Ethics in using students as subjects in software engineering experiments

University of Oulu
Faculty of Information Technology and Electrical Engineering/ M3S
Master's Thesis
Benjamin Ato Enyan
November, 2016

Abstract

The ethical issues that researchers encounter in experiment are usually debated upon. Ethical issues are scrutinized and reported in the law, medicine and psychology areas of study. Independent boards are commonly employed to vet the ethical aspect of experiments in the above mentioned areas of study. However, in the software engineering area of research, it is not entirely like that as researchers just follow the guidelines of their respective institutions. Ethical principles of right and wrong, codes of conducts, codes of rules, and external standards that govern an organization, must be considered and followed during the research.

Experiments must be designed on utmost ethical rules and values. These ethical rules must be followed and reported, as researchers conduct their experiments. Ethics in research is inevitable however; its application in software engineering might not be direct. Four fundamental principles of research ethics being informed consent, beneficence, confidentiality, and scientific value are the major pointers of ethical issues in SE experiment.

The focus of this thesis was to find out the ethically related issues reported by researchers while they conducted their research on students. This study used a systematic mapping review method, to gather and analyse reported experiments that were conducted with students as subjects, by classifying and identifying ethical issues that are reported in software engineering experiment over the past few years. The studies reviewed were published within the years 2000 and 2016, and from 12 journals and conference proceedings. The search strategy identified 2512 papers, of which 70 were identified as relevant or primary studies.

The findings revealed that researchers in SE research pay minimal attention to issues of ethics. No reporting guidelines developed over the years suggesting that researchers are obligated to report issues of ethics. However, one aspect seems to be considered most in the reports; beneficence, where researchers seek to motivate the students to participate in their study. In a nutshell, experiment in SE field of study does not have higher risk on its human subjects as compared to other fields of study. Ethical issues are hardly debated or never mentioned in the SE experiment publication. In the near future, experiment reporting guidelines, must be developed in order to promote or suggest to researcher to give details about ethical issues, especially when it involves human (students) subjects.

Keywords
Ethical issues, Experiment, Students as subjects, Software engineering experimentation, code of ethics, informed consent, voluntariness, confidentiality, Researcher

Supervisor
Adjunct Professor Raija Halonen, Muhammad Ovais Ahmad.

Foreword

I would want to express my in-depth gratitude to all the people who have supported me during my studies and also especially the writing of my thesis. This period has been very challenging but with your encouragement, I say thank you to all especially Pastor Dionysius Florack Eshun, Arisekola Akanbi, Pastor Collins Nsiah Peprah, Pastor Kaj Kiviniemi and the Ny Start Pentecostal church as a family. Special appreciation goes to Adjunct Professor Raija Halonen and Muhammad Ovais Ahmad as my supervisors for the time and suggestions. I truly appreciate your support. This is especially dedicated to Auxil, Kuukua and Ohenba.

Contents

1. Introduction

Decisions made by professional software engineers and researchers have effects on the world a lot, as many operations, transactions and processes in the world today are controlled by information systems. It, therefore, creates the urgency to find out how well these researchers uphold the code of ethics or conducts at any level of operation, as these researchers seek to find solutions to problems and meet client requirements in Software Engineering (SE). (Stephenson, 2010.)

However, the growth in the SE field largely depends on how fast researchers work to find solutions to existing and future problems. Also, how further into the future these researchers are able to project or predict. In the attempt to solve the problems, experiments are conducted. As in any other field, researchers in SE do investigate or experiment in order to verify and validate issues of concern and relevance. Information Systems have become essential in the lives of people. Ethics and morality play very important roles as organizations depend on software engineers to develop information systems and application. (Stephenson, 2010.)

Ethics are principles of right and wrong (Laudon & Laudon, 2010 pp. 151.) Kuhar (2006) defined ethics as a code of conducts or rules while Gotterbarn (2001), defines it as external standards which govern an organization. It is pointed out that experiments must be designed on utmost ethical rules and values. These ethical rules must be followed and reported by researchers as they conduct their experiments. However, the researchers' perspective of ethical and moral issue is a thing of concern to the organizations. If the level of understanding of ethics is not considered, ethically controversial systems would be developed and used sub-optimally (Charlesworth & Sewry, 2002; Svahnberg, Aurum & Wohlin, 2008b). (Abrahamsson, Kettunen & Fagerholm, 2010.)

1.1 Purpose

Experiment in any field, is expected to follow some code of conducts (Abrahamsson et al., 2010). Researchers are however, expected to acknowledge the code of ethics. In an experimental environment where human beings are used as subjects, there are ethical conducts provided by bodies such as the British Psychological Association (BPA), American Psychological Association (APA). Researchers must endeavour to follow these given protocols that pertain to experimentation with students as subjects (BPA, 2016).

It is expected of experimenters to mention rules of ethics observed, in their reports. This research will enlighten the understanding of moral and ethical issues in software development. It seeks to identify ethical issues in an experimental environment, where students are used as test subjects by researchers in SE. The study aims at finding out the reported ethical issues in using students as subjects in SE experimentation. It will highlight moral and ethical issues confronted in information system development, hence

enriching researcher's insight on the use of students as subjects and the ethical issues involved. This study takes a further look into the ethical issues in using students for experimentation in the SE field, as in most fields, experimentation must follow a set of rule thus ethics.

1.2 Motivation

Organizations or industries have confidence as they depend on software developers, who produce applications that meet the needs of clients and stakeholders (Stephenson, 2010). It is difficult to recruit professional software engineers to conduct experiments as compared to using students (Höfer & Tichy, 2007). It is therefore the students that assume the role of the software professionals, to enable scholars to conduct research. The use of students as subjects by researchers is widely accepted, but debated based on its validity and lapses.

However, students will continue to be used as subjects in various empirical studies in the SE. Höst, Regnell and Wohlin (2000) encourage the question about the use of students as test subjects. Researchers pilot their experiments with students before they are implemented in the industries. The student group that partakes in experiments gets benefits as well as the costs. Since the motivations of the students are usually different from that of the software professionals, this could cause some variations that may serve as a benefit or cost for them. (Carver et al., 2003; Staron, 2007a; Runeson, 2003.)

1.3 Research question

There are a couple ethical issues in the using of students in SE. However in an institution or set up software laboratory, where students are used as subjects for software engineering research purposes, it leaves one to question if there are ethically related issues. Students have effects and impact in SE. Experiments have effects on students as students input also impacts on the experiment. The impact experimenting on students have on them and vice versa raise ethical concerns of experimenting on students. (Carver et al., 2003.)

Abrahamsson, Kettunen and Fagerholm (2010 June), propose that research must be built on ethical rules and researchers must endeavor to abide by the research protocol and design. Staron (2007b), reports that the code of ethics provided to be followed when using human as subjects were followed during the experiment.

The ethical concerns in using students as subjects in SE experiment must be taken into account both in conducting and reporting of the experiment. It leads to asking questions if researchers report ethical issues they face in their experiments with students. (Singer &Vinson, 2000.) If researchers do follow the guidelines of using students as subjects then it leads us to the research question of**;** *what are the ethical issues reported in using students as subjects in software engineering experiments?*

1.4 Research method

The research method employed for this study is systematic mapping review. This method is termed as mapping study (Petersen, Feldt, Mujtaba & Mattsson, 2008) and also scoping study (Kitchenham & Charters, 2007). It provides a structure of the type of reports and results that are published by categorization. The scoping study is considered an evidence-based SE study that shares the same procedure as a systematic literature review except for criteria of data extraction. A mapping study is used to provide an overview of a research area by assessing the quantity of evidence that exists on a particular topic. (Kitchenham, Budgen & Brereton, 2011; Petersen et al., 2008.)

1.5 Structure

The remainder of the paper is structured as follows; Chapters two and three elaborate on the background of the study. An insight into ethically philosophy and ethics as a general subject is introduced in Chapter two and streamlined to the kind of ethics that applies to the SE experimentation. In Chapter three, ethics or morals in the use of students for empirical studies and various experiments are discussed. Chapter four discusses SE experiments and its related ethical issues.

Chapter five gives a breakdown of the systematic mapping review and why the chosen method and step by step implementation of it. Chapter six takes on the results and analysis. This section of the paper will also shed more light on research results and contributions, limitations of the study. Chapter seven discusses the findings or results and its implications. Chapter eight expresses the observations and conclusions made as a result of the finding of the study.

2. Background

The background of the study is found rooted in two perspectives; ethical philosophy and use of humans (students) as subjects in experiment. This chapter takes a look into ethics, the general definition and the identified types of it, ethical history and theories, the declarations of philosophers over the years. It further explains ethics in the concept of humanity and the moral upbringing or development of a person. Ethical concerns and the idea of ethics as related to experiment are considered in this chapter.

2.1 Ethical history and theories

Ethics as defined earlier, are principles of right and wrong (Laudon & Laudon, 2010, pp. 151). Ethics are explained as a set of codes or rules of conduct (Kuhar, 2006) or external standards according to Gotterbarn (2001), provided by a group or institution or society to which an individual may belong. It is also philosophically described as the manner or way someone may respond to another or anything (Crowel, 2012). In the field of ethics and moral philosophy the idea of right or wrong behaviour is systematized, defended and recommended. Ethical theories are classified into three types namely metaethics, normative ethics, and applied ethics. (Fieser, n.d.; PB, 2008.)

2.1.1 Metaethics

Metaethics is the study of the foundation and explanation of the perceptions on ethics. It covers the scope from moral semantics through to moral epistemology. It consists of issues such as metaphysical which concerns morality in independent humans and intellectually which is the mental support of human moral actions and judgement. (Fieser, n.d.; PB, 2008.)

Metaphysically, metaethics consists of the discovery of either moral appraisals are eternal truths in the spirit-like sphere or man's ideology. Plato classified moral values as spiritual objects (MacIntyre, 2003 pp. 18-36) and the medieval philosopher also see it as spirit-like objects therefore terming it as eternal law (Crisp, 2013 pp. 818-841). Samuel Clarke refers to it as spirit-like relationships according to PB, (2008). Moral relativism has two forms, individual relativism where it describes that people create personal moral standards and cultural relativism, which depict that morality is dependent on societal approval and not individual (Crisp, 2013 pp. 753-816). (Fieser, n.d.; PB, 2008.)

Thomas Hobbes had the view that all human actions are motivated by selfish desires even the act of charity is empowering and it is termed as psychological egoism (MacIntyre, 2003 pp.78-93). Even though instinctive selfishness may drive human actions, there is the psychological capacity to show benevolence. Some of the human actions are actually inspired by instinctive benevolences, which are termed psychological altruism. (Fieser, n.d.; PB, 2008.)

Moral judgment involves emotions and not reasoning. That all reasons will come to play but a distinctive emotional reaction is needed and moral assessment is not about factual descriptions. However, Kant argues that a pure moral action is inspired by

reasons and void of emotions and desires (MacIntyre, 2003 pp.22-126). Baier (1958, pp. 187-191) continues that all human moral actions are backed by reasons and some justification. Therefore a good moral decision is made by giving a proper reason to back a course of action. (Fieser, n.d.; PB, 2008; Baier, 1958, pp. 187-191.)

2.1.2 Normative ethics

Normative ethics include the derivation of moral standard that guide right and wrong activities in human. It depicts the search for an ideal measure for right or wrong and proper behaviour. For normative principles, the golden rule is considered as the basic and classic example. The golden rule establishes one principle to which all actions are judged. Good character attribute and foundational principles constitute the focus of other normative theories. Either single rule or a set of fundamental principles, normative ethics has one criterion for moral behaviour. Normative ethics consist of three (3) theories namely virtue, duty, and consequentialist theories. (Fieser, n.d.; PB, 2008.)

Morality, according to Fieser, (n.d.) is believed by many philosophers simply consist of following exactly defined rules of conduct. Virtue theories, however, puts a little attention on learning the rules but focuses on the importance of attaining a positive habit of character (Crisp, 2013 pp. 744-766). The four virtues propounded by Plato are called the cardinal virtues, consisting of wisdom, courage, temperance, and sincerity (Crisp, 2013 pp. 21-42). Virtue theory alongside advocating for good habits also teaches the avoidance of acquiring bad character. (Fieser, n.d.; PB, 2008.)

Virtues are good habits acquired to regulate human emotions as argued by Aristotle as stated in Fieser (n.d). After Aristotle, medieval theologians added to the list of virtues, three Christian virtues: faith, hope, and charity (MacIntyre, 2003 pp.38-53). MacIntyre (2007, pp. 181-203) argued that virtues are something grounded in social traditions. (Fieser, n.d.; PB, 2008.)

Under normative ethics, there are four duty theories. These theories are base human morality on the foundational principle of obligation. The first duty-based theory is classified into three folds: duties to God, duties to oneself, and duties to others. Duties towards God are both theoretical and practical. Duties towards oneself concern one's duties to the soul thus acquiring skill and duties to the body by not doing harm to the body. Duties to others are classified into two t absolute and conditional duties to other people. The absolute duties include avoiding to wrong other, treat people as equals and with respect, and advocate the good of others. Conditional duties refer to the principle of keeping promises as contractually agreed between people. (Fieser, n.d.; PB, 2008.)

