecht, C. C. (2005, January). Motivating content contributions to online communities: Toward a more comprehensive theory. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 193b-193b). IEEE.

Teo, H. H., Chan, H. C., Wei, K. K., & Zhang, Z. (2003). Evaluating information accessibility and community adaptivity features for sustaining virtual learning communities. *International Journal of Human-Computer Studies*, *59*(5), 671-697.

Thode, H. C. (2002). *Testing for normality* (Vol. 164). CRC press.

Time. (2006). *You — Yes, You — Are TIME's Person of the Year*. [online] Available at: http://content.time.com/time/magazine/article/0,9171,1570810,00.html [Accessed 12 Feb. 2016].

Tuten, T. L. (2008). *Advertising 2.0: Social Media Marketing in a Web 2.0 World: Social Media Marketing in a Web 2.0 World*. ABC-CLIO.

van Mierlo, T. (2014). The 1% rule in four digital health social networks: an observational study. *Journal of medical Internet research*, *16*(2), e33.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing & Management*, *40*(4), 677-691.

Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites?. *Journal of the American Society for Information Science and Technology*, *54*(1), 29-38.

Vaughan, L., & Wu, G. (2004). Links to commercial websites as a source of business information. *Scientometrics*, *60*(3), 487-496.

W3 (1990). *WorldWideWeb: Proposal for a HyperText Project*, at: http://www.w3.org/Proposal.html [Accessed on 25 Jan. 2016].

W3, (2015a). *Tim Berners-Lee*. [online] Available at: https://www.w3.org/People/Berners-Lee/ [Accessed 25 Jan. 2016].

W3, (2015b). *What is HyperText*. [online] Available at: https://www.w3.org/WhatIs.html [Accessed 12 Feb. 2016].

Walsham, G. (1995). The emergence of interpretivism in IS research.*Information systems research*, *6*(4), 376-394.

Wang, Y., & Fesenmaier, D. R. (2003). Assessing motivation of contribution in online communities: An empirical investigation of an online travel community. *Electronic markets*, *13*(1), 33-45.

Wang, Y., & Fesenmaier, D. R. (2004). Towards understanding members' general participation in and active contribution to an online travel community.*Tourism management*, *25*(6), 709-722.

Watson, J. B. (1913) Psychology as the behaviorist views it. *Psychological review*, 20(2), 158.

Weaver, K., & Olson, J. K. (2006). Understanding paradigms used for nursing research. *Journal of advanced nursing*, *53*(4), 459-469.

Webb, J. R. (1992) *Understanding and Designing Marketing Research*, London: academic Press.

Webesignerdepot. (2010). *The Psychology of Web Design. Webdesigner Depot*. Available at http://www.webdesignerdepot.com/2010/05/the-psychology-of-web-design/ [Accessed 28 Feb 2016]

Weltevreden, J. W. (2007). Substitution or complementarity? How the Internet changes city centre shopping. *Journal of Retailing and consumer Services*,*14*(3), 192-207.

Wikimedia Report Card, (2016). *Limn*. [online] Available at: http://reportcard.wmflabs.org/ [Accessed 4 Mar. 2016].

Wikipedia, (2016a). *Welcome unregistered editing*. [online] Available at: https://en.wikipedia.org/wiki/Wikipedia:Welcome_unregistered_editing [Accessed 9 Feb. 2016].

Wikipedia, (2016b). *Wikipedians*. [online] Available at: https://en.wikipedia.org/wiki/Wikipedia:Wikipedians [Accessed 4 Mar. 2016].

Wolverhampton (n.d.). Virtual learning via WOLF. [online] Available at: https://www.wlv.ac.uk/study-here/student-life/study-facilities/it-facilities/virtual-learning-via-wolf/ [Accessed 20 Jul. 2016]

Zhao, L., Lu, Y., Wang, B., Chau, P. Y. and Zhang, L. (2012). Cultivating the sense of belonging and motivating user participation in virtual communities: A social capital perspective, *International Journal of Information Management*.

Appendix 1. Participant Information Sheet

You are being invited to take part in a research study. Thank you for taking the time to read this document. In this document the reasons for carrying this research will be explained, together with your role in this research as a participant. Participation in this research is voluntary and you are under no obligation to participate.

What is the purpose of this study?

The purpose of this study is to determine the effects of "Sense of Belonging" on "Online User Behaviour". To do this, the online behaviour will be monitored for participants. This data will then be analysed to highlight different trends which exist within groups with different levels of sense of belonging. The data will be anonymous, and no personal or identifying data will be used for this study.

This study is being carried out as part of a PhD project at the University of Manchester. Hamed Mohagheghi-Fard is the PhD student and he is supervised by Dr Peter Fenn and Dr Oscar De bruijn. These people make up the research team and their contact details are listed at the end of this information sheet.

Why was I asked to take part?

This research uses a Web Application as the platform for collecting data. This Web Application is available online, and all of its users have been offered to take part in this experiment.

Do I have to take part?

Your participation is entirely voluntary. It is up to you decide whether or not to take part. You can consult friends and family before taking part in the experiment.

If you decide to take part, you can withdraw from the study at any time without giving a reason. The contact details of the research team are given at the end of the information sheet, in case you have any additional questions or if you decide to stop taking part.

