جامعة المجمعة
Majmaah University

JEAS

JOURNAL OF ENGINEERING AND APPLIED SCIENCES

A Refereed Academic Journal Published by the
Publishing and Translation Center at Majmaah University

Vol. 7 Issue (2) (November. 2020) ISSN: 1658 - 6638

مركز النشر والترجمة جامعة المجمعة
Publishing & Translation Center
Publishing and Translation Center - Majmaah University

IN THE NAME OF ALLAH,
THE MOST GRACIOUS,
THE MOST MERCIFUL

Kingdom of Saudi Arabia
Ministry of Education
Majmaah University

جامعة المجمعة
Majmaah University

JEAS

JOURNAL OF ENGINEERING AND APPLIED SCIENCES

A Refereed Academic Journal Published by the
Publishing and Translation Center at Majmaah Universtiy

Vol. 7, Issue (2) (November - 2020) ISSN: 1658 - 6638

مركز النشر والترجمة جامعة المجمعة
Publishing & Translation Center
Publishing & Translation Center - MU

About the Journal

Journal of Engineering and Applied Sciences (JEAS)

Vision

Pioneer journal in the publication of advanced research in engineering and applied sciences.

Mission

A peer-review process which is transparent and rigorous

Objectives

a) Support research that addresses current problems facing humanity.
b) Provide an avenue for exchange of research interests and facilitate the communication among researchers.

Scope

JEAS accepts articles in the field of engineering and applied sciences. Engineering areas covered by JEAS include:

Engineering areas

Architectural Engineering
Chemical Engineering
Civil Engineering
Computer Engineering
Electrical Engineering
Environmental Engineering
Industrial Engineering
Mechanical Engineering

Applied Sciences areas

Applied Mathematics
Applied Physics
Biological Science
Biomathematics
Biotechnology
Computer Sciences
Earth Science
Environmental Science

Computer Sciences areas

Computer Sciences
Information Technology
Information Sciences
Computer Engineering

Correspondence and Subscription

Majmaah University, Post Box 66, Al-Majmaah 11952, KSA
email: jeas@mu.edu.sa

Journal of Engineering and Applied Sciences

Journal of Engineering and Applied Sciences

Editorial

Scientific publishing has brought many challenges to authors. With increasing number of scientific journals, varying scopes, reviewing requirements, and cost of publishing to authors, finding the right journal to publish an article is a decision many authors must bitterly confront and resolve. The publication of scientific findings is an integral part of the life of researchers. The process of publishing has evolved to become an efficient system of decimating knowledge and collaboration among scientists. Science journals have institutionalized procedures to manage large volume of article submissions per year. In many cases, journals began to define narrower scopes for a dual purpose: managing submissions and delivering outstanding research.

Based on recent studies, the scientific publishing world consists of more than 25 thousand active journals in various disciplines and fields. Science Direct hosts 3,348 journals (as of February 2014). The Directory of Open Access Journals lists in its search engine more than 9,800 open access online journals.

According to recent estimates, the number of scientific journals grows by 3% per year worldwide. With this large number of journals, journals may find it harder to stay afloat.

In its inauguration, the board of editors is honored to introduce to the scientific community the Journal of Engineering and Applied Sciences - JEAS, another scientific journal from Majmaah University. The board has pledged a commitment to JEAS authors and readers to bring the most dynamic and vibrant journal management with better satisfaction.

Dr. Mohamed Alshehri

Contents

ORIGINAL ARTICLES

Journal of Engineering and Applied Sciences, Vol. 7, Issue (2) November 2020

Network Intrusion Detection Approach using Machine Learning Based on Decision Tree Algorithm

Elmadena M. Hassan*
Computer Science Department, FMCS, University of Gazira, Wad madani,
Sudan, madinamohammed@uofg.edu.sd
Mohammed A. Saleh
Computer Department, College of Science and Arts in Ar Rass Qassim University,
Kingdom of Saudi Arabia, m.saleh@qu.edu.sa
Awadallah M. Ahmed
Computer Science Department, FMCS, University of Gazira , Wad madani,
Sudan, awadallahd@uofg.edu.sd

Abstract

Computer security, as well cyber security, is safeguarding information systems from stealing, destruction, and misusing computer hardware, software, data, and the delivered services. In general, machine learning is the area of studying, which grants a computer system to grasp, although not explicitly programmed. Often, anomaly-based Intrusion Detection Systems (IDS) experiences high false alarms rates (FAR), and since many different mechanisms are used by the researchers to protect the system from high false alarms and least detection rates, the challenge is to reduce high false alarms and achieve high detection rate is remain, and therefore; a new approach need to be applied. The objective of this study is to specify a network traffic technique to distinguish the normal from abnormal attacks, and also, to use specific algorithm to reduce the high false alarms rate (FAR). The dataset used in this study is NSL-KDD, where the data are divided into two parts (60%) for training and (40%) for testing. The results show that the decision tree (DT) algorithm achieved high detection rate (DR) and low false alarms rate (FAR) in comparison with other machine learning algorithms. This study achieved rate of detection for random tree about (99.7%) and for J48 about (99.8%), but for naïve Bayes about (86.8%). Also, the rate of false alarm for random tree about (0.2%) and for J48 about (0.3%), but for naïve Bayes about (6%), and hence the researchers concluded that the decision tree algorithm accomplishes high detection rate (DR), and low false alarms rate (FAR) compared to other algorithms of machine learning.

Keywords:
Decision Tree Algorithm; NSL-KDD dateset; Anomaly Detection

1. Introduction

As the size and class of electronic network attacks increase, it becomes tough to discover intrusion into the network of interests. Advanced persistent threats (APTs) became a greater threat to all companies and nation states. Cybercriminals are still using more refined technologies to illicitly access systems, and corporations and staff are increasingly to adopting new, more refined technologies and networks in the workplace. All these elements considerably hinder the mission of defending the network and trends suggest that these problems are possible to be extended over time. Fighting these difficulties need totally different tools and strategies to discover and defend attacks.

Decision tree is a method from the sphere of data mining, can help in this mission. Decision trees give distinctive views into the matter of recognizing malignant activity and may help to create technology-specific techniques to prevent attacks [1].

Computer security, as well as cyber or IT security, is guarding information systems from stealing or destructing to the software, hardware, or information on them. In addition, it protecting from perturbation or misusing of the IT services they deliver. The type of attacks, such as buffer overflow, teardrop and ping of death [2].

The Intrusion detection system (IDS) is defined as a system for discovering intrusions which making an attempt to misuse the data or computing resources of a computer system. It has two detection models include signature based IDS and anomaly based systems [3].

The target of this research is to decrease the high rate of false alarms, and also to achieve high detection rate. With the aim of achieving this goal, the next points are set:

- Use specific algorithm to reduce the high false alarms.
- Specify an efficient technique to distinguish the normal from abnormal attacks.

The reset of this paper is organized as follows: section 2 represents the previous related works, section 3 denotes the proposed approach, an explanation for the dataset, and the detection metrics that will be used, section 4 presents the experimental setup, section 5 shows the results and discussions, and finally section 6 is the conclusion.

2. Previous Related Works

In [4] a study applied decision tree to explain how the decision tree algorithms are used to reveal attacks in result of students. It aims to make models for decision tree (DT), in order to detect abnormal circumstances mechanically in student results tests. The results of this research achieved reasonable performance in the evaluation, namely in accuracy, sensitivity and specificity.

In [5] a study anticipated novel machine learning algorithms for declining false positives alarms in intrusion detection system (IDS). This work aims to state necessary entering attributes for structuring consort intrusion detection system (IDS), which is cost-effective. The results of this research show that the intended approach reduces the quantity and ratio of false positives, and poise detection rates (DR) for several types of network intrusions.

In [6] a study developed intrusion detection system (IDS) to detect attacks in computer networks. This work aims to analyze KDD99 test dataset by utilizing particular algorithms of machine learning like J48, random tree, Bayes net, random forest to identify these algorithms precision throughout categorizing attacks to wide-ranging classes. The results of this work explained that the random tree and random forest algorithms are the utmost effective in acting the classification on test dataset. As well, it depicted that CFS method for decreasing the detection time and increases the accuracy rate.

In [7] a study introduced the involvement of the four attributes categories in terms of evaluating detection rate (DR) and false alarm rates (FAR) metrics. This research studied NSL-KDD dataset in perspective of four attributes classes rather than behavior of specific attribute. The research showed that it can assist dataset appropriateness, in which greater detection rate (DR) is gained with a lowest false alarm rates (FAR). The shortcoming of this study it does not suitable for online intrusion detection.

In [8] a new study uses different classification algorithms to discover the anomalies in the network traffic patterns. This work aims to utilize and analyze the NSL-KDD dataset to study the capacity of various classification algorithms in realizing the abnormalities in patterns of network traffic.

In [9] study explained effective approach for detection and classification of lung cancer-related CT scan images into benign and malicious category. This work aims to apply image processing and machine learning approaches for detection and classification of lung cancer. The results of this work found that accuracy of MLP classifier is higher with value of 88.55% in comparing with the other classifiers.

In [10] study provided confidentiality to the user while utilizing cloud-based web services. This work aims to suggest DNA-based encryption/decryption technique to embed the confidentiality in the communication between client and the cloud service provider. The results of this work observed that DNA-based encryption/decryption technique gives the option to selectively encrypt only the confidential information rather that the whole information as in the case of SSL.