Second duty-based access to ethics is rights theory. Rights are a justified petition that is against another person's behavior. Rights and duties are relative as the rights of one person are the duties of another. Moral rights are said to have four features associated with it. Rights must be natural, thus not invented or created by governments. It must be universal, do not change from country to country. Rights must be equal to all people, irrespective of gender, race, or handicap. It must be inalienable, thus one cannot hand over his or her rights to another person. (Fieser, n.d.; PB, 2008.)

The third duty-based theory stresses on a single principle of duty, that humans have moral duties to themselves and others. Kant believes that the morality of all human

actions is determined by appealing to the single fundamental principle of duty (MacIntyre, 2003 pp.122-126). The fourth duty-based theory emphasizes prima facie duty where human duties are part of fundamental nature of the universe. These moral duties include fidelity, reparation, gratitude, Justice, beneficence, elf-improvement, and nonmaleficence. (Fieser, n.d.; PB, 2008.)

Moral responsibility is measured by the consequence of human actions. Cost-benefit analysis of an action is solely used to determine correct moral conduct. In consequentialism, an act is morally decent if its consequences are more favorable than it unfavorable. Consequentialist normative principles require humans must tally both good and bad outcomes or consequence of an action and determine whether the overall good consequences outweigh the bad consequences. The action is morally decent if the good consequences are greater than the bad. Consequentialist theories are also termed teleological theories. (Crisp, 2013 pp. 443-481; Fieser, n.d.; PB, 2008.)

The most appealing aspect of consequentialism is that it pleas to publicly observable consequences of actions. Three subdivisions of consequentialism emerged as ethical Egoism, which implies that an action is morally right if the consequences of that action are more favourable than unfavourable only to the agent doing the deed. Ethical altruism is when an action is morally right if the results of that deed are more favourable than unfavourable to everyone except the person or actor. Utilitarianism refers to when an action is morally right if the consequences of that action are more favourable than unfavourable to everyone. (MacIntyre, 2003 pp 145-157; Fieser, n.d.; PB, 2008.)

According to Fieser (n.d.), Jeremy Bentham proposes two theories, act-utilitarianism and hedonistic utilitarianism. Act-utilitarianism depicts that human tally the consequences of every action to be performed and determine on a case-by-case basis whether an action is morally right or wrong. Hedonistic utilitarianism proposes that human should tally the pleasure and pain which results from actions to determine is conducts are moral. This theories raises come concerns, which are addressed by rule-utilitarianism and ideal utilitarianism respectively. (MacIntyre, 2003 pp.145-157; Crisp, 2013 pp. 292-309; Fieser, n.d.; PB, 2008.)

According to act-utilitarianism, an act could be morally permissible once the social benefit outweighs the misfortunes. The rule-utilitarianism however, depicts that behavioural code or rule is morally right if the consequences of adopting that rule are more favourable than unfavourable to everyone. The adopted rule is morally binding because it produces a favourable consequence for everyone. Hedonistic utilitarianism considers pleasurable consequences as the only factor ignoring other significant factors that are not necessarily pleasing or painful. Ideal utilitarianism however, proposes that human tally any consequence that is intuitively recognised as good or bad and not just pleasurable or painful. Preference utilitarianism considers tallying any consequence that fulfils our preferences. (MacIntyre, 2003 pp.145-157; Fieser, n.d.; PB, 2008.)

2.1.3 Applied ethics

Applied ethics consists of the analysis of specific, controversial moral issues. It has been divided into acceptable groups such as medical ethics, business ethics, environmental ethics, and sexual ethics. Two features must necessarily be present for an issue to be considered as applied ethical issue. Firstly, the issue must be controversial in

the sense that there are significant groups of people both for and against on the issue at hand. Secondly, an issue must be a distinctly moral issue. (Fieser, n.d.; PB, 2008.)

In applied ethics, principles are set to guide and navigate through controversial issues. The principles must not be narrowly channelled, such as a version of act-egoism that may focus on action's short-term benefit. These Principles must also be seen as having merit by people on both sides of the controversial ethical issue. Principles that plea to duty to God are not usually cited for the reason that it might not impact on a non-believer that is engaged in the issue. The following principles according to Fieser (n.d.) are the ones most commonly appealed to in applied ethical discussions:

- Personal benefit
- Social benefit
- Principle of benevolence
- Principle of paternalism
- Principle of harm
- Principle of honesty
- Principle of lawfulness
- Principle of autonomy
- Principle of justice
- Rights

The personal benefit acknowledges the degree to where an action results in beneficial consequences to the individual, while social benefit acknowledges the measure to where an action gives beneficial consequences to society. Principle of benevolence consists of helping those in need. The principle of paternalism refers to the act of assisting other people to purse their interests when they are not privileged to do it themselves. Principle of harm simply means do no harm to others. However, the principle of honesty implies that others are not to be deceived. Principle of lawfulness means that the law must not be violated. The Principle of autonomy recognises the freedom of a person over their actions or physical body. Principle of justice notices a person's right to due process, fair rewards for harm suffered or inflicted, and fair share of asset. Rights acknowledge a person's rights to life, safety, privacy, information, and expression. (Fieser, n.d.)

The principles mentioned above stands for an area of normative principles that are traditional and are deduced from consequentialist and duty-based approach. Personal benefit and social benefit are consequentialist as they plea to the consequence of an action as it affects the individual or society. The rest of the principles are duty-based, which are established on duties towards others. The principle of autonomy, justice, and the various rights are established on moral rights. (Fieser, n.d.; PB, 2008.)

2.2 Ethics on human concept of life

Ethics is divided into three (3) categories namely basic, specifics, and technicalities. The basic consist of moral standard, morality as a guide, means for survival, values, virtue, self-interest and harmony of interest. While the specifics are based on self-reliance, productiveness, integrity, honesty, pride, justice, benevolence, and rationality. The technicality aspect consists of Metaphysical Justice, Free Will, Courage and Trader Principle. Ethics is an agent of moral development and academic discipline. (Landauer & Rowlands, 2001; Resnik, 2015.)

Ethical concepts are only useful when applied to real-life problems with human experiences. It applies to a more typical aspect of life such as family, profession, and politics. Ethical principles are used with accuracy or rigor in life as it cannot take into consideration a specific condition, as a result of that; there is often uncertainty about the answers to moral problems until it comes to specific cases. Becoming socially good is mostly about how an individual's special interests are regulated. The individual must however use personal discretion or judgment as she or he forms a system of applied ethics of his or her own. (Fieser, n.d.; PB, 2008.)

For typical selfish reasons, the acting agent human being is better off living in a society with moral rule than one with such rules. Without these moral rules, people would be subjects to the vagary of other people's selfish interest. Selfishness alone will encourage someone to adopt a basic set of rules, which will give way to an enlightened community. These regulations will, however, establish the safety only if the rules are sanctioned. Therefore, for selfish reasons alone, means are devised to enforce these rules. (Crisp, 2013 pp. 73-91; Fieser, n.d.; PB, 2008.)

There are many issues discussed that are surrounded by controversies in basic human living such as business ethics, biomedical ethics, environmental ethics, sexual morality and social morality. Biomedical ethics focus on a range of issues that arises in clinical settings. Health care workers continually get into unusual position dealing with life and death situations. Medical ethics issues are extreme and diverse than other areas of applied ethics. There are other controversial issues such as prenatal issues, genetic manipulation of foetuses, abortion, patient's rights and physician's responsibilities, medical experimentation on humans, physician-assisted suicide and euthanasia. (Fieser, n.d.; PB, 2008.)

In the field of business ethics, moral controversies are examined with relation to the social responsibilities of capitalist business practices, the moral status of corporate entities, affirmative action, basic employee rights, drug testing, whistle blowing, insider trading, and deceptive advertising. Environmental ethical issues usually overlap with business and medical ethics. These issues may include pollution control, obligation to future generations, whether eco-systems are entitled to direct moral consideration, rights of animals, the morality of animal experimentation, management of environmental resources, and preserving endangered species. (Fieser, n.d.)

Sexual morality entails a lot of controversial issues such as sexual relations without love, polygamy versus monogamy, homosexuality, and extramarital affairs. Other issues are of social morality such as capital punishment, welfare rights, racism, recreational use of drugs, and nuclear war. (Fieser, n.d.; PB, 2008.)

2.3 Moral development

Resnik (2015), states that ethical norms are acquired from social settings and evolves as people grow. However, Rest (1984) deduces that, for a person to be morally matured, there are four factors or things that needed to be evident in that person. These four facts are the important psychological components that need to be developed in order for someone to be regarded as morally correct. These factors are moral sensitivity, moral judgement, moral motivation and moral character. (Rest, 1984; Vartiainen, 2006.)

Moreover, the above-mentioned factors are facts that give a person the recognition of being morally correct. One must have the ability to note and recognize the ethical issues as they occur. The individual must be able to make intelligent decisions and have better judgement on the ethical issue as it arises. Getting the zeal to make a move or the courage to take action is a good quality or attribute that an individual must possess. Also, the personality or character to uphold and sustain good morality perceived. The psychological strength the person gets to take action. (Rest, 1984; Narvaez & Rest, 1995; Vartiainen, 2006.)

Morals are influenced by society and culture, however; they turn to be personally conceived principles that are esteemed and upheld by the individuals themselves. It is also defined as the personal or individual character. Ethics, on the other hand, are status, code, and standard for acceptable behaviour or as a social system. Though morals and ethics seem defined differently, they are mostly used synonymously; some researchers have used them interchangeably. Therefore, it makes no difference if the contexts in which these terms are being used do not contradict. (Brian, 2005; Vartiainen, 2009.)

2.4 Experimental ethics

Integrity in research is very important. Researchers from across the world expect that results reported from experiments are trustworthy. Researchers without being bias are expected to present information and findings to the world accurately. The trust in academic research will endure when the scientific environment devotes attention to illustration and transmission of values that are tied with ethical research activities. (NAP, 2009.)

NAP (2009), emphasize the obvious obligations of researchers. It highlights that researchers have the burden to honour the trust of their colleagues. Colleagues of researchers across the globe place their trust in them. It is expected of them to have followed the rules and have the right thing done. A researcher has got an obligation to himself or herself. Researchers owe to themselves to be truthful in other to preserve their reputation and pedigree. Experimenters are obligated to act in ways that will serve the general interest of the public. Scientific experiments greatly influence the society with its outcomes. McLeod (2015) addressed a number of principles that researchers and experimenters must adhere to. These principles include beneficence and nonmaleficence, fidelity and responsibility, integrity and justice. These compliment the obligations discussed by NAP (2009). (NAP, 2009; McLeod, 2015.)

Research ethics may vary with places or locations as may be influenced globally and locally. However, for a common ground of understanding on what research ethics should be, it is therefore required of the researchers to have a dialogue with participants. Each party negotiates to bridge the gap of understanding of ethics in research (Skovdal & Abebe, 2012.) Elms (1998, pp. 146-160) highlights the issue of entrapment, psychological injury, influence and ultimate value in researches that involves humans as subjects.

Moral or ethical cases hardly result in a plain answer of right or wrong. Since no particular answer seems to be good or bad, it is, however, a matter of judgement. However, judgement in this sense means determining whether particular research is excused or not (McLeod, 2015). Ethics deals with a number of factors. Even though moral judgement had been a pivotal part, moral views, behaviour, emotions,

deliberation, reasoning are contributing factors that influence moral issues (Alfano & Loeb, 2014).

Smith (2003) propounds some vital points that should be covered by an experimentalist. Presenting and clearing any issue of intellectual property. Considering and being watchful of multiple roles of participants. Observing and adhering to informed permission rules. Respecting privacy and confidentiality of individuals involved. Getting access to more ethical resources and make use of them. (Smith, 2003.)

Factors, methods, and principals are essential and impactful as any other set of codes or rules of conducts in research. However, they do not cover every ethical issue. They are most of the time in argument especially based on perceptions. This leaves it to require a considerable explanation and deliberation. Therefore, the interpretation, evaluation, and application of research rules is a must learn for all experimentalist. It is important that researchers learn how to apply different and conflicting research rules. Decision making in various ethical issues depends on how well a researcher can interpret and apply the said rules. A lot of decisions in ethical cases consist of a forward implementation of the ethical rules. (Resnik, 2015.)

3. Software engineering ethics

This chapter explains software engineering ethics as in the three concept introduced by Gotterbarn (2001) plus ethics considered as a professional lifestyle into three categories: Firstly, ethics in engineering as the activities of software engineers. Secondly, as a set of guidelines and principles lay-down by an association, society, or organization. Thirdly, ethics in engineering described as the study of the relationship that exists between the activities and the principles provided. Lastly, ethics as professionalism.

3.1 Ethics as an activity

The daily lives and activities of the world's population mostly depend on information systems (Reed, 2000). The engineer behind the making of these systems, however, needs to adhere to the value of human beings (Boehm, 1981). The decisions that are made during construction, development, maintenance and design of computing artefact constitute the activities of software engineers (Gotterbarn, 2001). Developmental decisions made either by teams, management, individuals, or the professional society affect the lives of others and shapes people's societal values (Narayanan, 2013). Therefore, when engineers are constructing an artefact, the professional SE ethics activities are integral to a technical decision made during development (Gotterbarn, 2001).

However, these technical decisions must be outlined, instrumented, or guided by ethical decisions on human values (Boehm, 1981). This creates the urgency for the governing bodies and associations to ensure that their members follow laid-down guidelines. In other words, the organizations are to see to it that no harm is done to the public by any product or artefact designed by its members intentionally or unintentionally. (Jonsen, 1978.)