Can I use the website if I do not participate in the research?

Yes. You can use the website without participating in the research. In this case, no additional data will be collected or used in the study.

What will happen if I withdraw from the experiment?

You can withdraw from the research at any time. If you decide not to take part, any collected behavioural data which is associated with this research will be deleted and will not be used in the study.

Are there any benefits to my participation?

There are no direct benefits to your participation in this study. However, taking part will give us a better understanding of the patterns which exist in online behaviour, and the techniques which can be used to influence online behaviour.

Are there any risks involved in participating?

This is a benign research and there are no known risks associated with participation. During the experiment, you will be using the website the same as non-participants. Everyone who takes part in the study will remain anonymous, and you can stop taking part at any time.

Will my taking part in the study be kept confidential?

All of the data associated with your participation is completely confidential. Data will be stored on a password protected computer, and it will not be shown or presented to anyone outside the research team. We have to keep the data for 5 years, so that research reports can be made and the accuracy of information can be checked. During this time the data will be stored encrypted in a locked filing cabinet at the University of Manchester. After 5 years, all information will be destroyed.

Prior to the data analysis phase, all personal details will be removed so that no-one will know who you are. You are free to talk about your participation in this study with your family and friends if you wish to do so.

What happens if I decide to take part?

If you decide to participate in this study, your online behaviour will be stored while you use the website. This includes the pages you view, the time you spend on them, and also the website functionalities that you use during your visit.

What will happen to the results of the research?

Results from the study may be published in journals or presented at conferences. No information that could identify participants will be included in the results.

What do I need to do next?

You are not required to do anything, apart from using the website as you would normally do.

Consent:

Please note that by participating you are also giving consent.

Further information and contact details:

If you have any questions regarding the study, please contact a member of research team.

Hamed Mohagheghi-Fard, PhD Student

Email: hamed.mohagheghi-fard@postgrad.manchester.ac.uk

Dr Peter Fenn

Email: peter.fenn@manchester.ac.uk

Dr Oscar De bruijn

Email: o.de-bruijn@manchester.ac.uk

How do I raise any issues or concerns about the research?

If there are any issues regarding this research that you would prefer not to discuss with members of the research team, please contact the Research Practice and Governance Co-ordinator by either writing to 'The Research Practice and Governance Co-ordinator, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester M13 9PL', by emailing: Research-Governance@manchester.ac.uk, or by telephoning 0161 275 7583 or 275 8093.

Appendix 2. Privacy Policy

We are committed to protecting your privacy online. We use the information we collect about you to provide services to you when you visit epsPoetry.

It is your responsibility to read and understand this Privacy Policy. If you do not agree to all of this Privacy Policy and do not wish to be bound by it, you are not authorised to access or otherwise use epsPoetry.

What information we collect

Upon registration we ask you to submit personal information to us. You are under no obligation to provide such information. However, if you should choose to withhold requested information, we may not be able to provide you with certain services.

We also collect behavioural data to monitor, develop and improve epsPoetry, and your experience. Also, if you agree to participate in online behaviour research, the collected data may be used as part of the research.

How we store your information

Information you submit via epsPoetry is sent to a computer located in the University of Manchester.

By submitting information, you agree to the collection, transfer, storage or processing of information about you for the purposes described in this Privacy Policy. We will take all reasonable steps to ensure that information about you is treated securely in accordance with this policy and that all information you provide to us is stored on secure servers.

Limitation of liability

EpsPoetry may contain links to other websites where information practices may be different to ours. You should consult the other websites' privacy notices as we are not responsible for and have no control over information that is submitted to, or collected by, these third parties.

Legal Disclaimer

We are not responsible nor will be liable to you in any way for events beyond our direct control. We cannot guarantee nor do we represent that our performance will be error free, and to the maximum extent permitted by law we will not be liable for any direct,

indirect, incidental or consequential damages or other loss suffered by you and relating to the use or release of your personal information.

Updates

We reserve the right, from time to time, and at our sole discretion, to change or update this Privacy Policy.

All changes to this Privacy Policy will be published on this page. Upon publication, each change will become effective and you will be deemed to be aware of and bound by it. You should therefore review this Privacy Policy regularly to ensure that you are up-to-date with the current terms of the Privacy Policy.

Cookies

Cookies are pieces of information that a website transfers to your hard drive to store and sometimes track information about you. Most web browsers automatically accept cookies, but if you prefer, you can change your browser to prevent that. However, you may not be able to take full advantage of a website if you do so. Cookies are specific to the server that created them and cannot be accessed by other servers, which means they cannot be used to track your movements around the web. Although they do identify a user's computer, cookies do not personally identify users and passwords and credit card information are not stored in cookies.

We may use cookies to:

- identify who you are and to access your account information;
- estimate our audience size and patterns;
- track preferences and to improve and update epsPoetry.

For more information on cookies and how to disable them you can consult the information provided by the Interactive Advertising Bureau UK at www.allaboutcookies.org

Security

Due to the nature of the Internet, we can't guarantee the protection of your personal information and we can't be responsible for any outcomes resulting from a breach of security when epsPoetry is used. We're confident in our security, and it is always a top priority to ensure we do not get any problems.