3. Proposed Approach based on Decision Tree Algorithm

The proposed approach based on Decision Tree algorithm follows the next pseudo code:

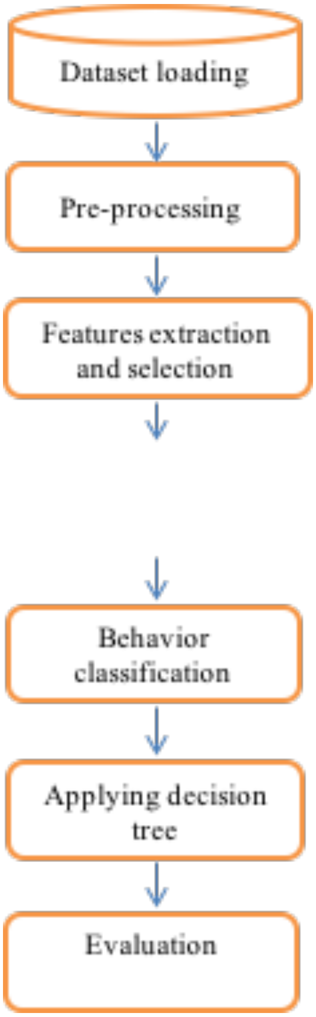
```
Step 1: Load the dataset into Weka.
Step 2: Apply pre-processing on the dataset.
Step 3: Divide the dataset into two labels called basic and traffic respectively.
Step 4: Calculate detection metrics (DR, FAR)
Step 5: Classify the behavior if it is normal or anomaly.
Step 6: Apply the decision tree algorithm.
Step 7: Evaluate detection metrics.
```

As well, it involves the following subsequent phases, as shown in Fig.1:

1. Dataset loading
2. Pre-processing

3. Features extraction and selection
4. Calculating detection metrics
5. Classification of behavior
6. Applying decision tree
7. Evaluation

Fig.1. The Proposed Approach based on Decision Tree Algorithm



The researchers use the decision tree algorithm for attack detection and reduce false alarms rate and to apply decision tree algorithm they following many steps include dataset loading in this phase we load the dataset in Weka workbench and the dataset will utilized in this paper is NSL-KDD dataset, pre-processing this phase is no need for pre-processing to be applied on the data used in this paper, in the feature extraction and selection phase they divide the both KDDTest+.arff and the KDDTrain+.arff into two labels called basic and traffic respectively. In calculating detection metrics phase, it calculates accuracy, Detection Rate (DR), False Alarm Rate (FAR), precision, and F-score. Indeed, these metrics are attained from the four principal components of a classification algorithm results that are presented in confusion matrix form, which demonstrates the actual instance classes versus predicted classes[11]. As presented in the table 1 below, these components are true negative (TN) component, false negative (FN) component, false positive (FP) component, and true positive (TP) component. In the classification of behavior phase, the researchers classify the behavior if it is normal or anomaly, and to perform this phase they apply the decision tree algorithms (J48, Random tree). In the reducing false alarms phase, the researchers apply the decision tree algorithm for reduce the false alarms rate and compare it with naïve Bayes to observe the effect of algorithms in reducing false alarms rate. The last phase is evaluating detection metrics, and in this phase the researchers evaluate the system depend on the rate of false alarms and detection for the decision tree.

To evaluate the gained results, the researchers need to calculate the accuracy, DR, FAR, precision and F-score. All these metrics are derived from the four basic result elements of any classification algorithm presented in the form of confusion matrix, which illustrates as in table 1.

Table 1. Confusion matrix for IDS

Confusion Matrix		Predicted Instances	
		Normal	Anomalous
Actual Instances	Normal	TN	FP
	Anomalous	FN	TP

2.1 NSL – KDD Dataset

NSL-KDD is a dataset offered to settle several of the ingrained matters of KDD99 dataset. While this version of KDD dataset experiences amount of the problems, mentioned by McHugh, that is not perfect illustrative of surviving actual networks due to the lack of universal datasets for network IDS, it widely used proficiently as benchmark dataset. Furthermore, the size of records of the NSL-KDD for train dataset and test dataset is suitable. Therefore, it is probable to apply tests for the whole dataset. The existing attack types of NSL-KDD dataset are: DOS, probing, U2R and R2L [12]. The NSL-KDD dataset holds the succeeding benefits compared to authentic KDD dataset:

- There are no repeated records in the train dataset, and hence; the classifier will not be prejudiced to many of repeated records.
- There is not any repetitive record in the planned test sets; so, learners' performance is not aligned with strategies that have higher detection rates in repeated records.
- The number of chosen records from every problem level collection is reversely related to the amount of records in the original KDD dataset.
- The numbers of records in the train dataset and test dataset are satisfactory, which keeps it reasonable to perform the experiences over the whole records, not to arbitrarily pick slight share.

NSL-KDD training dataset involves 4,900,000 single association vectors, each has 41 features, and classified as; either normal, or an attack. Table 2 shows the class of all 42 attributes of NSL-KDD dataset.

Table 2. Classes of KDD dataset attributes

No.	Label	Attribute Name	No.	Label	Attribute Name
1	B	duration	22	T	is_gust_login
2	B	protocol_type	23	T	Count
3	B	services	24	T	srv_count
4	B	flag	25	T	serror_rate
5	B	src_bytes	26	T	srv_serror_rate
6	B	dst_bytes	27	T	rerror_rate
7	B	land	28	T	srv_rerror_rate
8	B	wrong_fragmentt	29	T	same_srv_rate
9	B	urgent	30	T	diff_srv_rate
10	B	hot	31	T	srv_diff_host_rate

No.	Label	Attribute Name	No.	Label	Attribute Name
11	B	Num_faile_logins	32	T	dst_host_count
12	B	logged_in	33	T	dst_host_srv_count
13	B	num_compromised	34	T	dst_host_same_srv_rate
14	B	root_shell	35	T	dst_host_diff_srv_rate
15	B	su_attempted	36	T	dst_host_same_src_port_rate
16	B	num_root	37	T	dst_host_srv_diff_host_rate
17	B	num_file_creations	38	T	dst_host_serror_rate
18	B	num_shells	39	T	dst_host_srv_serror_rate
19	B	num_access_files	40	T	dst_host_rerror_rate
20	B	num_outbound_cmds	41	T	dst_host_srv_rerror_rate
21	B	is_hot_login	42	-	Class

2.2 Detection Metrics

This research utilized accuracy, detection rate (DR), precision, false alarms rate (FAR), F-measure, and recall as main detection metrics, which are based on calculating true positive (TP) rate value, false positive (FP) rate value, true negative (TN) rate value, and false negative (FN) rate value. True positive (TP) rate value denotes instance that is truly an attack, and classified correctly as an attack. FP represents those instances which are actually normal but classified as an attack. False positive (FP) rate value expresses instance that is in reality an attack, but classified wrongly as a normal. Where, true negative (TN) rate value symbolizes instance that is a normal, and classified as a normal, as well [13]. Accuracy represents how many instances were correctly classified, while precision represents out of all the instances classified as attacks, and how many were actually an attack. Recall represents how many attacks were correctly classified, percentage of attacks caught [14].

False alarm rate (FAR) is defined as the rate at which normal instances are wrongly classified as abnormally. Detection rate (DR) defines the proportion of accurately forecasted attacks to the whole number of real attacks. F-measure is defined as the coordinated mean of detection rate (DR) and the precision value.

4. Experimental Setup

This paper purposes to study and explain the function of 41 attributes of NSL-KDD dataset, and to utilize decision tree (DT) algorithm on detection rate (DR) and false alarm rate (FAR) for Intrusion Detection System (IDS) with relevance to two attributes class labels as in table 3.

Table 2. Classes of KDD dataset attributes

Attribute class	Abbreviation	Attributes
Basic	B	1 – 21
Traffic	T	22 – 41

The machine used in this research is HP laptop with 32 bit windows 7
operating system, corei3 and 4GB RAM. Furthermore, Weka 3.6.9 environment to implement the algorithms.

The NSL-KDD dataset was chosen for this experimental study whose attributes are tagged in two categories as present in table 3, where the data are divided into two parts (60%) for training and (40%) for testing. And this dataset we can find in (http://github.com/defcom17/NSL-KDD). This dataset holds a number of preparations, thus that its records is categorized into binary categories, such as normal/abnormal, or in one of five categories like normal, user to root (U2R) attack, denial of service (DOS) attack, probe attack, and , remote to local (R2L) attack. This study deals with the binary classification dataset detailed in Table 4.

Table 4. Instances of NSL-KDD dataset

	Normal Class Instances	Anomalous Class Instances	Total
KDDTrain+	67343	58630	125973
KDDTest+	9711	12833	22544

5. Results and Discussions

The results are presented as of confusion matrix of the three arrangements of datasets in Table 5 for naïve Bayes, Table 6 for random tree and Table 7 for J48.

Table 5. Result set for naïve Bayes algorithm

No	Attribute class combination	Naïve Bayes			
		TN	FN	FP	TP
1	BT	25362	3090	1626	20311
2	B	26675	9544	313	13857
3	T	24286	3267	2702	20134

Table 6. Result set for Random tree algorithm

No.	Attribute class combination	Random tree			
		TN	FN	FP	TP
1	BT	26931	67	57	23334
2	B	26766	119	222	23282
3	T	26735	299	253	23102

Table 7. Result set for j48 algorithm

No	Attribute class combination	J48			
		TN	FN	FP	TP
1	BT	26895	50	93	23351
2	B	26757	124	231	23277
3	T	26728	202	260	23199

The outline of results for DR is presented in Table 8 and for FAR is presented in Table 9. The essential measures applied in this paper are detection rate (DR) and false alarm rate (FAR). The classification results in detection rate (DR) and false alarm rate (FAR) form for the three cases of attribute classes mixtures are shown for Random tree, J48 and Naïve Bayes classifiers.