3.2 Ethics as principles

Secondly, SE ethics are principles or codes of conducts and guidelines for operating. Software engineers' decisions have an impact on people's lives. However, the ethical activities that play important roles in the technical decisions must be understood based on the kind of impact the decisions may have. However, decisions that are based only on technical issues may have either good or bad impact on the users' lives. Hence good decision made with vivid and careful consideration, is not only in technical aspect but ethical aspect as well. It is advised that technical decision is guided by the values. Technology and values work together, failure to recognize their relationship and incorporate them, makes a technology be regarded as lacking the principle of organization. (Reed, 2000; Gotterbarn, 2001.)

There are codes of conduct or ethics developed by professions and associations in other to explain their values. These codes as part of its numerous functions give directions or guidance to practitioners concerning ethical choices. Through the code of conduct, both the public and the practitioners are educated on the profession and its ethical values. (Davis, 1998; Herkert, 2000.) In order not to do any harm or damage (Jonsen, 1978), but

otherwise protect human values (Boehm, 1981), these codes are created. Therefore, the activities of practising professionals as a team, management, individual or organization must uphold the aims of the code, which is to preserve human values (Gotterban, 2001).

The IEEE - CS in collaboration with ACM came up with ethics to help maintain the professionalism of software engineers (ACM/IEEE-CS, 2000). The Ethical codes are classified into eight main principles listed as;

- Public
- Client and employers
- Products
- Judgement
- Management
- Profession
- Colleagues
- Self

The interest of the public is highly prioritized. Clients and employers interest must be respected by the actions taken. SE products should endeavor to meet higher standards or benchmark. Maintaining integrity and independence in professional judgment is required of software engineer. In software development and maintenance, management must advocate for ethical ways of doing things. SE experts must hold in high esteem the virtue and reputation of the profession. Software engineers must be truthful and helpful towards each other. SE profession requires the recognition of ethical practices. (ACM/ IEEE-CS, 2000.)

It is realized that the above-listed principles prioritize the public being the users of the artefact designed by software engineers. These principles are arranged in order of how an engineer's service is expected to be, giving more respect to the public that uses the designed artefact and value them. It indicates how the profession holds product users in high esteem. (Gotterbarn, Miller & Rogerson, 1999.)

3.3 Ethics as an academic discipline

Thirdly, SE ethics as a study is the interaction between the two categories namely ethical principles and technology. The study of ethics as a discipline examines the effect of interaction users of the products or citizen of an environment and the artefact itself. However, this discipline is said to illustrate the bond found between the principles and technical practice. SE ethics as a discipline refines the relationship identified between the principles and technology. It is also said to further explore the relationship between the given guidelines and the technical practices. (Gotternbarn, 2001; Bird & Sieber 2005.)

Furthermore, ethics is not only considered as a professional practice issue but also academia (Billington, 2006; Fleischmann, 2006). To recognize and address ethical issues is a primary skill that is required of every engineering professional. It is however of essence and relevance that science and engineering professionals have this competence to ensure a better working environment and service to the public. (Bird, 2003; Fleischmann, 2006.)

However, some institutions have SE ethics as a course of study for engineering students (Billington, 2006; Fleischmann, 2004; Lau, 2004.) Engineering ethics must be part of the academic discipline and as a matter of fact, it can be taken as a standalone course or as part of another course. Training engineers with ethics as part of their study may not be enough due to some external factors at the workplace. Organizations and firms that the trained engineers work for must organize workshops that will, educate and refresh them on ethics. Since ethics is not only a professional practice issue but also as academia. (Loui, 1997; Bird, 2003; Bouville, 2008.)

3.4 Ethics as professionalism

Professionalism in this context focuses on the integrity of the profession. Profound professionalism is therefore required of the software engineers, to avoid any unethical situation. It is also expected of them to uphold the profession in high esteem, by maintaining its integrity. In spite of all these ethics, if software engineers do not get to understand the values and standards of the profession, there is the possibility of letting things go haphazardly. The codes of ethics provided by every profession should be understood by the professional as part of his professional work (Gotterbarn, 2001). Bird, (2003) stated that an engineer must have the competence to recognize and address ethical concerns. This competence must be a primary skill for an engineering and science professional. (Gotterbarn, 2001; Bird, 2003.)

Spinello (1997, pp. 35-50) implied that there are connections that exist between ethics and professionalism. Mastering and getting acquainted to a complex knowledge, an effective and detailed intellectual training is required. The kind of services provided must have a significant bearing or contribution to the society or environment. There is the prediction of independent judgement in a task done based on expertise. The codes of ethical conducts are set to regulate the behavioural standard of the software engineer. Basically by these codes, one should act professionally and abide by the responsibilities adorn on them as professional ethics. (Burmeister & Weckert, 2003; Spinello, 1997, pp. 35-50.)

The professional ethics of software engineering uniquely defers from other disciplines of engineering. In addition to the professional ethics responsibilities, a software engineer has the technical responsibility as well, which is the procedures followed when solving problems, not only just problem but technical problems. It is specially tagged as quality or professionalism, which includes the idea of knowing and abiding by good professional standards. SE ethics are good as they shape the engineer in all professional activity. However, there are other ethics that apply to the section of experimentation in SE. (Gotterbarn, 2001.)

4. Software engineering experiment and ethical issues

Conducting an experiment entails a lot. One of the main challenges is recruiting subjects for participation. This chapter discusses software engineering experiment, using students as subjects and ethically related issues encountered in experiments. It further discusses issues of confidentiality, informed consent, beneficence, and scientific value.

4.1 Software engineering experiment

SE, according to the IEEE means applying the principles of engineering to the software development field (SEI, 2016.) It is also described as a well-articulated life cycle by SEI, (2014). SE is described as technical discipline that develops software by Gotterbarn (2001). Experiments involving software systems, products, processes, and resources are purposed to validate a theory or hypothesis by collecting any available data (TFD, 2015). SE does not consist of the technical view of developing software systems only, but managerial issues including directing programming teams, scheduling, and budgeting (Webopedia, 2016).

However, the growth in the SE field largely depends on how fast the researchers work to find solutions to existing and future problems. Also, how further into the future these researchers are able to project or predict. In the wake of solving problems and predicting the future, experiments are conducted. Researchers in SE do investigate or experiment to further it in any aspect of concern and relevance. (Reed, 2000.) Experimentation has been the act of verifying, disputing or establishing the authenticity of a hypothesis (TFD, 2015).

SE needs a cycle of model developing, learning and experimentation as any other engineering study. In valuing and making choices between tools, techniques, methods and languages as a software engineer, experimentation is the best way. Experimentation is purposed to introduce practitioners, researchers, students, and teachers to empirical studies. In SE, this introduction is done by means of controlled experiments. Empirical studies in SE field take many forms and shapes with its own different way of presenting the result. (Basili, Selby & Hutchens, 1986; Kuzniarz, Staron & Wohlin, 2003.)

Basili et al. (1986) developed a model for experimentation processes; definition, planning, operation, and interpretation. Wohlin et al. (2012 pp. 85-157), propose the experiment processes as scope, planning, execution, analysis, and interpretation. Experimentation in the SE domain is performed on professionals and students as subjects. For the purpose of this study, it's only the use of students in SE experimentation that will be discussed. (Basili et al. 1986; Wohlin et al. 2012 pp. 85-157.)

4.2 Using students as subjects

In SE literature, quite a reasonable amount of reported empirical studies have imployed students as subjects. Using students for experiments has primarily been the basis of

validating research hypothesis. It also serves the purpose of piloting; theories are tested on students before implemented in the industries. There are other reported experiments that are run both industrially and with students. Students on all accounts are termed as inexperienced compared to the working professionals on the field. (Carver et al., 2003; Sjøberg, 2005.)

Using students for experiments has primarily been the basis of validating research hypothesis. This is mainly so because, the researchers' have these subjects of theirs research as their students. Researchers indulge the services of their students, making it an easy way of getting subjects in SE experimentation. Recruiting subjects for an experiment is not an easy task. However, with researchers own students the burden is minimized. It is preferred that students are used in certain conditions instead of professionals. Junior staff or less experienced staff in the industries is represented by students, and purposefully for educational reasons. (Sjøberg et al. 2002; Sjøberg et al. 2005.)

Höfer and Tichy (2007) implied that the frequent use of students in experiments points out the problem in recruiting professionals to conduct experiments on. Sjøberg et al (2005) reported that 87% of the investigated articles cited between 1993 and 2002 used students as subjects. This student body includes graduates and undergraduates. Höfer and Tichy (2007) further confirmed statistically that the percentage of using students as subjects in experimental studies conducted in SE outweighs the use of professionals.

Few reasons are given, for the use of students partake in experimentation as subjects while reviewing empirical endeavours in SE. Researchers need ample information to convince professionals to adopt new ways of doing things, and this information is obtained from experimenting and observing students. (Tichy, 2000.) Several studies have reported using students as subjects in their experiment (Kyritsis, Gulliver & Feredoes, 2006; Purchase, Welland, McGill & Colpoys, 2004; Land, Tan & Bin, 2005November).

Students in SE experimentation act as a stepping stone for the actual experiments in the industry with the professionals. This paves a way for a better organized and improved way of conducting the experiment in the industry with the professionals. Involving the student is to get a first-hand experience and result by piloting the research in an academic environment ahead conducting the said procedure in the firms with professionals. This gives a basis and evidence that help validate the hypothesis. Conducting experiments with students is less expensive and less risky as compared to the professionals in SE. (Sjøberg et al., 2005; 2002; Tichy, 2000; Carver et al. 2003.)

Students are used as subjects for experimentation in every field. Students used as test subjects in the SE experimentation and they are considered the best surrogates for professionals (Staron, 2007b). The impacts students have in software engineering field are undisputed. Experiments have effects on students as well as raises ethical concerns (Carver et al., 2003). Comparative study on students and professionals as subject concluded by suggesting and confirming the relevance of using students in empirical SE studies and that there are no compelling differences in terms of the correctness between students and professional. Results however from students are generalized and acceptable. (Höst et al 2000; Runeson, 2003.)

However, the learning experiences of the students used as subjects are affected due to the fact that they are subjected to different treatment (Staron, 2007b). In other words, the learning of these students is altered positively or negatively. Students are used as subjects for experimentation in every field. Students are used as test subjects in the SE experimentation. They are considered the best surrogates for professionals (Staron, 2007b). Carver et al. (2003) identified issues in using students in the category of information sharing or hiding, use of data collected, students grading based on their participation and how ethical it is to base student evaluation on how they performed in some empirical studies. (Staron, 2007b; Carver et al. 2003.)

4.3 Issues in using students

Using students certainly presents or raises a few ethical issues. Researchers deal with ethically related issues in the industrial run experiments, the piloting experiments conducted with students also are met with issues (Singer &Vinson, 2002). These issues need to be considered when conducting experiments with students. It is seen as a matter of importance in the experiment. Carver et al (2003) identified ethical issues that might come up when using students in empirical studies but solely in SE education. The use of students as subjects in experiments leaves some ethical questions to be discussed. (Carver et al 2003; Singer &Vinson, 2002.)

In using students, there are many ethical issues to be considered. Mainly issues that occur when using students as subjects are the different exposure given to them. These exposure students used as subjects get alter their learning experiences, positively or negatively. Students that partake in these experiments gain some insight and experience that add to their knowledge and skill. Treating students differently due to experiments poses some threats. In other for these threats to be eliminated or minimized, subjects must see experiments as a helping mechanism that stimulates their learning process. (Staron, 2007a; 2007b.)

Experimenters and researchers have a code of conduct that must be followed. Every experiment depending on the field of study has a different set of ethical principles to be followed. Researchers adopt codes of ethics from other sectors or organizations to complement the activities. (Staron, 2007b.) Carver et al (2003) point out the merits and demerits students get in being subjects in experiments. Staron (2007b), continued by discussing the threats posed during experiments while using students.

Previous studies in this area have simply been to analyse and adopt from the psychology, medicine, and law to be used in all discipline of computing. However, the attention has been on participants and how they must be treated, being students or professionals. Singer and Vinson (2002) pointed out that research ethics are inevitable but its application in software engineering might not be direct. Davison et al. (2001) express that these guidelines can impose on researchers administrative burden. (Davison. et al. 2001; Singer & Vinson, 2001; 2002; Davis, 2001; Storey, Phillips, & Maczewski, 2001.)

Vinson and Singer (2008) propose four principles: informed consent, scientific value, beneficence and confidentiality which are the major pointers to ethical issues in SE experiment. Informed consent, beneficence, and confidentiality will be the basis of this study, which is further explained in the next sub chapter. Informed consent has been the primary ethical issue in software engineering studies. Vinson and Singer (2008)

deduced four common principles that are relevant for empirical software engineering research practices by reviewing codes of ethics from social science, biomedical, professional organizations, and government funding bodies. (Singer & Vinson, 2002; Vinson & Singer, 2008.)

4.4 Principles and ethical issues in experiment

The common principles deduced are informed consent, scientific value, confidentiality, and beneficence. The principles exploit either experimenters are monitoring the subjects' behaviour directly, or indirectly, or whether the subjects are just giving the code for examination. It is also applicable to any subject, whether students, volunteers, employees or organizations. The ethical principles of informed consent, scientific value, beneficence and confidentiality in experiment are discussed in details. (Vinson & Singer, 2008.)