Table 8. Detection rate (DR) for random tree, j48 and naïve Bayes algorithms

No.	Attribute class combination	Detection Rate (%)		
		Random tree	Naïve Bayes	J48
1	BT	99.7	86.8	99.8

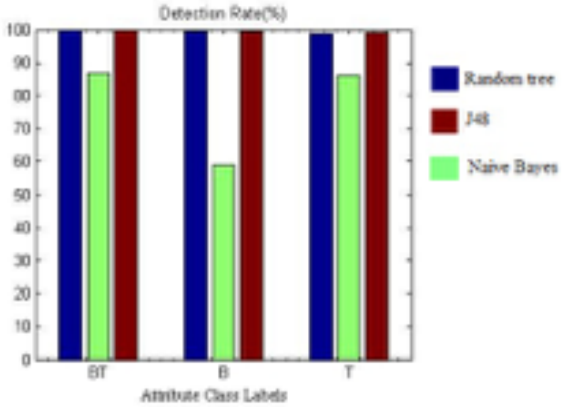
No.	Attribute class combination	Detection Rate (%)		
		Random tree	Naïve Bayes	J48
2	B	99.5	59.2	99.5
3	T	98.7	86	99.1

Table 9. False alarms rate (FAR) for random tree, j48 and naïve Bayes algorithms

No.	Attribute class combination	False Alarms Rate (%)		
		Random tree	Naïve Bayes	J48
1	BT	0.2	6	0.3
2	B	0.8	1.2	0.9
3	T	0 .9	10	1

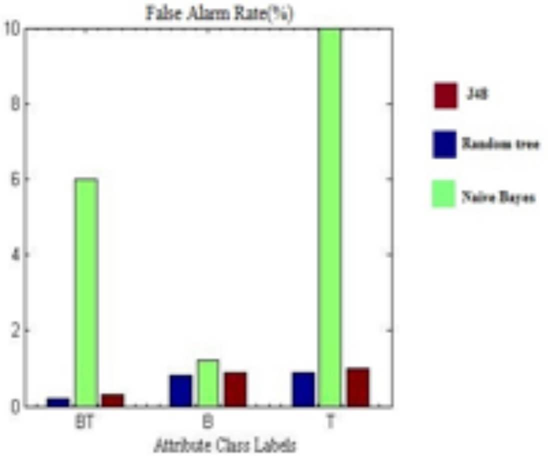
Fig. 2 and Fig. 3 present plots for the three classification algorithms with respect to one and two labeled attribute combinations respectively.

Fig. 2. Detection rate (DR) distributions for Random Tree, J48 and Naïve Bayes Algorithms



Taken into account the analysis of DR, Fig. 2 plot depict the DR for basic, traffic and BT attribute label combination. The random tree and J48 represent high detection rate for basic, traffic and BT labels but the naïve Bayes has low detection rate for three labels. Hence it can be concluded from Fig. 2 that the random tree and J48 have important contribution to achieving high detection rate (DR), whereas the naïve Bayes achieving low detection rate (DR).

Fig. 3. False alarms rate (FAR) for Random Tree, J48 and Naïve Bayes Algorithms



Taken into account the analysis of FAR, Fig. 3 plot depicts the FAR for basic, traffic and BT attribute labels combination. The random tree and J48 represents better FAR for basic, traffic and BT labels than Naïve Bayes. Hence it can be concluded from Fig. 3 that the random tree and J48 have significant contribution towards the reduction of the FAR whereas the naïve Bayes increase the FAR.

6.Conclusion

The fast advancement of data mining algorithms and strategies has led to machine learning shaping a distinct field of technology. It can be seen as a subclass of the artificial intelligence field, where the most important plans are the ability of a system to learn from its own activities.

Decision tree (DT) be attached to a class or category of supervised learning algorithms. In contrary to disparate supervised learning algorithms, decision tree (DT) algorithm is applied to resolving classification and regression problems, too. NSL-KDD dataset is one of the commonly exercised datasets for performance testing of Intrusion Detection System (IDS). In this paper, researchers used decision tree (DT) algorithm to achieve high detection rate (DR) and low false alarms rate (FAR). They applied decision tree algorithm and naïve Bayes on Weka. The decision tree achieved better FAR for basic, traffic and BT labels than Naïve Bayes. Also, the decision tree gained high detection rate for basic, traffic and BT labels but the naïve Bayes has low detection rate. Hence, they conclude with that the decision tree (DT) algorithm achieved high detection rate (DR) and low false alarms rate (FAR), which is better than other algorithms of machine learning.

In the future work researchers recommend applying other machine learning algorithm for five classes (normal, U2R, R2U, DOS and probe) to reduce the false alarms rate.

References

[1] Jeff Markey and Antonios Atlasis, “Using decision tree analysis for intrusion detection,” SANS Institute InfoSec Reading Room, 2011.

[2] Christina Mei-Fang Lee, An evaluation of machine learning techniques in intrusion detection., 2007.

[3] Ja Jabez and B Muthukumar, “Intrusion detection system (IDS): anomaly detection using outlier detection approach,” Procedia Computer Science, vol. 48, pp. 338--346, 2015.

[4] Hamza O Salami, Ruqayyah S Ibrahim, and Mohammed O Yahaya, “Detecting Anomalies in Students’ Results Using Decision Trees,” International Journal of Modern Education and Computer Science, vol. 8, no. 7, p. 31, 2016.

[5] Dewan Md Farid and Mohammad Zahidur Rahman, “Attribute weighting with adaptive NBTree for reducing false positives in intrusion detection,” arXiv preprint arXiv:1005.0919, 2010.

[6] Chibuzor John Ugochukwu and EO Bennett, “An Intrusion Detection System Using Machine Learning Algorithm,” International Journal of Computer Science and Mathematical Theory, vol. 4, no. 1, pp. 2545--5699, 2018.

[7] Preeti Aggarwal and Sudhir Kumar Sharma, “Analysis of KDD dataset attributes-class wise for intrusion detection,” Procedia Computer Science, vol. 57, pp. 842--851, 2015.

[8] L Dhanabal and SP Shantharajah, “A study on NSL-KDD dataset for intrusion detection system based on classification algorithms,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446--452, 2015.

[9] Gur Amrit Pal Singh and P. K. Gupta, "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans,", Neural Computing and Applications, 2018.

[10] Gunjan Gugnani S. P. Ghrera P. K. Gupta Reza Malekian B. T. J. Maharaj, "Implementing DNA Encryption Technique in Web Services to Embed Confidentiality in Cloud," Proceedings of the Second International Conference on Computer and Communication Technologies pp 407-415, 2015.

[11] Gupta, PK, Tyagi, Vipin, Singh, S.K., "Predictive Computing and Information Security," Springer Singapore, 2017.

[12] Rahul P. Tolankar, Vaibhav P. Sawalkar, and Niraj N. Kasliwal, "Review on IDS in Cloud Enviroment By Using FC - ANN," in National Conference on Innovative Trends in Science and Engineering, vol. 4, 2016, pp. 382 - 386.

[13] Wen Yu, Haibo He, and Nian Zhang, Advances in Neural Networks-ISNN 2009: 6th International Symposium on Neural Networks, ISNN 2009 Wuhan, China, May 26-29, 2009 Proceedings.: Springer, 2009, vol. 5552.

[14] Anubhavnidhi Abhashkumar and Roney Michael, "implementation an intrusion detection system using a decision tree".

Journal of Engineering and Applied Sciences, Vol. 7, Issue (2) November 2020

The Use of SNP Genotyping for QTL/ Candidate Gene Discovery in Plants

Salman F. Alamery
Biochemistry Department, College of Science, King Saud University,
Riyadh, P.O.Box 2455-11451, Saudi Arabia.

Abstract

Many plant species have been sequenced, which has enabled the identification of DNA sequence variations such as novel single-nucleotide polymorphisms (SNPs). SNP-based genotyping technologies have been developed over the last decade and play important roles in detecting SNPs in a cost and time effective manner, thereby facilitating plant genetic and genomic studies. The availability of genome and genetics resources and increase in sequencing data has led to SNPs discovery and array development using various approaches in many plant species. The development of SNP arrays for genotyping requires three steps: identification of a large number of SNPs, validation, and final selection. The discovery of SNPs and array development at various densities have been successfully developed and utilized in many diploid and polyploid plant species. SNP genotyping arrays have been useful for many applications such as cloning, association mapping, analyses of species diversity, evolution, genomic selection, and comparative genomics. The assaying of large segregating or mutant populations with an SNP array allows the accurate, efficient, and rapid determination of the genotypes of many individuals and genetic variations. Genotyping assays are an efficient way of generating a vast amount of genotypic data in crop species to produce highly informative SNPs for marker-assisted selection in breeding programs as well as functional genomic studies. This review highlights the SNP genotyping in the application of identification of SNPs in the discovery of quantitative trait locus (QTL) or candidate genes for important crop traits and dissection of other complex traits.

Keywords:
Single-Nucleotide Polymorphisms; Genotyping; Quantitative Trait Locus;
Candidate Genes; Crop Plants

1. Introduction

Genotyping is the process of determining the differences in DNA sequence between individuals. It involves the generation of allele-specific products and identification of the status of genotypes. Genotyping has become extremely important in association studies to determine genes or sequence variants linked to specific traits or diseases. It is also essential for gene mapping, analyses of species diversity and evolution, and marker-assisted selection [1]. Genotyping studies can be designed to identify DNA sequence differences at the single-nucleotide polymorphism (SNP) level. Hence, SNP genotyping is now widely applied in plant research such as candidate genes discovery, quantitative trait locus (QTL) analysis, linkage mapping and comparative genomics [2].

A single nucleotide polymorphism (SNP) is defined as single base-pair change at specific position in the genome and the most common genetic variations in plant genomes. SNP allelic variations in a population are often associated with a particular trait phenotype [3]. Therefore, SNPs have potential to be used as markers for genetic studies in functional genomics and breeding [4].