4.4.1 Informed consent issues

The principle of informed consent on the side of the research subjects in terms of participating in a study must follow the principle of respecting individual's sovereignty. In actual sense, possible subjects have got the right to choose if they will partake in the projects before it begins. The full components of informed consent are debatable, however, it is very clear that it must at least consist of the following; comprehension, disclosure, voluntariness, actual consent, and competence. (Fleuhr-Lobban, 1994; Faden & Beauchamp, 1986, pp. 4 - 33; Singer & Vinson, 2008.)

Disclosure points out the information that researcher ought to make available to the subjects to enable them to consider and make knowledgeable decision by accepting to partake in the study or not. The information to be provided must include the research procedure, the expected gains for the subjects and the world, options to partake in the case of students as subjects, the voluntary type of participation, the risk to the subjects, the research purpose, the handling of confidential information, a statement of assurance to answer the subjects' question, and the type of data that will be gathered and its use. The whole idea is to get the potential subject well informed. Get the necessary information about the study and understand its effect on them. (Faden & Beauchamp, 1986, pp. 4 – 33; Sieber, 2001; Patrick, 2006; Singer & Vinson, 2008.)

Comprehension mandates the experimenter to make available the necessary information in the simplest way for the understanding of the subject. Therefore, the information must be void of technical vocabulary and terms especially outside the subject's scope. Competence concerns the subjects' capacity to make the reasonably informed decision to get involved in the study. Voluntariness stipulates that informed consent ought to be acquired under circumstances void of intimidation and improper influence. The consent of the potential subject must be voluntary and wilful. The subjects as part of voluntariness must have the right and choice to opt-out or terminate participation at any given as they so wish. However, the subjects' decision to partake must express an active authorization, but not an implied acceptance or a sheer formality. (Faden & Beauchamp, 1986, pp. 4 - 33; Davis, 2001; Vinson & Singer, 2008.)

4.4.2 Issues of beneficence

The level of beneficence issues begin from the measured combination of harms, perils, and gain to the potential subjects and the society partaking in the study. Researchers are entreated to increase the benefits for study subjects. Benefits come to light as early as from the research topic and the possible harm from the research methods. The moment the research question is selected or formed, researchers can select methods that will increase beneficence but will reduce the peril of damage to the subjects. The said principle applies to a group of subjects such as an ethnic group, organization or companies, and or socio-economic groups not only to an individual subject. Harm in this context does not only refer to physical harm but includes entities such as fatigue, loss of decency, self-esteem, personal freedom, interrupting of daily activities, boredom, and financial harm. The maximum risk of harm in empirical software engineering arises from the violation of confidentiality. (McNeill, 1993; Vinson & Singer, 2008.)

4.4.3 Issues of scientific value

The study must be done with a valid methodology for the results to faithfully reflect reality else it will yield no benefit. A study that is without benefit should not commence. Scientific value is classified into two categories: the efficacy and effectiveness of the study, and the relevance of the research topic. Understanding the standard research and statistical methodologies in this context refers to competence. Empirical software engineering is yet undergoing an immense deal of development in terms of methodologies. Methodological development presents complications for weighing scientific value. It is challenging to estimate the validity of a new methodology since it is recent. It is strenuous to rate the scientific value of the advancement of the said method. By replicating a well-established result, a new methodology can be validated. (Freedman, 1987; McNeill, 1993; Singer & Vinson, 2008.)

4.4.4 Issues of confidentiality

The right subject to anticipate that information given to the researcher will remain confidential is referred to as the principle of confidentiality. Subjects' identities are to be concealed and protected by researcher whether they are organizations or individuals. Information that does not have a direct correlation with the research must be considered private and kept confidentially. (Vinson & Singer, 2008.)

Data privacy, data anonymity, and anonymity of participants are the three components that form confidentiality. Limitations that are enforced on the access granted to the data gathered from respective subjects are termed as data privacy. This data must be securely stored, and password protected or key locked to keep data privacy. Getting access to the collected data must be limited to core people who are normally members of the project team. (Patrick, 2006; Vinson & Singer, 2008.)

If possible, personal or organizational facts that may be used to identify a subject must not be collected to safeguard the anonymity of the gathered data. To avoid a possible breach of confidentiality, personally identifiable information must not be collected as it may require informed consent. Apart from names, individual characteristics could serve as identifiers by experienced experts. Anonymised data could also be applied when reporting accumulated data like medians and standard deviations, which are difficult in

empirical software engineering studies as there are often a few subjects used. (Vinson & Singer, 2001; Vinson & Singer, 2008.)

The anonymity of participant is achieved by disguising the existence of the subjects. The subjects identity the must be hidden from other competitors, colleagues, clients, managers, professors, and the public. Identities of these subjects must be protected as it may impact on their career and studies. Researchers must be sensitive to conceal the characteristics that identify a participating company in the research as competitors and clients influence companies and organizations' business or decision making or strategies. The manager or the professor must not have the knowledge as to either an employee or student refused to partake, considering the professor or the manager is not the researcher. (Vinson & Singer, 2008.)

Data collection is best done in private, except observational studies, which would not be easily accomplished in open office cubicle setting. Researchers are advised to use misleading pseudonyms if an identifier is paramount for clarity. The respective names of subjects and organizations must not be written unless otherwise requested by the subject. Executives of companies sometimes request the identity published, in such case, experimenters must inform of the implications and potential risk. (Vinson & Singer, 2008.)

All parties that are associated to the study either researchers, subjects, managers, research assistants or professors, the subject of confidentiality must be prioritized and emphasized. Loss of trust between researchers and the subject populace would be as a result of breaches in confidentiality. This breakdown in trust may leave the research without access to the subject populace. It is of utmost importance to the researcher for the subjects' confidentiality to be protected and their data and also to inform them of limited confidentiality. (Vinson & Singer, 2008.)

5. Method

The method selected for this study was mapping review, as it suits the search to answer the research question. This chapter explains the research method used, also it give a step to step process on how the data was gathered and extracted. The procedure includes defining the scope of the study, conducting search for primary studies, screening for relevant papers, developing classification scheme, extracting data and mapping of the study.

5.1 Systematic mapping review

Systematic reviews are differentiated by scientific literature into two namely: conventional systematic review and mapping study. The conventional systematic review focuses on aggregate results about the effectiveness of a treatment, intervention, or technology, and mapping study however, seeks answers to causal or relational research questions. Mapping study aims to identify researches that are related to a particular topic and to answer much broader and exploratory questions that are connected to the trends in study. (de Magalhães, da Silva, Santos & Suassuna, 2015; Fernández-Sáez, Genero, & Chaudron, 2013.)

A scoping study is a method used in building classification schemes. It is used to build a structure of any field of interest, including any SE field. In this method, analyzing the result is mainly centered on the recurrences of publication. The scope of the search field is determined and control by the researcher. A systematic mapping research presents the summary of results visually. (Petersen et al, 2008.) This type of study gives a structure of the form of experiment reports and its published results by categorizing them. Researchers can be overwhelmed by the amount of data they have to work with during this process, and mapping can help to synthesize primary findings and exhibit the complexity of the study topic in a more manageable way. (Petersen et al, 2008; Budgen, Turner, Brereton & Kitchenham, 2008.)

A mapping study is used to present an outline of a research area by assessing the quantity of evidence that exists on a particular topic. It is described as the preliminary study that precedes Systematic Literature Review (SLR). SLR and mapping study are manually intensive. The process of sorting through existing literature is labor exhaustive. (Brereton, Kitchenham, Budgen, Turner & Khalil, 2007; Babar & Zhang, 2009; Riaz, Sulayman, Salleh & Mendes, 2010; Marshall & Brereton, 2013.)

By categorizing and identifying whether there were ethical issues reported, this study reported on SE experiments that had students as subjects. This current study employed SMR to establish the quantity (Petersen et al, 2008) of studies that have reported on ethics while using students as subjects. In identifying the evidence that is available on the reporting of ethics in the use of students as subjects, the systematic review mapping or scoping study was employed. Categorize the results into two those that reported on ethical issues and those that did not. Go further into the ones that reported on ethical issues to classify the kind of ethics. The following procedure as suggested by Petersen et al (2008) was adapted and applied.

Systematic mapping (SM) is purposed at providing an overview and identifying the quantity of evidence in a particular area of research. The results attainable and type of research in a particular research topic is a reason for conducting an SM. To gain success in SM, it must be planned properly (Petersen et al, 2008; Nakagawa, Feitosa & Felizardo, 2010.)

Search strings are used to identify primary studies from digital databases. The research question drives the structure of which the search should follow. In formulating the search strings, advice that was given by Kitchenham (2004b), and Kitchenham and Charters (2007) were taken into account and followed. Inclusion and exclusion criteria are used to eliminate irrelevant reports or papers. Papers that were relevant are the ones that seek to answer the research question (Petersen et al 2008). Keywording helps in developing the classification scheme making sure that relevant papers are not overlooked. It makes sure that the scheme considers or takes into account the existing studies. (Petersen et al 2008; Paternoster et al 2014.)

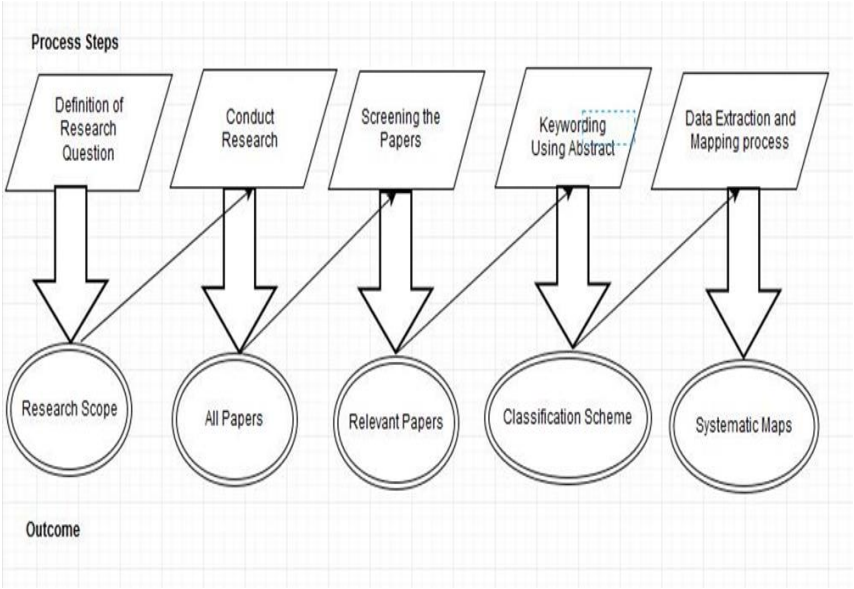

Figure1. Systematic Mapping Process as adapted from Petersen et al (2008)

The study followed the processes illustrated in Figure 1, as search was conducted to answer the research question.

5.2 Definition of research question and scope

The study addresses the research question; what ethically related issues are reported in using students as subjects in software engineering experiments? In answering the research question, one must find out the number and gather all reported experiments that were conducted with students. Then group them into two; those that reported ethical issue and those that did not. The group that reported on ethically related issue is sorted into the types of ethics. However, only experiments done on students and published from 2000 to May 2016 were covered.

5.3 Search for primary studies

The search was conducted from the set of reputable journals and conference papers. However an attempted search for experiment reports in the ACM Digital libraries alone had 454,593 hits making it difficult to proceed on that due to time constraint and only one researcher on it. The following journals are hand-picked by the researcher of which most are of the Science Direct database. To avoid dealing with duplicates of papers, major databases like ACM Digital libraries and IEEEXplore were not used directly but individual journal and conference proceedings. The following journals and conference proceeding were used in this study;

1. Journal of Systems and Software
2. Information and Software Technology
3. Information Sciences
4. International Journal of Human-Computer Studies
5. Information Systems
6. Computers in Human Behavior
7. IEEE Transactions on Software Engineering
8. International Conference on Information Technology and Computer Science (ITCS)
9. International Conference on Software Engineering (ICSE)
10. International Symposium on Empirical Software Engineering (ISESE)
11. International Symposium on Empirical Software Engineering and Measurement (ESEM)
12. MIS Quarterly

Google scholar was used to aid in citing or referencing article properly, snowballing through references of papers for more relevant and related ones. The search strings were formulated using Boolean operators “OR” and “AND” to conduct the search (Kitchenham, 2004 September). In other to get the targeted data, a combination of search words and keywords were developed. These strings for the search were developed from the research questions. Upon advisement, the search strings were kept as simple as possible to prevent false positives. It was directed to help better the search. Search strings developed for the study are presented in Table 1.

Table 1. Search strings for the study.

Search strings
(”students as subjects” OR “students” OR “students as participants”) AND (“experiment” OR “empirical study”) (“ethics” OR “Ethical issues” OR “legal issues” OR “experimental ethics” OR “moral issues” OR “ethical value”)

The search for papers was started with the above-mentioned search strings, but there as need to change the strategy due to a realization that, no reports of SE experiments were stating the specific terms “ethics, ethical issue or legal issues”. However, the original search strings were then modified. Firstly the search for “experiments” OR “empirical studies” from the listed journals and conference proceedings carried out. Within the reported experiment papers, the search for the ones that used students as subjects was done. It was noticed that some reports did not use students but terms like “undergraduates”, “graduates”, and “post-graduates”. Further search for students and

these terms within the results was done. The search strings was later modified, the final search strings are showed in Table 2.