It has been reported that the frequency of SNPs in plant species is about 1 SNP in every 100–300 base pairs [5]. For instance, 37,000 SNPs have been identified in Arabidopsis Columbia (Col) and Landsberg erecta (Ler) [6]. This provides powerful tool for genotyping assays enabling these SNPs to be associated with of economically important traits [6].

The development of genomic and bioinformatic databases in plant species has dramatically increased the identification of SNP polymorphisms and scoring of the variation in specific targets. More importantly, a large number of potential SNPs and their surrounding sequences would provide a foundation for highly multiplexed automated genotyping analysis [7]. The availability of genomic resources permits the identification of SNPs and their association with particular phenotype in which it can be employed for marker-assisted selection and candidate gene approach [8, 9].

2.SNP array development and selection

SNP arrays called “SNP chips” are a type of DNA microarray designed with a large number of SNPs (up to one million). This approach involves high-throughput genome scan providing a time and cost-efficient tool for genotyping individuals at millions of different positions across the genome [10]. These arrays have become commonly used in plant genetic studies such as gene mapping, genome-wide association, genomic selection, etc. [11].

The completion of human genome sequencing and genotyping arrays were initially developed in human before other organisms. The SNP Genotyping has been a powerful tool for comprehensive genome-wide association studies in human [1]. It also has significant impact on the genetic analysis of human disease and cancer. This facilitated understanding complex diseases and further characterized the human genome [2]. In addition, SNP genotyping was also played a role in animal genetic and breading studies such as in genome selection and identification of disease resistance genes [3]. However, the genotyping array is still limited to certain plant species and challenging due to the genome nature and size. SNP identification and calling is more difficult and complex. Even though the level of diversity and functionality in plant species compared to animal and human, there is still ongoing works on adopting this technology for functional genomics studies [4].

Various approaches have been used to

identify SNP genotyping arrays in many plant species depending on the availability of genome sequences and genomic/ genetic resources. The primary method is the direct sequencing of DNA fragments. This is the most reliable approach and has been proven to successfully detect SNPs in many organisms [12]. It is appropriate when sequence information is limited or when polymorphisms of interest in a specific genotype or candidate gene are being investigated [13]. Direct sequencing can be used as a validation tool to differentiate true polymorphisms from sequencing errors. This requires the sequencing of enough individuals to achieve sufficient the accuracy and coverage to distinguish real SNPs from sequencing artifacts [7, 14]. However, this is generally costly and time-consuming and is impractical for large-scale genetic studies.

The mining of expressed sequence tags (ESTs) have also been used for identifying SNPs. ESTs are short sequences of cDNA. The number of identified ESTs range from tens of thousands for species that have been little investigated to over a million in well studied plant species [15]. For instance, in Arabidopsis, 1.5 million ESTs have been identified, whereas in Brassica species; namely B. napus, B. rapa and B. oleracea 643,944, 213,605 and 179,213 have been identified, respectively. The EST sequences provide sufficient redundancy for screening of the presence of polymorphisms using bioinformatic analysis tools, allowing subsequent identification of SNPs polymorphism [16]. This approach mainly provides limited SNPs located only in transcribed regions (coding and UTR regions), resulting in a small number of SNPs, but it offers a low-cost source of informative and abundant SNPs and high quality sequence data.

RNA sequencing (RNA-seq) is another cost effective approach for SNPs discovery within coding regions of the genome[17]. It can be used to identify SNPs in the transcriptome and help to reduce genome complexity. Geraldes, Pang [18] were able to identify over 0.5 million putative SNPs in 26,595 genes in Populus trichocarpa. RNA-seq has been found to provide more accurate functional annotation due to the enrichment of expressed genes, but, it has not been widely used due to its cost and laboriousness [19].

The most comprehensive approach for discovering SNPs is next-generation sequencing (NGS). It provides cheap and reliable large-scale SNPs identification. SNPs discovery by NGS is not limited to protein-coding sequences leading to the identification of SNPs widely distributed across the genome and enabling determination of the correlation between phenotypes and SNPs in non-coding regions. This has enabled the discovery of thousands of SNPs in closely related species [20]. NGS-derived SNPs have been reported in many plant species [21-23]. SNPs identification by NGS requires a complete genome sequence as a reference, although it can be achieved without a full reference sequence using de novo read assembly mapping [23,24] . The genome sequences of a number of plant species with relatively small genome are available as suitable reference genomes [25]. Genotyping-by-sequencing (GBS) has newly emerged as approach for sequence-based genotyping. The strength of this approach is that sequence polymorphism [mainly SNPs] and genotyping are completed at the same time. This approach enables the search for SNPs and presence/ absence variations in diverse species with and without reference genomes [26, 27].

This approach has been demonstrated to be robust across a range of species and producing high number of molecular markers [28]. However, it is limited in terms of the number of individuals for whom NGS data are available, requires bioinformatic analysis, and is technically challenging [4].

With the advancement of sequencing technologies and SNP discovery approaches, computational tools and databases for SNP markers have been developed such as AutoSNPdb [29], TreeSNPs [30], dbSNP [31], and NABIC SNP [32]. Such databases provide more detail about SNP sequences, names, gene definitions, locations, functions, and associations among others.

The development of SNP arrays requires three steps: identification of a large number of SNPs, validation, and final selection. The discovery of SNPs aims to identify as many SNPs as possible. This requires background of bioinformatics analyses, including sequence or reads mapping or alignment, reads trimming, SNPs calling, and SNPs filtering [11, 33]. SNP arrays can also include SNPs developed from specific gene sequences “candidate genes of interest” selected from databases. Various SNP analyses can yield different numbers of variants among species [34]. Therefore, SNP validation is essential to select true SNPs polymorphisms and to decrease the false positive SNPs.

Given that sequencing data remains prone to inaccuracies at rate as high as one error every 100 base pairs [35]. SNPs selections can be significantly improved by increase the stringency of SNPs detection and meeting the requirement of accurate SNP calling [36]. A number of important filtering criteria should be considered including the following:

(1) Information of the SNP flanking sequences is a key element. This refers to sequence length on either side of the SNP, for instance, 50 bp for Illumina. There should be no SNPs present within 50 bp on either side of the SNP.

(2) Sequence depth over the entire mapping assembly. A minimum sequence /read depth per individual/genotype represents less sequencing error at SNPs.

(3) The allele depth defined as a number of different sequences in which an allele appeared for both variant and reference alleles. SNPs were selected where the minor allele was present in more than one genotype. SNPs present in less than two genotypes are excluded.

(4) The general considerations for array SNP selection include the presence of repetitive or palindromic sequences, GC content, SNP depth, SNP types and SNP frequency.

(5) Location and distribution of SNPs

throughout the genome and their genetic effects.

(6) Minimum quality score measuring the probability that a base is called incorrectly.

(7) Minimum coverage of reads, supporting the presence of the allele in a given SNP.

(8) Maximum coverage of reads, which should be less than the average read depth of all SNPs.

(9) Average copy number of the SNP flanking region, which should be less than two.

(10) Sequences with more than 4 SNPs per 100 bp should be removed to avoid selecting any SNPs located in hypervariable regions.

Furthermore, during selection, SNPs can be classified according to their physical position on the chromosomes and best Blast hits. According to their relative location, SNPs are classified into different categories because they are part of coding sequences, within exons, introns, between exons or within promoter sequences. This would help to integrate the information with existing genetic mapping for further genetic studies.

The majority of SNPs are most likely to be false positive, leading to SNPs being undetermined or incorrectly genotyped [37]. filtering criteria for SNP selection and reduction of genomic complexity would increase the SNP genotyping efficiency by achieving a higher proportion of correctly genotyped SNPs relative to incorrectly genotyped ones [37].

Approaches for SNP identification involving the discovery and validation of predicted SNP polymorphisms have been optimized and developed for simple genomes in diploid plant species [13, 38]. However, SNP identification in polyploid crops with complex genomes has remained very challenging. The higher levels of genome complexity and presence of polymorphisms between subgenomes in polyploid crops constitute major additional challenges for SNP prediction [2].

Polyploidy is extremely common in the plant kingdom, particularly, in important crops such as potato (Solanum tuberosum L.), alfalfa (Medicago sativa L.), durum wheat (Triticum durum Desf.), cotton (Gossypium hirsutum L.) and canola (Brassica napus). The major concern with plant species is the complexity of their genomes, due to highly repetitive sequence, homologous genome and intragenomic duplication. These factors influence the efficiency of SNPs identification and impede the accurate discrimination of candidate SNPs between homologous and paralogous sequences. In polyploid species, it is very common to find polymorphic SNP within a single genotype due to the presence of either homoeologous loci from individual subgenomes or paralogous loci from duplicated regions of the genome [12]. In Brassica napus, most polymorphisms between subgenomes are homoeologous making it difficult to assign them to the A or C subgenome [Kaur et al., 2012]. Hence, stringent mapping parameters are often essential to avoid false SNP calls [11, 39].

Polyploid species are characterized by a large genome size and complex structure, consisting of subgenomes of two homologous genomes from progenitor diploid species. They also contain high level of repetitive DNA sequences. For example, Brassica napus is an allotetraploid that has undergone extensive duplication events over the course of evolution. High levels of intragenomic duplication lead to significant intragenomic sequence paralogy and a complex gene family. Owing to this genome duplication, a gene could have a second copy at another position in the genome [40]. Therefore, duplication events in Brassica influence the alignment comparison process of NGS sequencing assembly and SNPs calling. Thus, there is a need for rigorous discrimination between the origin of identified SNPs [2]. Maize is an ancient polyploid species with large regions of genomic duplication. It has been shown that a considerable number of SNP markers have a pattern of shifting of the clusters to one side or five clusters. This is indicative of detection of more than one locus that correspond to the duplicated regions identified in the maize genome [14].

The discovery of SNPs and array development at various densities have been successfully achieved and applied in many diploid and polyploid plant species [34]. Examples of these include wheat 20K, 820K and 660K SNP arrays [41-43], Maize 600K SNP array [44], Brassica napus 60K and 6K SNP arrays [45, 46], rice 700K SNP array [47], Apple 8K and 480K SNP array [48,49] , peanut 58K SNP arrays [50], strawberry 90K SNP array [51], sugarcane 345K, 76K and 84K SNP arrays [52-54], cotton 63K SNP Array[55] and oat 6K SNP array [56]. However, there are additional obstacles to SNP identification and utilization for genotyping in polyploid species, so this field of study has progressed slowly due to the complex nature of the genomes and polyploid inheritance [11, 57].

Against this background, to reduce the genome complexity, additional genetic information about progenitor relationships with allopolyploid specie is extremely useful to assist in discriminating SNPs polymorphism through the ability to compare polyploid-derived sequence genotypes to diploid counterparts in which strict alignment and assembly criteria should be applied [2]. SNP arrays have significantly accelerated the molecular studies in plant genomics. As a large number of SNP array platforms have been developed, there are still demands to address their limitations. The efficiency and accuracy of genotyping rate is still a challenging. SNP selection and validation requires careful consideration to increase the success rate and to produce high proportions of correctly genotyped SNPs. Another issue is the number of informative polymorphic SNP in a population. Only a small number of SNPs can be suitable for further analysis as the majority of SNPs cannot be detected, genotyped or being monomorphic. This is due to significant numbers of false positive SNPs. Moreover, the level of sequence diversity and structural genome variations including translocations and copy number variation may reduce both the efficiency of SNP discovery and the ability of correctly identifying the allelic state of each individual [58, 59]. The main drawback of SNP arrays is that they still remain expensive. There is a huge cost associated with design and validation of the array. The array technique requires specific equipment and high labour cost, Thus resulting in the limit use by most researchers. Commercially available SNP arrays with variable number of fixed SNPs have a limited use and cannot be modified. They are often customized to be specific to certain species or populations [4].

With recent advancements of sequencing technologies, SNPs prediction and validation by sequencing could replace the genotyping arrays. The genotyping arrays are species-specific, expensive to design. It only covers all known SNP variants and not intended to discover rare or novel variants. In contrast, sequencing generates more coverage and the entire genome is scanned to find and genotype new SNP variants. The sequencing approach is less expensive, more power and computational process is becoming more routine. In addition, sequencing is ideal for uninvestigated species with limited genomic resources. Sequencing is powerful tool for genetic diversity and genomic selection with speed and efficiency. In the future, sequencing could become more widespread and ideal option for SNP discovery for more genetic and genomics applications [5, 6]

3. SNP genotyping and candidate gene association

SNP genotyping arrays are a powerful approach to detect SNPs in a cost- and time-effective manner and more importantly facilitate genetic and genomic studies such as QTL mapping and association studies. The discovery of a vast number of SNPs provides an ideal makers for the novel SNPs array to study the genetic mapping of QTL associated with candidate genes [26, 34, 60].

The candidate genes for particular traits in plants have been the particular focus in order to facilitate the reverse genetic approach and gene discovery or the identification of gene polymorphism that could be useful for marker-assisted selection [55,61, 62]. Thus, SNP discovery and genotyping in segregating populations with specific trait generally enable us to find a strong correlation between many more SNPs and genes controlling the trait in a population [2]. This approach can be successful if there is a known, strong functional SNP and phenotypic data and quantitative loci.

Cloning and characterization of candidate genes within QTL in plant genetics and breeding have been difficult and not feasible through genetic mapping and map based cloning due to large genetic and physical distances of QTL, low resolution of available genetic maps and complexity of the genome [63]. Conventionally, QTL is determined by linkage mapping approaches where two parents and their segregating population are screened for polymorphisms associated with the trait of interest. The linkage of a particular region to a given phenotype can be determined by the frequency of recombinants exhibiting phenotypic variations for a trait of interest [4]. In fact, SNP genotyping solves the problem of positional map cloning of a large number of candidate gene association studies in the case that many candidate genes are weak candidates with a low level of polymorphism. This is because candidate gene studies only examine one or a few SNPs which carry limited polymorphism information on the overall population variation within that QTL. All above, if a certain region corresponds to a QTL, the gene may be a candidate gene. This approach could greatly facilitate the map cloning approach[64].

Recently, the mapping of QTL of interest has been made possible by taking advantage of SNP genotyping technologies combined with NGS data. Candidate gene-based approaches have been used for identifying SNPs for specific candidate genes which might provide connection between allele function and specific trait variations. In addition, scanning may be performed to identify regions associated with a particular phenotype, while gene-wide scanning for the SNP distribution and association to a segregating population would significantly enhance large-scale efforts to identify all loci controlling multiple traits for crop improvement [7].

SNP genotyping plays an essential role in identifying a number of candidate genes / QTL linked to important crop traits and dissecting other complex traits [65]. For example, a candidate gene and QTL cluster associated with four fiber traits on chromosome A07 of cotton was identified [66]. In wheat, a candidate gene based SNP marker was developed to determine the locus that controls resistance to leaf rust, stripe rust, and powdery mildew diseases [67]. In addition, a study by Wu, Zhao [68] found 39 candidate genes linked to three loci DSRC4, DSRC6, and DSRC8 associated with resistance to sclerotinia stem rot in B. napus. Moreover, putative SNP markers were mapped to the marker flanking region linked to Fhb, a gene involved in resistance to fusarium head blight (FHB) disease [69]. This would enable fine mapping towards cloning of the Fhb1 gene.

Many studies have been conducted to identify SNP markers linked to a QTL region to produce high resolution QTL and narrow down the target locus for the analysis of candidate genes. For example, a high-density SNP map gave a better resolution RFLP/SSR-based QTL by condensing two QTL regions for grain weight in rice within 123 kb. Moreover, a few SNPs were functionally associated with the variation in yield [70]. In soybean, SNP genotyping narrowed down the QTL region for aphid resistance gene, Rag1 from 12 cM to 115 kb with two linked SNP markers and additional candidate genes being identified [71]. In another study, SNP markers of candidate genes for flowering time in Brassica were identified. [72]. Also, Three QTLs for flowering time were linked to three SNPs in the promoter of the BrFLC2 gene causing low expression. These markers facilitated comparative mapping between B. rapa and Arabidopsis.

SNP genotyping arrays with a large number of SNPs are useful for the analysis of many traits or individuals at high genetic resolution, leading to candidate gene mapping. The idea is that SNP markers at candidate genes selected for their association with a phenotype are physically mapped and the genotypic data are compared across contrasting plant genotypes. Hence, a cluster of SNPs situated in a small physical distance within QTL region are most likely the potential candidate genes. This strategy has a good advantage to the problems associated with map-based cloning and physical mapping [73, 74]. Taking advantage of SNP genotyping technologies combined with NGS data, candidate gene-based approaches have been used for the discovery of SNPs linked to candidate genes/QTL which might provide a direct correlation of allele function with specific trait variation [7].

4.Use of Illumina Infinium genotyping assay as an example

High throughput SNP genotyping platforms have been developed to assay up to 1 million SNPs which played a decisive role in the success of genomic studies. Many SNP genotyping platforms are currently available commercially such as Illumina GoldenGate and Infinium, TaqMan or GeneChip from Affymetrix. These technologies can be applied to genotype SNP markers and require a preliminary step of SNP discovery, technical expertise and an expensive laboratory set-up.

The Illumina Infinium assay is a whole-genome SNP genotyping assay that has been proven to be successful and efficient for many plant crops [23, 45, 75]. It is capable of multiplexing from about 6,000 up to 1 million SNPs. It enables genome-wide analysis offering an ideal method to identify candidate polymorphic SNPs for QTL mapping. The Infinium assay develops BeadChips with customized SNP content. Each SNP locus is assayed and analyzed independently for each sample yielding high intensity estimate and accurate genotype calls.

In this paper, we briefly describe the use of 6K B. napus custom SNP array with 5306 SNPs as implemented in the Infinium assay [Illumina Inc., San Diego, USA]. Details of the array and the Infinium assay are described previously [Dalton-Morgan et al., 2014]. These SNPs were distributed over the 19 chromosomes of B. napus genome. This array was applied to genotype B. napus cultivars and map populations in order to study the SNP genotypes and identify the candidate SNPs located on QTL or chromosome regions associated with candidate genes. Samples were analyzed with the Infinium II assay protocol according to the manufacturer's instructions. Genotyping module in GenomeStudio Illumina with the default parameters was utilized for SNP data analysis clustering and genotype call. The clusters were reviewed and manually edited if needed. SNPs that were difficult to score were either manually adjusted or eliminated from the analysis completely.