Table 2. Final search string of the study

Final search strings
("experiment" OR "experiments" OR "empirical studies") AND ("students" OR "undergraduates" OR " graduates" OR "post-graduates")

The study continued by reading to ascertain student participation in the reported experiment. The search resulted in 2512 papers reporting on an experiment in the 12 journals and conference proceedings.

5.4 Screening for relevant papers

Papers were screened through in order to get relevant ones. The criteria were totally influenced by the research question. It was considered useful to include papers that made mention of the main focus of this study, experiment, students as subjects in its title or abstract. It was taken into account papers that mention ethics too. Some papers may have frequently used it in abstract but did not further address it. However this exercise brought the number of papers from 2512 to 562, reporting with students as participant or subjects. Table 3 illustrates the criteria that were used in screening for relevant papers for this study.

Table 3. Inclusion and Exclusion criteria for the study.

Inclusion
1. published in journals or conference proceedings 2. peer-reviewed 3. published during and after 2000 to 2016 as report of an experiment 4. software engineering or information system based
Exclusion
1. not peer-reviewed 2. not written in English 3. published before 2000 4. not related to computer science, software engineering, and information system experiment 5. related to industry-university collaboration projects 6. paper refers to students as subjects but didn't actually use students

The inclusive criteria included peer reviewed publications in journals or conference proceedings that are published during and after 2000 to May 2016, that report of an experiment conducted and that are Software engineering or Information system based. Rejected publications are ones that were not peer-reviewed and written in English and also were published earlier; before 2000. If it was not in the field of study related to computer science, software engineering, and information system experiment it was omitted. Rejected publications included industry-university collaboration projects works, as the main focus was on reported experiments. The term students as subjects are used in some publications, which are not reporting on experiments and were however rejected.

5.5 Keywording and classification scheme

This was very useful as it forms the core concept in the area of study. First of all, the abstracts of the papers found relevant (562) were read. However, reading the abstract and introduction was not enough. Quite a number of the reports were not very clear in the abstract about which kind of participants were employed.

Further reading into the experiment settings was however necessary. The reading went further to spot ethically related issues in the papers. This is where researchers devote a portion of their chapter to describe their subjects, recruitment process, honorarium, etc. Most ethical issues are discovered and reported in that aspect of the chapter. After going through, 70 papers actually made mention of any ethical related issue. Keywords and concepts that relate to the contributing paper are identified. They are confidentiality, which consists of anonymity, and confidentiality. Informed consent includes disclosure, termination, comprehension, competence, and volunteerism. Beneficence split into benefit and harm, scientific value deals with the importance of research and validity of results. These keywords gathered from papers are put together to create an understanding of the contribution of the study. Further read through the conclusion of some of the papers, if the abstract, introduction, and experiment settings do not give the whole picture or expected data. The finalized keywords were used to make the categories, which build the map

5.6 Data extraction and mapping of studies

At this stage data is extracted into the classification scheme. The schemes are explained in details in the next chapter. After the classification scheme was made, the relevant 70 papers were sorted. The following data was extracted from each of the papers deemed qualified.

1. First Author's surname
2. Year of publishing.
3. Text (Statement that merit Ethical recognition in the report)

A short text is extracted from the paper to rationalize why it belongs to that category. However, not all the documents have the third part extracted and presented here. Experiments conducted with small and large number of subjects or participants was considered, meaning no limit to experiment size.

6. Results and analysis

In this chapter, the result is presented and briefly analyzed. It will also briefly explain the classifications and pointers, which are detailed and explained in Chapter 4. This chapter presents the outcome of the search conducted, results categorized and mapped into schemes. The distributions of the number of papers per journal or conference proceedings are presented in Table 4.

Table 4. Number of papers found with respect to Journals and Conference proceeding.

Journal or conference	**All papers**	**potentially relevant studies**	**Relevant Papers**
Journal of Systems and Software	378	106	7
Information and Software Technology	217	88	16
Information Sciences	149	19	2
International Journal of Human-Computer Studies	382	78	25
Information Systems	86	32	1
Computers in Human Behavior	268	54	3
IEEE Transactions on Software Engineering	217	85	11
International Conference on Software Engineering (ICSE)	166	28	0
International Conference on Information Technology and Computer Science (ITCS)	88	3	0
International Symposium on Empirical Software Engineering (ISESE)	33	23	3
International Symposium on Empirical Software Engineering and Measurement (ESEM)	78	34	0
MIS Quarterly	450	12	2
Total	2512	562	70

Quite a small amount of the articles found, reported on possible ethically related issues. The percentage of potentially relevant papers per each journal and conference proceedings is illustrated in Table 5.

Table 5. Percentage of potentially relevant papers per Journal and Conference proceedings.

Journal or Conference	**All papers**	**Potentially relevant**	**%**
Journal of Systems and Software	378	106	28.04
Information and Software Technology	217	88	40.55
Information Sciences	149	19	12.75
International Journal of Human-Computer Studies	382	78	20.42
Information Systems	86	32	37.21
Computers in Human Behavior	268	54	20.15
IEEE Transactions on Software Engineering	217	85	39.17
International Conference on Software Engineering (ICSE)	166	28	16.87
International Conference on Information Technology and Computer Science (ITCS)	88	3	3.41
International Symposium on Empirical Software Engineering (ISESE)	33	23	69.70
International Symposium on Empirical Software Engineering and Measurement (ESEM)	78	34	43.59
MIS Quarterly	450	12	2.67
Total	2512	562	22.57

The initial search for the studies ended up with 2512 papers however, 1950 papers could not make the cut after they were screened through some of the inclusion and also exclusion criteria. In seeking to answer the research question, 70 papers were found. Tables 6 and 7 present the percentages of the primary studies that were found with respect to all papers found and the potentially relevant papers.

Table 6. Percentage of relevant papers with respect to all papers.

All papers	**Relevant Papers**	**%**
2512	70	2.79

Table 7. Percentage of relevant studies with respect to the potentially relevant paper.

Potentially relevant	**Relevant Papers**	%
562	70	12.46

Comparatively, Tables 6 and 7 indicate that a small percentage of the papers that were found (all papers) and potentially relevant papers, to have reported ethically related issues thus 2.79 and 12.46 respectively.

A brief reminder the main pointers discussed in chapter 4 informed consent, confidentiality and beneficence for sake of this study. Table 8 illustrates the breakdown and summary of Vinson and Singer (2008) idea of ethical principles.

Table 8. Ethical principles propounded by Singer and Vinson (2008).

Informed Consent	Beneficence	Scientific Value	Confidentiality
Disclosure Comprehension Competence Voluntariness Termination	Benefit Harm	Importance of Research Validity of Results	Anonymity Confidentiality

Informed consent: the participant consenting or agreeing to partake in the experimenter's idea. The informed consent as discussed in detail consists of the following;

- Disclosure: How subjects are educated on the aims, process, merits and demerits of the experiments.
- Comprehension: How researchers prepare the subjects before commencing with the experiment.
- Competence: How subjects can understand and do what is expected.
- Voluntariness: How subjects freely consent to partake in the experiment.
- Termination: how freely a subject can opt out of the experiment if he or she wishes to.

Beneficence consists of benefit and harm, which is basically how and what the subject gain or losses for participating

Confidentiality: this consists of Anonymity and confidentiality.

- Anonymity: that data collected cannot be used to identify a subject.
- Confidentiality: deals with limiting those who can get hold of the data and how they are presented.

Each relevant paper that met the screening criteria is assigned a code. See Appendix A for details of paper codes and it sources.

6.1 Informed consent

Out of the 70 papers, 51 papers reported of issues related to informed consent thus 72.86 percent of it, with 7 reporting on 2 categories. The distribution of it, with respect to the contents of Informed Consent is shown in Table 9. (PS31) and (PS32) reported that participants were made to read and sign consent forms before commencing. Being it an indication that subjects volunteered or mandatory in a course of study, are placed under Voluntariness, as it depicts subjects participation. However, (PS08) and (PS33) are papers that reported on issues where subjects opted out or their response omitted from the results. Table 9 shows the papers that reported issues related to informed consent.

Table 9. Breakdown of papers per informed consent.

Disclosure	Comprehension	Competence	Voluntariness	Termination
(PS31); (PS32);	(PS02); (PS06); (PS13); (PS17); (PS19); (PS21); (PS22); (PS25); (PS42); (PS52); (PS53); (PS54); (PS60); (PS63); (PS64); (PS70); (PS38)	(PS07); (PS13); (PS15); (PS16); (PS21); (PS22); (PS24); (PS30); (PS34); (PS35); (PS36); (PS49); (PS65); (PS68); (PS69); (PS61);	(PS02); (PS03); (PS04); (PS05); (PS12); (PS14); (PS25); (PS26); (PS27); (PS28); (PS29); (PS51); (PS54); (PS55); (PS56); (PS57); (PS58); (PS65); (PS63); (PS62); (PS18);	(PS08); (PS33)

In some cases of voluntariness, (PS08) and (PS10) reported that participants did not know it was an experiment as it was part of a course. Students performed tasks as subjects without the knowledge of it at that very moment. Others too reported experiment as being a pre-requisite for a course, leaving them with no choice.

Figure 2 shows that 2 papers reported on disclosure (PS31) and (PS32) had subjects read and sign consent forms before taking part in the experiment. (PS32) also reported specifically that an online consent form was read and agreed by participants before partaking in the experiment. Four papers reported on competence while fourteen on voluntariness and 1 on termination.

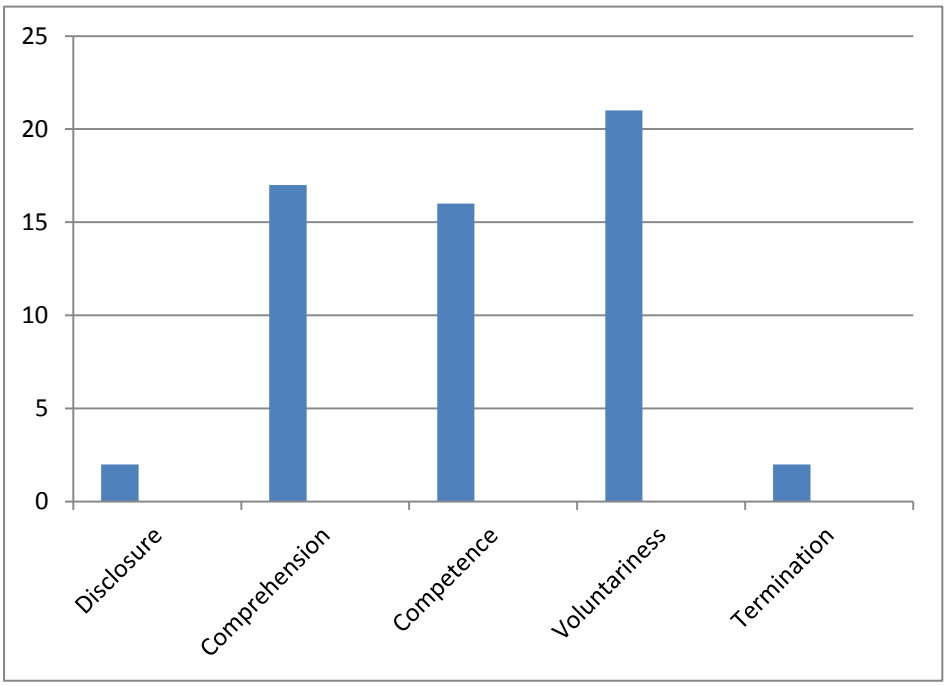

Figure 2. Number of papers over informed consents.

Comprehension had 17 papers that reported on it. For Competence and Voluntariness the number of papers that reported on them was 16 and 21 respectively. (PS08) and (PS33) reported of subjects opting out or data collected from them were omitted for some reason.

6.2 Beneficence

Benefits reported ranges from extra points in a course of study to an honorarium. Cash paid to students are the most reported benefits students get for being used as subjects. Table 10 shows how the reported benefits were distributed out of 27.

Table 10. Papers that reported on Beneficence related issue.

Benefit	Harm
(PS01); (PS09); (PS10); (PS11); (PS12); (PS17); (PS20); (PS23); (PS36); (PS37); (PS39); (PS40); (PS41); (PS43); (PS44); (PS45); (PS46); (PS47); (PS48); (PS50); (PS54); (PS55); (PS59); (PS65); (PS66); (PS67); (PS62);	

Total number of papers that reported on the beneficence related issues that student subjects got from their experiment was 27. Table 10 shows that no paper reported of any harm in terms of ethics in the experiment and all on benefits. Table 11 shows how the reported benefits were distributed out of 27. Some papers however, expressed subjects'

benefits as intellectual gain. Table 11 gives a complete overview of how the papers stated these benefits.

Table 11. Details of rewards for participants.

Reward	means of expression
Cash	honorarium, cash, paid, an amount of
Vouchers	gifts cards, vouchers, lunch
Credits	grade, credits, course requirement, extra points, extra credits
Others	chocolate box, practical experience, intellectual gain

Figure 3 presents the number of papers that reported in the various categories under the benefits or rewards as presented by researchers. The reward system is used in most fields to motivate participation. This is done for subjects during experiments to keep them motivated. Cash payment, vouchers, credits, and others are represented in the diagram.

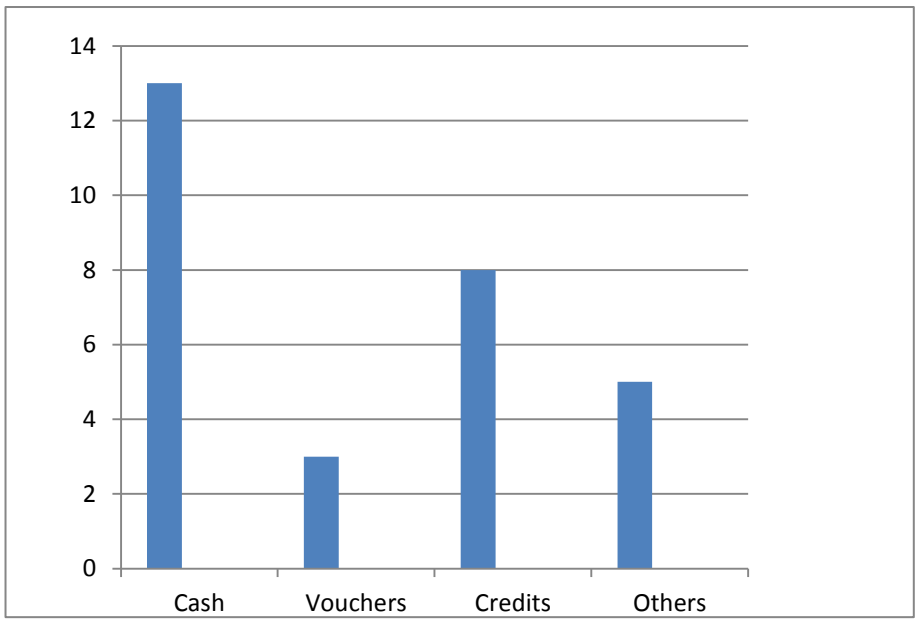

Figure 3. Number of rewards for participants.

Figure 3 gives a complete overview of the number of the rewards as reported. Thirteen papers reported to have paid their subjects with cash, three rewarded them gift cards and lunch etc. Eight, awarded points and credits to subjects in an ongoing course for their participation. Five papers motivated or rewarded student subjects in other ways.

6.3 Confidentiality

In the aspect of confidentiality, (PS38) was found to report on something related to it but was not very clear and was given in the appendix A. Be it anonymity and/or confidentiality, researchers do not frequently report on how they handle that. Table 12 presents some extracted direct statement from some of the reported experiment.

Table 12. Statements from experiment report.

Authors	Statements
(PS08)	mandatory course
(PS49)	no competence or experience in the area of the experiment before conducted on
(PS09)	they were interested in gaining practical experience
(PS10)	granted extra points in their final evaluation at the end of the course
(PS11)	grades on quality of specifications as part of their final course grades
(PS12)	partial fulfilment of the course requirements
(PS01)	subjects were just completing a course
(PS55)	rewarded with extra credit points for their participation
(PS56)	voluntarily participated in the experiment
(PS57)	all subjects participate on a voluntary basis
(PS58)	all participants were unpaid volunteers who had a professional interest in...
(PS59)	the subjects were rewarded with a free lunch

Table 12 shows statements used by authors in their reports that express ethically related issues. These phrases in the experimental report do not follow a format or a special grammar tense. Statements like mandatory course, gaining practical experience, extra points, course requirements, and free lunch were used by researchers in reporting issues. Researchers however chose a way of expressing it.

7. Discussions

This section discusses and analyzes the maps and outcomes of the study by answering the research question with the data gathered. The threat to validity is also discussed in this section.

7.1 Ethically related issues software engineering reports

The most prominent ethical issue in various studies has been informed consent. Experimenters and participants are expected to have an open dialogue to keep ethical concerns raised at a mutual level and understanding. The open process must include confidentiality agreements, declarations, explanations and a vivid description of the experimental procedures. The participant must be given the option to opt out of the experiment whenever pleases. (Singer & Vinson, 2002; Vinson & Singer, 2008; Sieber, 2001.)

Researchers should make it a must to get acquainted with ethical principles. They must have the knowledge of it and how to implement them in their respective research works. Following these set of rules or regulations is not is not sufficient. Decisions made with regards to ethics must be in consideration of the principles provided by Vinson and Singer (2008). The main ethical issue in terms of informed consent is voluntariness, when students are subjects. The researchers usually have these subjects of their research for students.

Recruiting subjects for an experiment is not an easy task; however with researchers' own student minimizes the burden. Sjøberg et al. (2002; 2005), stated how students are preferred to be used in certain conditions instead of professionals. Most researchers indulge the services of their students, making it easier to get a subject for experiment. Vinson and Singer (2008) stated that subjects must participate willingly without coercion. The study has it recorded that some subjects did not have a choice. Participants in works such as PS08 and PS01 had no choice because the experiment was an integral part of a mandatory course of study. However, majority of the papers had their subjects to participate voluntarily.

From the study, it is clear that some students did not have any choice but to participate in the experiment, as it was required to complete a course. In some aspect, it formed part of their grades and points for final grading. This exercise leaves them no choice but to comply with their professors in their research. In another case of voluntariness, when the experiment is part of a course that students performed tasks as subjects without even knowing they were experimented upon at that very moment as reported by (PS08) and (PS10) reported. Others too reported experiment as being a pre-requisite for a course, leaving them no choice. (Sjøberg et al. 2002; Sjøberg et al. 2005.)

Staron (2007b) stated that subjects are affected by the experiments. Their learning experience is affected and however builds up their competence. Students that partake in these researches and studies gain some insight and experience that add to their knowledge and skill (Staron, 2007b). In (PS09) and (PS58), participants took part in the

experiment solely because of their interest in a particular profession and gaining practical experience. They reported that students volunteered to participate because of their interest in a particular profession as the experiment was directed in that field.

None of the reports stated anything specifically in any aspect of confidentiality. This study affirms the literature that, there are ethical issues in using students as subjects but seldom reported. Considering the importance or significance of findings from SE experiments and how they are translated and implemented in developing systems and applications in the industry around the world, it is paramount importance to notice how this study reveals that ethical issues in using students as subjects prevalent, but are not often highlighted or mentioned in most research works.

7.2 Threats to validity

The scope was planned to cover published research from the selected journals and conference proceedings from January 2000 through May 2016. Streamlining the search would pave a way for possibly missing a report that seemingly qualifies for this study. Papers that could be selected and classified for this study are counted and reported upon because their year of publication was out of the time frame considered for this study. A wider scope of the search in terms of years and journals or databases would have limited its threats to validity. Researcher's bias was also a threat as the experience and background of the researcher could have played a role or impacted on the decision to consider certain reports or finding from reports as relevant to the study. The inclusion and the exclusion criteria used in the data gathering for this study could be preserved as a threat to the validity of this work, as it could be argued that inclusion criteria is not robust enough to be unanimously accepted.

8. Conclusion

The purpose of the current study was to identify the ethical issues reported in SE experiments. It was realized that experimenters do not address the ethical aspect of experiment when reporting in SE field. Ethical issues in using students as subjects are rarely reported in SE domain. It was utterly impossible to come out with keywords for classification and categorization scheme. This was so because in the abstracts, introductions and conclusions of reported experiments in the SE area, ethical issues are not reported. Reporting in SE experiment however, focuses on the research question, methodology (when, how, and where), results and its implications (scientifically). (Wohlin, Höst & Henningsson, 2003.)

Organizations have confidence as it depends on software developers and researchers, who produce applications and solutions that meet needs. An industry also recognizes the effects that designers may have on an artefact. This effects that developers have, intentionally or unintentionally, could be based on the ground of ethical or moral values of the engineers. Decision-making becomes very critical and important issue. However, rules and boundaries set by the institution are very important. (Stephenson, 2010.)

In software engineering literature, there seems to be less attention on research ethics. This could be as a result of researcher's unawareness of this aspect. It could also be attributed to the poor style of reporting. As suggested by Moløkken-Østvold (2005), it could be any of these or both. Hall and Flynn (2001) pointed out the less attention given to ethics in research and it is affirmed by Sjøberg et al. (2005). Reporting and debating on ethical issues in the software engineering experimental reports is practically not available (non-existent).

In this area, previous studies have simply been to analyse and adopt from fields such as psychology, medicine, etc. to be used in all discipline of computing. However, the attention and focus has only been on participants and how they must be treated, be it students or professionals but not in reporting. (Davison. et al. 2001; Singer & Vinson, 2002, Singer & Vinson, 2001; Davis, 2001; Storey, Phillips & Maczewski, 2001.)

Papers like (PS67) and (PS17) reported on the ethical issue in a footnote while (PS38) in the appendix. Others reported on it in the chapter that describes the experimental settings. This means that there is no definite format or way research ethics is reported and discussed in software engineering experiments. This important aspect of research or experiment has received very little or no attention in SE.

There are two main factors that contribute to this negligence. Firstly, in SE experimentation, the main point of interest is the result, the interpretation of the results and its effect on the Computing environment and the world at large. Fellow researchers across the world take a keen interest in the result and how to develop upon it. Secondly, experimentation in SE domain has not been considered as a dangerous thing, because it presents no or very little harm as compared to other fields of study. These have however contributed to researchers just focusing on the results and the implication of their study to the world.

This study was purposed to bring to attention an important issue that has been ignored. The study was to report on ethical issues that are mentioned by researchers in their reports. Also to create awareness on ethics as it seems neglected in SE experiment. However, it came to notice that ethical issues in research are not reported. This study is an affirmation that this vital aspect of research or experiment is neglected or given less attention when reporting in SE field. It has also been proven that there is no laid down structure for reporting on experiments that suggest it to the researchers.

If ethics in using humans are not critically scrutinized in the SE field, for reported experiments to merit in all aspects, it is expected that ethical issues are debated and scrutinized and reported in SE. (Moløkken-Østvold, 2005.) I am of the view that reporting ethical issues confronted are essential and will enrich the report. Future studies could consider developing a definite format or template for reporting experiment and the SE society could ensure and emphasize the need to include ethical issues in all research works that deals with humans.

References

Abrahamsson, P., Kettunen, P., & Fagerholm, F. (2010, June). The set-up of a software engineering research infrastructure of the 2010s. In *Proceedings of the 11th International Conference on Product Focused Software* (pp. 112-114).

ACM/IEEE-CS. (2000). Software Engineering Codes of Ethics. Accessed: 22.11.2015. Available: http://ieeezplore.ieee.org/stamp.jsp?tp=&arnumber=796142

Alfano, M. & Loeb, D. (2014). Moral Judgments and Intuitions. Accessed: 27.10.2015. Available: http://plato.stanford.edu/entries/experimental-moral/

Babar, M. A., & Zhang, H. (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. In *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement,* pp. 346-35.

Baier, K., 1958, The Moral Point of View: A Rational Basis of Ethics. Cornell University Press, New York, pp. 187-191 Accessed: 20.04.2016. Available: https://www.uta.edu/philosophy/faculty/burgess-jackson/Baier.pdf

Basili, V. R., Selby, R. W., & Hutchens, D. H. (1986). Experimentation in software engineering. *IEEE Transactions on Software Engineering*, (7), 733-743.

Berry, D. M., & Tichy, W. F. (2003). Comments on" Formal Methods Application: An Empirical Tale of Software Development". *IEEE Transactions on Software Engineering*, *29*(6), 567.

Billington, D. P. (2006). Teaching Ethics in Engineering Education through Historical Analysis. *Science and Engineering Ethics*, 12, 205-222

Bird, S. J. (2003). Ethics as a Core Competency in Science and Engineering. *Science and Engineering Ethics* 9, 443-444

Bird, S. J. & Sieber J. E. (2005). Teaching Ethics in science and Engineeringg; Effective Online Education. *Science and Engineering Ethics*, 11, 323/-328

Boehm, B.W. (1981). Software Engineering Economics, Prentice Hall, NJ.

Bouville, M. (2008). On Using Ethical Theories to Teach Engineering Ethics. *Science and Engineering Ethics*, 14, 111-120

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, Vol. 80, no. 4, pp. 571-583.

Brian, S. (2005). Teaching Research Ethics: Can Web-Based Instruction Satisfy Appropriate Pedagogical Objectives? *Science and Engineering Ethics*, 11, 347–366

Budgen, D., Turner, M., Brereton, P., & Kitchenham, B. (2008). Using mapping studies in software engineering. In *Proceedings of Psychology of Programming Interest Group, Vol.8,* pp. 195-204.

Burmeister, O. K., & Weckert, J. (2003). Applying the new software engineering code of ethics to usability engineering: A study of four cases. *Journal of Information, Communication and Ethics in Society*, *1*(3), 119-132.

Carver, J., Jaccheri, L., Morasca, S., & Shull, F. (2003, September). Issues in using students in empirical studies in software engineering education. In *Software Metrics Symposium, 2003. Proceedings. Ninth International* (pp. 239-249). IEEE.

Charlesworth, M., & Sewry, D. A. (2002, September). Ethical issues in enabling information technologies. In *Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology* (pp. 163-171). South African Institute for Computer Scientists and Information Technologists

Crisp, R. (2013). *The Oxford handbook of the history of ethics*. Oxford University Press. pp. 21- 841.