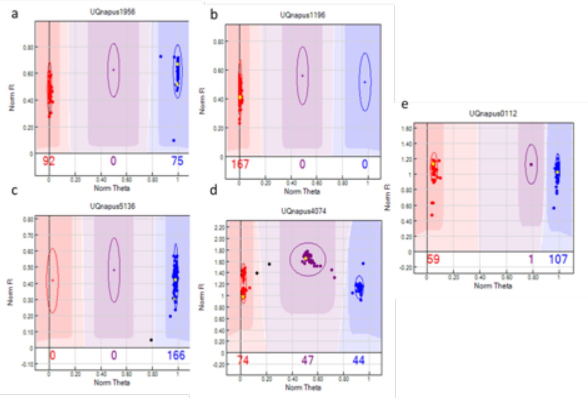
The SNP genotype results show consistent and clear clustering for most of the 5306 SNPs across the mapping population with 91% of SNPs successfully genotyped and clustered, taking into account both monomorphic and polymorphic SNPs

(Fig 1 and Table 1). Successful SNP markers primarily produced distinct clustering patterns with three possible genotypes (AA, AB, BB), detecting polymorphic or monomorphic loci. A good-quality SNP genotyping cluster shows three clearly defined and tight clusters with the homozygotes and heterozygotes. The cluster definition is relatively simple in diploid species. However, it is more complex in crop plants that are polyploid such as B. napus because SNP clusters can often detect more than three overlapping clusters. This is sometimes seen in polyploid species as five clusters (AAAA, AAAB, AABB, ABBB, and BBBB) instead of the typical three clusters (AA, AB and BB). Therefore, SNP clusters in polyploid cases often make the distinction of the allelic status more challenging since the clusters need to be defined more carefully and mostly in a manual fashion.

Table 1. Classification of SNP marker results from B. napus Infinium SNP genotyping assay on the population in this study

SNP position	No. of total SNPs	Proportion of SNPs		
		Failed	Monomorphic	Polymorphic
Genome-wide	5306	487 (9%)	3535 (67%)	1283 (24%)
Chromosome A7	285	32	178	74
Scaffold 3 on A7	109	12	44	57

Fig 1. Examples of successfully clustered SNPs based on the GenomeStudio software. Clusters display the area where the three different genotypes with homozygous allele A (red), heterozygous AB (purple) and homozygous allele B (blue) are called. A) Two clusters with AA and BB genotypes and a monomorphic state for the parents (both in yellow). B) One cluster with AA genotype and a monomorphic state for both parents and all individuals. C) One cluster with BB genotype and a monomorphic state for both parents and all individuals. D) Three clusters represent the genotypes AA, AB and BB, and a polymorphic state for parents. E) Two clusters with AA and BB genotypes and a polymorphic state for both parents. The black dot outside of the cluster represents failed samples and thus is scored as "no call".



Two different indexes were initially used to evaluate the quality of the raw data before the optimization of SNP clustering. The call rate index showed that more than 90% of SNPs were successfully genotyped for the mapping population. In addition, the GenTrain score index gives an indication of the quality of the SNP clustering. According to Illumina, for an SNP to be retained, a minimum GenTrain score of 0.15 is advisable. In the present study, an SNP had to get a minimum GenTrain score of 0.15 and had to be segregating in the related mapping population to be declared successful. In the DH mapping population, the majority of clustering SNPs had a GenTrain score of more than 0.4, indicating that SNP genotyping was highly reliable with a low rate of missing data. Low GenTrain score may indicate ambiguous cluster separation.

Of the 5306 SNPs assayed, 1284 (24%) were polymorphic between the two parental lines and thus could be used for clustering (Table 1). This provides a large number of novel markers that will be used in genetic mapping. Another 3535(67%) SNPs from this assay were monomorphic between the two parental lines, meaning that they clustered together, having the same base at that specific locus. These SNPs are uninformative in genetic studies. Only 487(9%) of the 5306 SNPs failed to give a genotype or could not be easily clustered and were not investigated further in this study. This probably indicates false SNPs, resulting from possible sequencing mistakes.

The Infinium assay is sensitive to the number of allelic copies [target locus + duplicated or paralogous alleles] being assayed (Fig 2 and 3). This can be a problem, particularly in polyploids such as B. napus, where two homozygous clusters or more than three clusters are most likely due to the existence of non-unique genomic regions. This is because B. napus is a highly duplicated genome. It is anticipated that the amphidiploid nature of the B. napus genome will usually result in homeologous pairs of genes, originating from the A and C genomes. This sometimes results in polymorphisms detected at a given position, one corresponding to the A genome and the other to the C genome. The Infinium assay cannot distinguish the status of SNP alleles from homologues (76). This makes automated SNP genotyping more challenging in a complex genome. The absence of cluster separation can be due to a non-allele-specific match of the primers, for example (Fig 3).

Fig 2. Vertically separated clusters generally polymorphic for a different locus than the source of the targeted SNP. This may indicate the presence of a third polymorphic allele.

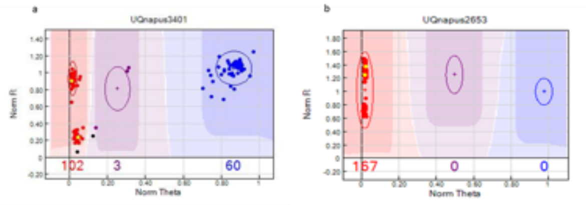
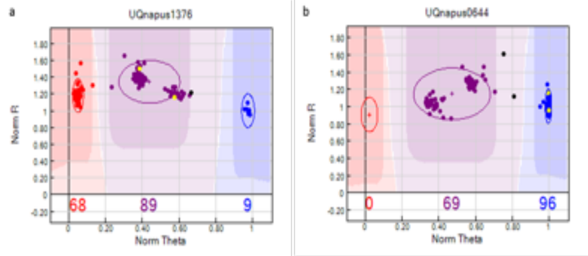


Fig 3. An example of more than two genotype clusters observed due to the polyploid nature of the B. napus genome



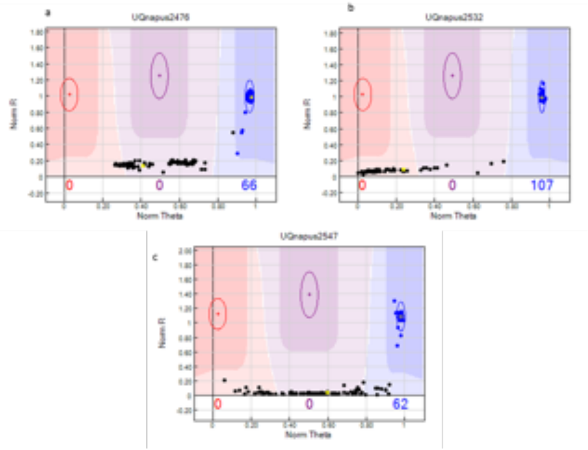
The majority of SNPs on chromosome A7 (252 out of 285) gave a clear genotype (Table 1). Of these, 74 were successfully polymorphic. Fifty seven of these SNPs were on scaffold 3, scaffold with the largest length in chromosome A7, representing the QTL region underlying a disease resistance candidate gene [77]. The SNPs located the within genes in scaffold 3 were generally more polymorphic than the overall set of SNPs in chromosome A7 suggesting their potential as candidate genes.

The results also showed that the 57 SNPs markers were distorted, with a 2:1 segregation ratio, and the direction of distortion was mainly toward the maternal parent. In addition, the number of genotypes corresponding to the resistant parent was almost twice that corresponding to the susceptible parent, with averages of 102 and 58, respectively. This suggests that segregation of the resistance might be controlled by multiple genes. This finding is consistent with a previous study [78], which found a distorted segregation ratio for phenotype. This distortion might be the result of chromosomal abnormalities biasing the estimation of the recombination fraction between markers on the linkage group. This may subsequently lead to the loss of informative data in QTL mapping, if these markers are not handled carefully.

As the population is double haploid (DH), most of the SNPs were expected to show two clear main clusters, representing the two homozygous genotypes. A small additional cluster in the middle of the graph corresponding to heterozygous genotypes may represent a third allele or an additional null allele. Doubled haploids are genetically homozygous lines. However, DH samples sometimes behave as heterozygotes. This might be because the SNP was initially identified as polymorphic in sequence comparisons but behaved monomorphically in the analyzed sample or was genotyped as heterozygous reflecting structural chromosomal rearrangements: duplication, translocation or transpositions. Taken the findings together, our analyses of the DH mapping population showed distinguishable segregating allelic variants at a single locus underlying the QTL.

Three SNPs were null and showed the presence of an insertion or deletion surrounding the SNP locus (Fig 4), presumably due to a mutation or the presence of a third allele. Further sequencing and validation of these SNPs might give an indication of whether the SNP allele status has an association with a candidate gene.

Fig 4. Individuals with homozygous deletion cluster at the bottom of the graph; their genotype calls are missing/null. This may indicate the presence of chromosomal deletion or a third allele. The BB cluster remains intact.



5. Conclusion

This work demonstrates how NGS technologies associated with SNP genotyping could be a feasible strategy for the genotyping of thousands of SNPs and correlating them with QTL or candidate genes. It also provides an alternative approach for identification of candidate gene. The results can be further analyzed for genetic mapping and association in the B. napus segregating population. Hence, the genotyping quality and physical position data obtained for the SNPs can be useful for consensus genetic maps, positional cloning, or association mapping. Despite genotyping assay are being efficient to generate a vast amount of genotypic data in polyploid crops, analysis of SNP calls is somewhat challenging in polyploids due to the multiallele combinations in the genotypes. However, high density SNP arrays would be efficient to produce highly informative SNPs for marker-assisted selection in breeding programs as well as functional genomic studies.

References

[1] Syvanen, A.-C., 2005. Toward genome-wide SNP genotyping, Nature Genetics 378,

[2] Kaur, S., Francki, M. G., and Forster, J. W., 2012. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species, Plant Biotechnology Journal. 10, no. 2, pp. 125-138,

[3] Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., and Lander, E. S., 1998. Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome, Science. 280, no. 5366, pp. 1077-1082

[4] Deschamps, S., Llaca, V., and May, G. D., 2012. Genotyping-by-Sequencing in Plants, Biology. 1, no. 3, pp. 460-483,

[5] Gupta, P. K., Roy, J. K., and Prasad, M., 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants, Current Science. 80, no. 4, pp. 524-535

[6] Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L., 2002. Arabidopsis Map-Based Cloning in the Post-Genome Era, Plant Physiology. 129, no. 2, pp. 440-450

[7] Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics, Current Opinion in Plant Biology. 5, pp. 94 - 100,

[8] Westermeier, P., Wenzel, G., and Mohler, V., 2009. Development and evaluation of single-nucleotide polymorphism markers in allotetraploid rapeseed (Brassica napus L.), Theoretical and Applied Genetics. 119, no. 7, pp. 1301-1311,

[9] Edwards, D., and Batley, J., 2010. Plant genome sequencing: applications for crop improvement, Plant Biotechnology Journal. 8, no. 1, pp. 2-9,

[10] von Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T. E., Alves, P. C., Lyons, L. A., Mattucci, F., Randi, E., Cragnolini, M., Galián, J., Hegyeli, Z., Kitchener, A.