Crowel S., (2012). Why is Ethics First Philosophy? Levinas in Phenomenological Context. Accessed: 11.12.2013. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0378.2012.005500.x/pdf

Davis, M. (1998). Thinking like an engineer: Studies in the ethics of a profession. New York: Oxford University Press. pp.143.

Davis, M. (2001). When is a Volunteer Not a Volunteer? *Empirical Software Engineering, 6*(4): pp. 349- 352.

Davison, R. M., Kock, N., Loch, K. D., & Clarke, R. (2001). Research ethics in information systems: would a code of practice help? *Communications of the Association for Information Systems*, *7*(1), 4.

de Magalhães, C. V., da Silva, F. Q., Santos, R. E., & Suassuna, M. (2015). Investigations about replication of empirical studies in software engineering: A systematic mapping study. *Information and Software Technology*, *64*, 76-101.

Elms A. C. (1998). Social Psychology and Social Relevance, Chapter 4. Little, Brown and Company Boston. pp. 146-160. Accessed: 21/12/2015. Available: http://www.ulmus.net/ace/library/obedienceresearchethics.html.

Faden, R. R. & Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press. pg, 4 - 33

Fernández-Sáez, A. M., Genero, M., & Chaudron, M. R. (2013). Empirical studies concerning the maintenance of UML diagrams and their use in the maintenance of code: A systematic mapping study. *Information and Software Technology*, *55*(7), 1119-1142.

Fieser J., (N/A), Ethics, Internet Encyclopedia of Philosophy. Accessed: 15.7.2016. Available: http://www.iep.utm.edu/ethics/

Fleischmann, S. T. (2004). Essential Ethics – Embedding Ethics into an Engineering Curriculum. *Science and Engineering Ethics*, 10, 369-381

Fleischmann, S., T. (2006). Teaching Ethics: More than an Honor Code. *Science and Engineering Ethics*, 12,381-389

Fleuhr-Lobban, C. (1994). Informed consent in anthropological research: we are not exempt. *Human Organization*, *53*(1), 1–10.

Freedman, B. (1987). Scientific value and validity as scientific requirements for research: a proposed explication. *IRB: Ethics and Human Research, 9*(6), 7–10.

Gotterbarn, D. (2001). Software Engineering Ethics. Ethics. Encyclopedia of Software Engineering

Gotterbarn, D., Miller, K., & Rogerson, S. (1999). Computer Society and ACM Approve Software Engineering Code of Ethics, *Computer*, 31(10), Oct: 84-88.

Hall, T., & Flynn, V. (2001). Ethical issues in software engineering research: a survey of current practice. *Empirical Software Engineering*, *6*(4), 305-317.

Herkert, J. R. (2000). Engineering Ethics Education in the USA Content, Pedagogy and Curriculum. *European Journal of Engineering Education*, 25(4), 303-313.

Höfer, A., & Tichy, W. F. (2007). Status of empirical research in software engineering. In *Empirical Software Engineering Issues. Critical Assessment and Future Directions* (pp. 10-19). Springer Berlin Heidelberg

Höst, M., Regnell, B., & Wohlin, C. (2000). Using students as subjects—a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Engineering*, *5*(3), 201-214.

Jonsen, A. R. (1978). Do No Harm. *Annals of Internal Medicine, 88*(6), 827-832.

Kitchenham, B. A. (2004, September). Systematic reviews. In *Software Metrics, 2004. Proceedings. 10th International Symposium on* (pp. xii-xii). IEEE.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Version 2.3. *URL http://www. dur. ac. uk/ebse/resources/Systematic-reviews-5-8. pdf*.

Kitchenham, B. A., Budgen, D., & Brereton, O. P. (2011). Using mapping studies as the basis for further research–a participant-observer case study. *Information and Software Technology, 53(6*), 638-651.

Kuhar, M. J. (2006). Should Codes of Ethics include expectations of others? *Science and Engineering Ethics*, 12,413-414

Kuzniarz, L., Staron, M., & Wohlin, C. (2003). Students as study subjects in software engineering experimentation. In *Third Swedish Conference on Software Engineering Research and Practise in Sweden: Proceedings (SERPS'03)* (pp. 19-24).

Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2016). Environmental factors and features that influence visual search in a 3D WIMP interface. *International Journal of Human-Computer Studies*, *92*, 30-43.

Landauer, J., and Rowlands, J. (2001). Importance of philosophy. Assessed: 4.4.2014. Available: http://www.importanceofphilosophy.com/Ethics_Main.html

Lau, A. S. (2004). Teaching Engineering Ethics to First-Year College Students. *Science and Engineering Ethics, 10*,359-368

Laudon, K. C., & Laudon, J. P. (2010). Management information systems: managing the digital firm. New Jersey, 8.pp. 151

Land, L. P. W., Tan, B. C., & Bin, L. (2005, November). Investigating training effects on software reviews: a controlled experiment. In *Empirical Software Engineering, 2005. 2005 International Symposium on* (pp. 11-pp). IEEE.

Loui, M. C. (1997). Preface to a special section: practicing and teaching ethics in engineering and computing. *Science and Engineering Ethics, 3*, 423 – 431

MacIntyre, A. (2003). *A Short History of Ethics: a history of moral philosophy from the Homeric age to the 20th century*. 2nd edition. Routledge. pp. 18 – 1571

MacIntyre, A., (2007). After Virtue. Third edition, Notre Dame: Notre Dame University Press. pp. 181-203

Marshall, C., & Brereton, P. (2013, October). Tools to support systematic literature reviews in software engineering: A mapping study. In *Empirical Software Engineering and Measurement, 2013 ACM/IEEE International Symposium on* (pp. 296-299). IEEE.

McLeod, S. A. (2015). Psychology Research Ethics. Accessed: 12.12.2015 Available: www.simplypsychology.org/Ethics.html

McNeill, P. (1993). *The Ethics and Politics of Human Experimentation*. New York: Cambridge University Press.

Moløkken-Østvold, K. (2005, January). Ethical concerns when increasing realism in controlled experiments with industrial participants. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 264a-264a). IEEE.

Nakagawa, E. Y., Feitosa, D., & Felizardo, K. R. (2010, May). Using systematic mapping to explore software architecture knowledge. In *Proceedings of the 2010 ICSE Workshop on Sharing and Reusing Architectural Knowledge* (pp. 29-36). ACM.

NAP, (2009). National Academy of Sciences. On Being a Scientist: Third Edition. Washington, DC: The national Academies Press. Available at: http://www.nap.edu/catalog.php?record_id=12192.) Accessed: 1.1.2016

Narvaez, D., & Rest, J. (1995). The four components of acting morally. In W. Kurtines & J. Gewirtz (Eds.), Moral behavior and moral development: An introduction (385-400). New York: McGraw-Hill.

Narayanan, A. (2013). Ethical dilemmas faced by software engineers: A request for real-world examples. Accessed: 12.1.2016 Available: https://freedom-to-tinker.com/blog/randomwalker/ethical-dilemmas-faced-by-software-engineers-a-request-for-real-world-examples/

Paternoster, N., Giardino, C., Unterkalmsteiner, M., Gorschek, T., & Abrahamsson, P. (2014). Software development in startup companies: A systematic mapping study. *Information and Software Technology*, *56*(10), 1200-1218.

PB, (2008), The basics of philosophy. Accessed: 21.06.2016. Available: http://www.philosophybasics.com/branch_ethics.html

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June). Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering* (Vol. 17, No. 1).

Reed, K. (2000). Software engineering—A new millennium? *IEEE Software, 17*(4), 107.

Rest, J. (1984). "The Major Components of Morality," In W. M. Kurtines, J. L. Gewirtz (eds.) Morality, Moral Behavior, and Moral Development. New York: A Wiley-Interscience Publication. pp. 24–38.

Resnik, D.B. (2015). What is Ethics in Research & Why is it Important? Accessed: 1.1.2016. Available: http://www.niehs.nih.gov/research/resources/bioethics/whatis/

Riaz, M., Sulayman, M., Salleh, N., & Mendes, E. (2010). Experiences conducting systematic reviews from novices' perspective. In *Proceedings. of EASE* Vol. 10, pp. 1-10.

Runeson, P. (2003, April). Using students as experiment subjects–an analysis on graduate and freshmen student data. In *Proceedings of the 7th International Conference on Empirical Assessment in Software Engineering*.–Keele University, UK (pp. 95-102).

SEI, (2014). What is software Engineering? Accessed: 09.12.2015. Available: http://www.softwareengineerinsider.com/articles/what-is-software-engineering.html#.VnCYRF5LdNu

Sieber, J. (2001). Protecting Research Subjects, Employees and Researchers: Implications for Software Engineering. *Empirical Software Engineering*, 6(4): pp. 329-341

Singer, J., & Vinson, N. (2002). Ethical issues in empirical studies of software engineering.

Singer, J., & Vinson, N. (2001). Why and How Research Ethics Matters to You. Yes, YOU! *Empirical Software Engineering*, 6, 4:287-290.

Sjøberg, D. (2005). Use of Professionals in Experiments, Letter to Editor, *IEEE Software*, 22(5): 9-10,

Sjoberg, D. I., Anda, B., Arisholm, E., Dyba, T., Jorgensen, M., Karahasanovic, A., ... & Vokác, M. (2002). Conducting realistic experiments in software engineering. In *Empirical Software Engineering, 2002. Proceedings. 2002 International Symposium n* (pp. 17-26). IEEE.

Sjøberg, D.I.K., Hannay, J.E., Hansen, O., Kampenes, V.B., Karahasanović, A., Liborg, N.-K., & Rekdal, A.C. (2005). A Survey of Controlled Experiments in Software Engineering, *IEEE Transactions on Software Engineering*, 31(9): 733–753,

Skovdal, M., & Abebe, T. (2012). Reflexivity and dialogue: methodological and socio-ethical dilemmas in research with HIV-affected children in East Africa. *Ethics, policy & environment*, *15*(1), 77-96.

Smith, D. (2003). Five principles for research ethics. *Monitor on psychology*, *34*(1), 56. Available: http://www.apa.org/monitor/jan03/principles.aspx, Accessed: 17.12.2015

Spinello, R. A. (1997). Frameworks for Ethical Analysis Case Studies in Information and Computer Ethics. Upper Saddle River: Prentice Hall. pp. 35-50

Staron, M. (2007a). Using Experiments in Software Engineering as an Auxiliary Tool for Teaching--A Qualitative Evaluation from the Perspective of Students' Learning Process. In *Proceedings of the 29th international conference on Software Engineering* (pp. 673-676). IEEE Computer Society.

Staron, M. (2007b). Using Students as Subjects in Experiments--A Quantitative Analysis of the Influence of Experimentation on Students' Learning Process. In *Software Engineering Education & Training, 2007. CSEET'07. 20th Conference on* (pp. 221-228).

Stephenson, J. H. (2010). Ethics and morality in software development: A developer's perspective (Doctoral dissertation, Capella University).

Storey, M. A., Phillips, B., & Maczewski, M. (2001). Is it Ethical to Evaluate Web-based Learning Tools using Students?. *Empirical Software Engineering*, *6*(4), 343-348.

Svahnberg, M., Aurum, A., & Wohlin, C. (2008, October). Using students as subjects-an empirical evaluation. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement* (pp. 288-290). ACM.

TFD. (2015). The free dictionary: Experimentation. Accessed: 10.02.15. Available: http://www.thefreedictionary.com/experimentation

Tichy, W.F. (2000) "Hints for Reviewing Empirical Work in Software Engineering," *Empirical Software Engineering, 5*(4), pp. 309-312,

Vartiainen, T. (2006, February). Moral conflicts perceived by students of a project course. *In Proceedings of the 6th Baltic Sea conference on Computing education research*: Koli Calling 2006 (pp. 77-84). ACM.

Vartiainen, T. (2009). Moral problems perceived by industry in collaboration with a student group: balancing between beneficial objectives and upholding relations. *Journal of Information Systems Education*, 20(1), 51.

Vinson, N. G., & Singer, J. (2008). A practical guide to ethical research involving humans. In *Guide to Advanced Empirical Software Engineering* (pp. 229-256). Springer London.

Vinson, N., & Singer, J. (2001). Getting to the Source of Ethical Issues. *Empirical Software Engineering.6* (4):293-297.