C., Lambinet, C., Lucas, J. M., Mölich, T., Ramos, L., Schockert, V., and Cocchiararo, B., 2020. Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels, Molecular Ecology Resources. 20, no. 3,

[11] You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J., 2018. Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array, Frontiers in Plant Science. 9, pp. 104-104,

[12] Ganal, M. W., Altmann, T., and Röder, M. S., 2009. SNP identification in crop plants, Current Opinion in Plant Biology. 12, no. 2, pp. 211-217,

[13] Edwards, D., Forster, J., Cogan, N. I., Batley, J., and Chagné, D., “Single Nucleotide Polymorphism Discovery,” Association Mapping in Plants, N. Oraguzie, E. A. Rikkerink, S. Gardiner et al., eds., pp. 53-76: Springer New York, 2007.

[14] Durstewitz, G., Polley, A., Plieske, J., Luerssen, H., Graner, E. M., Wieseke, R., and Ganal, M. W., 2010. SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species Brassica napus, Genome. 53, no. 11, pp. 948-956

[15] Schmid, K. J., Sörensen, T. R., Stracke, R., Törjék, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B., 2003. Large-Scale Identification and Analysis of Genome-Wide Single-Nucleotide Polymorphisms for Mapping in Arabidopsis thaliana, Genome Research. 13, no. 6a, pp. 1250-1257

[16] Batley, J., Barker, G., O’Sullivan, H., Edwards, K. J., and Edwards, D., 2003. Mining for Single Nucleotide Polymorphisms and Insertions/Deletions in Maize Expressed Sequence Tag Data, Plant Physiology. 132, no. 1, pp. 84-91,

[17] Barbazuk, W. B., Emrich, S. J., Chen, H. D., Li, L., and Schnable, P. S., 2007. SNP discovery via 454 transcriptome sequencing, The Plant journal : for cell and molecular biology. 51, no. 5, pp. 910-918,

[18] Geraldes, A., Pang, J., Thiessen, N., Cezard, T., Moore, R., Zhao, Y., Tam, A., Wang, S., Friedmann, M., Birol, I., Jones, S. J. M., Cronk, Q. C. B., and Douglas, C. J., 2011. SNP discovery in black cottonwood (Populus trichocarpa) by population transcriptome resequencing, Molecular Ecology Resources. 11 Suppl 1, pp. 81-92

[19] Rogier, O., Chateigner, A., Amanzougarene, S., Lesage-Descauses, M.-C., Balzergue, S., Brunaud, V., Caius, J., Soubigou-Taconnat, L., Jorge, V., and Segura, V., 2018. Accuracy of RNAseq based SNP discovery and genotyping in Populusnigra, BMC Genomics. 19, no. 1, pp. 909

[20] Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A., 2009. Next-gen-

eration sequencing technologies and their implications for crop genetics and breeding, Trends in biotechnology. 27, no. 9, pp. 522-530,

[21] Yamamoto, T., Nagasaki, H., Yonemaru, J.-i., Ebana, K., Nakajima, M., Shibaya, T., and Yano, M., 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms, BMC Genomics. 11, no. 1, pp. 267,

[22] Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D., 2008. Sequencing of natural strains of Arabidopsis thaliana with short reads, Genome Research. 18, no. 12, pp. 2024-2033

[23] Kumar, S., Banks, T. W., and Cloutier, S., 2012. SNP Discovery through Next-Generation Sequencing and Its Applications, International Journal of Plant Genomics. 2012, pp. 15,

[24] You, F., Huo, N., Deal, K., Gu, Y., Luo, M.-C., McGuire, P., Dvorak, J., and Anderson, O., 2011. Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence, BMC Genomics. 12, no. 1, pp. 59,

[25] Feuillet, C., Leach, J. E., Rogers, J., Schnable, P. S., and Eversole, K., 2011. Crop genome sequencing: lessons and rationales, Trends in Plant Science. 16, no. 2, pp. 77-88,

[26] Hirakawa, H., Shirasawa, K., Ohyama, A., Fukuoka, H., Aoki, K., Rothan, C., Sato, S., Isobe, S., and Tabata, S., 2013. Genome-Wide SNP Genotyping to Infer the Effects on Gene Functions in Tomato, DNA Research,

[27] Poland, J. A., and Rife, T. W., 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics, Plant Gen. 5, no. 3, pp. 92-102, 2012/11,

[28] He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., and Li, Z., 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, Frontiers in Plant Science. 5, pp. 484-484,

[29] Duran, C., Appleby, N., Clark, T., Wood, D., Imelfort, M., Batley, J., and Edwards, D., 2009. AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants, Nucleic Acids Research. 37, no. Database issue, pp. D951-D953,

[30] Clément, S., Fillon, J., Bousquet, J., and Beaulieu, J., 2010. TreeSNPs: a laboratory information management system (LIMS) dedicated to SNP discovery in trees, Tree Genetics & Genomes. 6, no. 3, pp. 435-438

[31] Smigielski, E. M., Sirotkin, K., Ward, M., and Sherry, S. T., 2000. dbSNP: a database of single nucleotide polymorphisms, Nucleic Acids Research. 28, no. 1, pp. 352-355,

[32] Kim, C.-K., Won, S.-Y., Seol, Y.-J., and Lee, T.-H., 2015. NABIC SNP: an integrated database for SNP markers, Bioinformation. 11, no. 7, pp. 369-372,

[33] Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C. M., Pandey, M. K., Vaughn, J., Abernathy, B., Barkley, N. A., Hovav, R., Burow, M., Nayak, S. N., Chitikineni, A., Isleib, T. G., Holbrook, C. C., Jackson, S. A., Varshney, R. K., and Ozias-Akins, P., 2017. Genome-wide SNP Genotyping Resolves Signatures of Selection and Tetrasomic Recombination in Peanut, Molecular plant. 10, no. 2, pp. 309-322,

[34] Montanari, S., Bianco, L., Allen, B. J., Martínez-García, P. J., Bassil, N. V., Postman, J., Knäbel, M., Kitson, B., Deng, C. H., Chagné, D., Crepeau, M. W., Langley, C. H., Evans, K., Dhingra, A., Troggio, M., and Neale, D. B., 2019. Development of a highly efficient Axiom™ 70 K SNP array for Pyrus and evaluation for high-density mapping and germplasm characterization, BMC Genomics. 20, no. 1, pp. 331-331,

[35] Appleby, N., Edwards, D., and Batley, J., "New Technologies for Ultra-High Throughput Genotyping in Plants," Plant Genomics, Methods in Molecular Biology™ J. P. Gustafson, P. Langridge and D. J. Somers, eds., pp. 19-39: Humana Press, 2009.

[36] Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., Micheletti, D., Rosyara, U. R., Cattonaro, F., Vendramin, E., Main, D., Aramini, V., Blas, A. L., Mockler, T. C., Bryant, D. W., Wilhelm, L., Troggio, M., Sosinski, B., Aranzana, M. J., Arús, P., Iezzoni, A., Morgante, M., and Peace, C., 2012. Development and Evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm, PLoS ONE. 7, no. 4, pp. e35668,

[37] Trebbi, D., Maccaferri, M., Heer, P., Sørensen, A., Giuliani, S., Salvi, S., Sanguineti, M., Massi, A., Vossen, E., and Tuberosa, R., 2011. High-throughput SNP discovery and genotyping in durum wheat (Triticum durum Desf.), Theoretical and Applied Genetics. 123, no. 4, pp. 555-569

[38] Chagné, D., Batley, J., Edwards, D., and Forster, J., "Single Nucleotide Polymorphism Genotyping in Plants," Association Mapping in Plants, N. Oraguzie, E. A. Rikkerink, S. Gardiner et al., eds., pp. 77-94: Springer New York, 2007.