Webopedia. (2016). Software engineering. Accessed: 1.12.2015. Available: http://www.webopedia.com/TERM/S/software_engineering.html

Wohlin, C., Höst, M., & Henningsson, K. (2003). Empirical research methods in software engineering. In *Empirical methods and studies in software engineering* (pp. 7-23). Springer Berlin Heidelberg.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Experimentation in software engineering. Springer Science & Business Media. pp. 85-157

Appendix A. Reference for the Primary Study

Study code	Source
PS01	Benander, A. C., Benander, B. A., & Sang, J. (2000). An empirical analysis of debugging performance–differences between iterative and recursive constructs. *Journal of Systems and Software, 54*(1), 17-28.
PS02	Ghanbari, H., Similä, J., & Markkula, J. (2015). Utilizing online serious games to facilitate distributed requirements elicitation. *Journal of Systems and Software, 109*, 32-49.
PS03	Jørgensen, M., Faugli, B., & Gruschke, T. (2007). Characteristics of software engineers with optimistic predictions. *Journal of Systems and Software, 80*(9), 1472-1482.
PS04	Ghabi, A., & Egyed, A. (2015). Exploiting traceability uncertainty among artifacts and code. *Journal of Systems and Software, 108*, 178-192.
PS05	Müller, M. M. (2005). Two controlled experiments concerning the comparison of pair programming to peer review. *Journal of Systems and Software, 78*(2), 166-179.
PS06	Baker, A., Navarro, E. O., & Van Der Hoek, A. (2005). An experimental card game for teaching software engineering processes. *Journal of Systems and Software, 75*(1), 3-16.
PS07	Andrzejczak, C., & Liu, D. (2010). The effect of testing location on usability testing performance, participant stress levels, and subjective testing experience. *Journal of Systems and Software, 83*(7), 1258-1266.
PS08	Glezer, C., Last, M., Nachmany, E., & Shoval, P. (2005). Quality and comprehension of UML interaction diagrams-an experimental comparison. *Information and Software Technology, 47*(10), 675-692.
PS09	Kamsties, E., von Knethen, A., & Reussner, R. (2003). A controlled experiment to evaluate how styles affect the understandability of requirements specifications. *Information and Software Technology, 45*(14), 955-965.
PS10	Cruz-Lemus, J. A., Genero, M., Caivano, D., Abrahão, S., Insfrán, E., & Carsí, J. A. (2011). Assessing the influence of stereotypes on the comprehension of UML sequence diagrams: A family of experiments. *Information and Software Technology, 53*(12), 1391-1403
PS11	Kabeli, J., & Shoval, P. (2005). Comprehension and quality of analysis specifications—a comparison of FOOM and OPM methodologies. *Information and Software Technology*, 47(4), 271-290.
PS12	Ng, T. H., Yu, Y. T., Cheung, S. C., & Chan, W. K. (2012). Human and program factors affecting the maintenance of programs with deployed design patterns. *Information and Software Technology, 54*(1), 99-118
PS13	Saleh, F., & El-Attar, M. (2015). A scientific evaluation of the misuse case diagrams visual syntax. *Information and Software Technology, 66*, 73-96.
PS14	Han, D., Yang, Q., Xing, J., Li, J., & Wang, H. (2016). FAME: A UML-based framework for modeling fuzzy self-adaptive software. *Information and Software Technology, 76*, 118-134.
PS15	Collins, R. W., Hevner, A. R., Walton, G. H., & Linger, R. C. (2008). The impacts of function extraction technology on program comprehension: A controlled experiment. *Information and Software Technology, 50*(11), 1165-1179.
PS16	Juristo, N., Vegas, S., Solari, M., Abrahão, S., & Ramos, I. (2013). A process for managing interaction between experimenters to get useful similar replications. *Information and Software Technology, 55*(2), 215-225.
PS17	Jørgensen, M., & Sjøberg, D. I. (2001). Impact of effort estimates on software project work. *Information and software technology, 43*(15), 939-948.
PS18	Abrahão, S., & Poels, G. (2007). Experimental evaluation of an object-oriented function point measurement procedure. *Information and Software Technology, 49*(4), 366-380.

Study Code	Source
PS19	Ng, T. H., & Cheung, S. C. (2005). Enhancing class commutability in the deployment of design patterns. *Information and Software Technology, 47*(12), 797-804.
PS20	Bipp, T., Lepper, A., & Schmedding, D. (2008). Pair programming in software development teams–An empirical study of its benefits. *Information and Software Technology, 50*(3), 231-240.
PS21	Sabaliauskaite, G., Kusumoto, S., & Inoue, K. (2004). Assessing defect detection performance of interacting teams in object-oriented design inspection. *Information and Software Technology, 46*(13), 875-886.
PS22	Yin, Z., Dunsmore, A., & Miller, J. (2004). Self-assessment of performance in software inspection processes. *Information and Software Technology, 46*(3), 185-194.
PS23	Reynoso, L., Manso, E., Genero, M., & Piattini, M. (2010). Assessing the influence of import-coupling on OCL expression maintainability: A cognitive theory-based perspective. *Information Sciences, 180*(20), 3837-3862.
PS24	Yu, N., & Kong, J. (2016). User experience with web browsing on small screens: Experimental investigations of mobile-page interface design and homepage design for news websites. *Information Sciences, 330*, 427-443.
PS25	Chung, W. (2006). Studying information seeking on the non-English Web: An experiment on a Spanish business Web portal. *International Journal of Human-Computer Studies, 64*(9), 811-829.
PS26	Hsu, Y. C., & Schwen, T. M. (2003). The effects of structural cues from multiple metaphors on computer users' information search performance. *International Journal of Human-Computer Studies, 58*(1), 39-55.
PS27	Tractinsky, N., Cokhavi, A., Kirschenbaum, M., & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International journal of human-computer studies, 64*(11), 1071-1083.
PS28	Zhao, J., Soukoreff, R. W., & Balakrishnan, R. (2015). Exploring and modeling unimanual object manipulation on multi-touch displays. *International Journal of Human-Computer Studies, 78*, 68-80.
PS29	Kwon, J. H., Powell, J., & Chalmers, A. (2013). How level of realism influences anxiety in virtual reality environments for a job interview. *International journal of human-computer studies, 71*(10), 978-987.
PS30	Kalnikaitė, V., & Whittaker, S. (2010). Beyond being there? Evaluating augmented digital records. *International journal of human-computer studies, 68*(10), 627-640.
PS31	Purchase, H. C., Welland, R., McGill, M., & Colpoys, L. (2004). Comprehension of diagram syntax: an empirical study of entity relationship notations. *International Journal of Human-Computer Studies, 61*(2), 187-203.
PS32	Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2016). Environmental factors and features that influence visual search in a 3D WIMP interface. *International Journal of Human-Computer Studies, 92*, 30-43
PS33	Hoffmann, L., & Krämer, N. C. (2013). Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. *International Journal of Human-Computer Studies*, *71*(7), 763-774.
PS34	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.
PS35	Tran, C. D., Ezzedine, H., & Kolski, C. (2013). EISEval, a generic reconfigurable environment for evaluating agent-based interactive systems. *International Journal of Human-Computer Studies*, *71*(6), 725-761.
PS36	Roth, S. P., Tuch, A. N., Mekler, E. D., Bargas-Avila, J. A., & Opwis, K. (2013). Location matters, especially for non-salient features–An eye-tracking study on the effects of web object placement on different types of websites. *International journal of human-computer studies*, *71*(3), 228-235.
PS37	Whitley, K. N., Novick, L. R., & Fisher, D. (2006). Evidence in favor of visual representation for the dataflow paradigm: An experiment testing LabVIEW's comprehensibility. *International journal of human-computer studies*, *64*(4), 281-303.

Study Code	Source
PS38	Stumpf, S., Rajaram, V., Li, L., Wong, W. K., Burnett, M., Dietterich, T., ... & Herlocker, J. (2009). Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies*, *67*(8), 639-662.
PS39	Nova, N., Girardin, F., & Dillenbourg, P. (2010). The effects of mutual location-awareness on group coordination. *International journal of human-computer studies*, *68*(7), 451-467.
PS40	Lindley, S. E., & Monk, A. F. (2008). Social enjoyment with electronic photograph displays: Awareness and control. *International Journal of Human-Computer Studies*, *66*(8), 587-604.
PS41	McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies*, *69*(1), 70-79.
PS42	Jan, J. C., Chen, C. M., & Huang, P. H. (2016). Enhancement of digital reading performance by using a novel web-based collaborative reading annotation system with two quality annotation filtering mechanisms. *International Journal of Human-Computer Studies*, *86*, 81-93.
PS43	Shirazi, A. S., Clawson, J., Hassanpour, Y., Tourian, M. J., Schmidt, A., Chi, E. H., ... & Van Laerhoven, K. (2013). Already up? using mobile phones to track & share sleep behavior. *International Journal of Human-Computer Studies*, *71*(9), 878-888.
PS44	Janssen, J. H., IJsselsteijn, W. A., & Westerink, J. H. (2014). How affective technologies can influence intimate interactions and improve social connectedness. *International Journal of Human-Computer Studies*, *72*(1), 33-43
PS45	Butavicius, M. A., Lee, M. D., Pincombe, B. M., Mullen, L. G., Navarro, D. J., Parsons, K. M., & McCormac, A. (2012). An assessment of email and spontaneous dialog visualizations. *International Journal of Human-Computer Studies*, *70*(6), 432-449.
PS46	Bauerly, M., & Liu, Y. (2006). Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies*, *64*(8), 670-682.
PS47	McKechnie, S., & Nath, P. (2016). Effects of new-to-market e-store features on first time browsers. *International Journal of Human-Computer Studies*, *90*, 14-26.
PS48	Shahar, N., Meyer, J., Hildebrandt, M., & Rafaely, V. (2012). Detecting system failures from durations and binary cues. *International Journal of Human-Computer Studies*, *70*(8), 552-560.
PS49	Hattori, L., D'Ambros, M., Lanza, M., & Lungu, M. (2013). Answering software evolution questions: An empirical evaluation. *Information and Software Technology*, *55*(4), 755-775.
PS50	Rozado, D., El Shoghri, A., & Jurdak, R. (2015). Gaze dependant prefetching of web content to increase speed and comfort of web browsing. *International Journal of Human-Computer Studies*, *78*, 31-42.
PS51	Tanin, E., Shneiderman, B., & Xie, H. (2007). Browsing large online data tables using generalized query previews. *Information Systems*, *32*(3), 402-423.
PS52	Shan, Y. (2016). How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. *Computers in Human Behavior*, *55*, 633-641.
PS53	Sha, L., Looi, C. K., Chen, W., Seow, P., & Wong, L. H. (2012). Recognizing and measuring self-regulated learning in a mobile learning environment. *Computers in Human Behavior*, *28*(2), 718-728.
PS54	Kim, H. C., & Hyun, M. Y. (2016). Predicting the use of smartphone-based Augmented Reality (AR): Does telepresence really help?. *Computers in Human Behavior*, *59*, 28-38.
PS55	Hendrix, D., Cross II, J. H., & Maghsoodloo, S. (2002). Corrections to" The Effectiveness of Control Structure Diagrams in Source Code Comprehension Activities". *IEEE Transactions on Software Engineering*, *28*(6), 624.
PS56	Vitharana, P., Zahedi, F. M., & Jain, H. (2003). Knowledge-based repository scheme for storing and retrieving business components: a theoretical design and an empirical analysis. *IEEE transactions on Software Engineering*, *29*(7), 649-664.

Study Code	Source
PS57	Cornelissen, B., Zaidman, A., & van Deursen, A. (2011). A controlled experiment for program comprehension through trace visualization. *IEEE Transactions on Software Engineering*, *37*(3), 341-355.
PS58	Perepletchikov, M., & Ryan, C. (2011). A controlled experiment for evaluating the impact of coupling on the maintainability of service-oriented software. *IEEE transactions on software engineering*, *37*(4), 449-465.
PS59	Dig, D., Manzoor, K., Johnson, R. E., & Nguyen, T. N. (2008). Effective software merging in the presence of object-oriented refactorings. *IEEE Transactions on Software Engineering*, *34*(3), 321-335.
PS60	Ricca, F., Di Penta, M., Torchiano, M., Tonella, P., & Ceccato, M. (2010). How developers' experience and ability influence web application comprehension tasks supported by UML stereotypes: A series of four experiments. *IEEE Transactions on Software Engineering*, *36*(1), 96-118.
PS61	Arisholm, E., Briand, L. C., Hove, S. E., & Labiche, Y. (2006). The impact of UML documentation on software maintenance: An experimental evaluation. *IEEE Transactions on Software Engineering*, *32*(6), 365-381.
PS62	Aranda, A. M., Dieste, O., & Juristo, N. (2016). Effect of Domain Knowledge on Elicitation Effectiveness: An Internally Replicated Controlled Experiment. *IEEE Transactions on Software Engineering*, *42*(5), 427-451.
PS63	Mäntylä, M. V., & Lassenius, C. (2009). What types of defects are really discovered in code reviews?. *IEEE Transactions on Software Engineering*, *35*(3), 430-448.
PS64	Krein, J. L., Prechelt, L., Juristo, N., Nanthaamornphong, A., Carver, J. C., Vegas, S., ... & Eggett, D. L. (2016). A Multi-Site Joint Replication of a Design Patterns Experiment Using Moderator Variables to Generalize across Contexts. *IEEE Transactions on Software Engineering*, *42*(4), 302-321.
PS65	Anda, B., & Sjoberg, D. I. K. (2003, September). Applying use cases to design versus validate class diagrams-a controlled experiment using a professional modeling tool. In *Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on* (pp. 50-60). IEEE.
PS66	Abdelnabi, Z., Cantone, G., Ciolkowski, M., & Rombach, D. (2004, August). Comparing code reading techniques applied to object-oriented software frameworks with regard to effectiveness and defect detection rate. In *Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on* (pp. 239-248). IEEE.
PS67	Land, L. P. W., Tan, B. C., & Bin, L. (2005, November). Investigating training effects on software reviews: a controlled experiment. In *Empirical Software Engineering, 2005. 2005 International Symposium on* (pp. 11-pp). IEEE.
PS68	Yoo, Y., & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS quarterly*, 371-390.
PS69	Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS quarterly*, 401-426.
PS70	Vegas, S., Apa, C., & Juristo, N. (2016). Crossover Designs in Software Engineering Experiments: Benefits and Perils. *IEEE Transactions on Software Engineering*, *42*(2), 120-135.