[39] Byers, R., Harker, D., Yourstone, S., Maughan, P., and Udall, J., 2012. Development and mapping of SNP assays in allotetraploid cotton, Theoretical and Applied Genetics. 124, no. 7, pp. 1201-1214

[40] Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I., 2005. Chromosome triplication found across the tribe Brassiceae, Genome Research. 15, no. 4, pp. 516-525

[41] Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., and Chen, F., 2020. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat, Plant Biotechnology Journal. 18, no. 6, pp. 1354-1360,

[42] Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., King, J., West, C., Griffiths, S., King, I., Bentley, A. R., and Edwards, K. J., 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool, Plant Biotechnology Journal. 14, no. 5, pp. 1195-1206,

[43] Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L., 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Ap proach, PLoS ONE. 7, no. 2, pp. e32253,

[44] Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T. M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K. F. X., and Schön, C.-C., 2014. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array, BMC Genomics. 15, no. 1, pp. 823

[45] Mason, A. S., Higgins, E. E., Snowdon, R. J., Batley, J., Stein, A., Werner, C., and Parkin, I. A. P., 2017. A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array, Theoretical and Applied Genetics. 130, no. 4, pp. 621-633

[46] Raman, H., Dalton-Morgan, J., Diffey, S., Raman, R., Alamery, S., Edwards, D., and Batley, J., 2014. SNP markers-based map construction and genome-wide linkage analysis in Brassica napus, Plant Biotechnology Journal. 12, no. 7, pp. 851-860

[47] McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., Singh, N., DeClerck, G., Agosto-Perez, F., Korniliev, P., Greenberg, A. J., Naredo, M. E. B., Mercado, S. M. Q., Harrington, S. E., Shi, Y., Branchini, D. A., Kuser-Falcão, P. R., Leung, H., Ebana, K., Yano, M., Eizenga, G., McClung, A., and Mezey, J., 2016. Open access resources for genome-wide association mapping in rice, Nature Communications. 7, no. 1, pp. 10532

[48] Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R. P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J. D., Iezzoni, A., Mockler, T., Wilhelm, L., Van de Weg, E., Gardiner, S. E., Bassil, N., and Peace, C., 2012. Ge-

nome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple, PLoS ONE. 7, no. 2, pp. e31745,

[49] Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., Poncet, C., Micheletti, D., Kerschbamer, E., Di Pierro, E. A., Larger, S., Pindo, M., Van de Weg, E., Davassi, A., Laurens, F., Velasco, R., Durel, C.-E., and Troggio, M., 2016. Development and validation of the Axiom®Apple480K SNP genotyping array, The Plant Journal. 86, no. 1, pp. 62-74

[50] Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., Chitikineni, A., Chavarro, C., Chen, X., Upadhyaya, H. D., Vishwakarma, M. K., Leal-Bertioli, S., Liang, X., Bertioli, D. J., Guo, B., Jackson, S. A., Ozias-Akins, P., and Varshney, R. K., 2017. Development and Evaluation of a High Density Genotyping ‘Axiom_Arachis’ Array with 58 K SNPs for Accelerating Genetics and Breeding in Groundnut, Scientific Reports. 7, no. 1, pp. 40577

[51] Bassil, N. V., Davis, T. M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., Mahoney, L., Wood, D., Alperin, E. S., Rosyara, U. R., Koehorst-vanc Putten, H., Monfort, A., Sargent, D. J., Amaya, I., Denoyes, B., Bianco, L., van Dijk, T., Pirani, A., Iezzoni, A., Main, D., Peace, C., Yang, Y., Whitaker, V., Verma, S., Bellon, L., Brew, F., Herrera, R., and van de Weg, E., 2015. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria× ananassa, BMC Genomics. 16, no. 1, pp. 155

[52] Aitken, K. S., Farmer, A., Berkman, P. J., Muller, C., Wei, X., Demano, E., Jackson, P., Magwire, M., Dietrich, B., and Kota, R., “Generation of a 345K sugarcane SNP chip.” pp. 1165-1172.

[53] Yang, X., Song, J., You, Q., Paudel, D. R., Zhang, J., and Wang, J., 2017. Mining sequence variations in representative polyploid sugarcane germplasm accessions, BMC Genomics. 18, no. 1, pp. 594

[54] Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Hoffmann, H. P., de Souza, A. P., Garcia, A. A. F., and Carneiro, M. S., 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane, BMC Genomics. 18, no. 1, pp. 72-72,

[55] Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., Frelichowski, J., Giband, M., Hague, S., Hinze, L. L., Kochan, K. J., Riggs, P. K., Scheffler, J. A., Udall, J. A., Ulloa, M., Wang, S. S., Zhu, Q.-H., Bag, S. K., Bhardwaj, A., Burke, J. J., Byers, R. L., Claverie, M., Gore, M. A., Harker, D. B., Islam, M. S., Jenkins, J. N., Jones, D. C., Lacape, J.-M., Llewellyn, D. J., Percy, R. G., Pepper, A. E., Poland, J. A., Mohan Rai,

K., Sawant, S. V., Singh, S. K., Spriggs, A., Taylor, J. M., Wang, F., Yourstone, S. M., Zheng, X., Lawley, C. T., Ganal, M. W., Van Deynze, A., Wilson, I. W., and Stelly, D. M., 2015. Development of a 63K SNP Array for Cotton and High-Density Mapping of Intraspecific and Interspecific Populations of Gossypium spp, G3: Genes|Genomes|Genetics. 5, no. 6, pp. 1187,

[56] Tinker, N. A., Chao, S., Lazo, G. R., Oliver, R. E., Huang, Y.-F., Poland, J. A., Jellen, E. N., Maughan, P. J., Kilian, A., and Jackson, E. W., 2014. A SNP Genotyping Array for Hexaploid Oat, The Plant Genome. 7, no. 3, pp. plantgenome2014.03.0010,

[57] Clevenger, J., Chavarro, C., Pearl, Stephanie A., Ozias-Akins, P., and Jackson, Scott A., 2015. Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations, Molecular plant. 8, no. 6, pp. 831-846

[58] Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., and Wen, C., 2020. A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties, Scientific Reports. 10, no. 1, pp. 5623

[59] Vanderzande, S., Zheng, P., Cai, L., Barac, G., Gasic, K., Main, D., Iezzoni, A., and Peace, C., 2020. The cherry 6+9K SNP array: a cost-effective improvement to the cherry 6K SNP array for genetic studies, Scientific Reports. 10, no. 1, pp. 7613

[60] Chong, X., Zhang, F., Wu, Y., Yang, X., Zhao, N., Wang, H., Guan, Z., Fang, W., and Chen, F., 2016. A SNP-Enabled Assessment of Genetic Diversity, Evolutionary Relationships and the Identification of Candidate Genes in Chrysanthemum, Genome biology and evolution. 8, no. 12, pp. 3661-3671,

[61] Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing, C., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A. R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.-C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K. J., Hayden, M., and Akhunov, E., 2014. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array, Plant Biotechnology Journal. 12, no. 6, pp. 787-796,

[62] Dalton-Morgan, J., Hayward, A., Alamery, S., Tollenaere, R., Mason, A. S., Campbell, E., Patel, D., Lorenc, M. T., Yi, B., Long, Y., Meng, J., Raman, R., Raman, H., Lawley, C., Edwards, D., and Batley, J., 2014. A high-throughput SNP array in the amphidiploid species Brassica napus shows diversity in resistance genes, Functional & Integrative Genomics. 14, no. 4,

pp. 643-655, 2014/12/01,

[63] Mauricio, R., 2001. Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology, Nature Reviews Genetics. 2, no. 5, pp. 370-381,

[64] Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., Tang, X., Zhou, F., Deng, X. W., and Zhang, Q., 2014. A High-Density SNP Genotyping Array for Rice Biology and Molecular Breeding, Molecular plant. 7, no. 3, pp. 541-553

[65] Majeed, S., Rana, I. A., Atif, R. M., Ali, Z., Hinze, L., and Azhar, M. T., 2019. Role of SNPs in determining QTLs for major traits in cotton, Journal of Cotton Research. 2, no. 1, pp. 5

[66] Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., Deng, D. D., Hinchliffe, D. J., Jones, D. C., and Fang, D. D., 2016. A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton, BMC Genomics. 17, no. 1, pp. 903-903,

[67] Lagudah, E., Krattinger, S., Herrera-Foessel, S., Singh, R., Huerta-Espino, J., Spielmeyer, W., Brown-Guedira, G., Selter, L., and Keller, B., 2009. Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens, Theoretical and Applied Genetics. 119, no. 5, pp. 889-898

[68] Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., Zhang, C., Fan, C., Wang, Y., and Zhou, Y., 2016. Genome-wide Association Study Identifies New Loci for Resistance to Sclerotinia Stem Rot in Brassica napus, Frontiers in Plant Science. 7, pp. 1418-1418,

[69] Bernardo, A., Ma, H., Zhang, D., and Bai, G., 2012. Single nucleotide polymorphism in wheat chromosome region harboring Fhb1 for Fusarium head blight resistance, Molecular Breeding. 29, no. 2, pp. 477-488

[70] Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J., and Zhang, Q., 2011. Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers, PLoS ONE. 6, no. 3, pp. e17595,

[71] Kim, K.-S., Hill, C., Hartman, G., Hyten, D., Hudson, M., and Diers, B., 2010. Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538, Theoretical and Applied Genetics. 121, no. 3, pp. 599-610

[72] Li, F., Kitashiba, H., Inaba, K., and Nishio, T., 2009. A Brassica rapa Linkage Map of EST-based SNP Markers for Identification of Candidate Genes Controlling Flowering Time and Leaf Morphological Traits, DNA Research. 16, no. 6, pp. 311-

323,

[73] McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., and Hirschhorn, J. N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges, Nature Reviews Genetics. 9, no. 5, pp. 356-369

[74] Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S., and Van Tassell, C. P., 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle, PLoS ONE. 4, no. 4, pp. e5350,

[75] Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., and He, Z., 2017. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives, Molecular plant. 10, no. 8, pp. 1047-1064

[76] Hu, Z., Huang, S., Sun, M., Wang, H., and Hua, W., 2012. Development and application of single nucleotide polymorphism markers in the polyploid Brassica napus by 454 sequencing of expressed sequence tags, Plant Breeding. 131, no. 2, pp. 293-299,

[77] Tollenaere, R., Hayward, A., Dalton-Morgan, J., Campbell, E., Lee, J. R. M., Lorenc, M. T., Manoli, S., Stiller, J., Raman, R., Raman, H., Edwards, D., and Batley, J., 2012. Identification and characterization of candidate Rlm4 blackleg resistance genes in Brassica napus using next-generation sequencing, Plant Biotechnology Journal. 10, no. 6, pp. 709-715,

[78] Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., Rehman, A., Lindbeck, K., Luckett, D., Wratten, N., Batley, J., Edwards, D., Wang, X., and Raman, H., 2012. Molecular mapping of qualitative and quantitative loci for resistance to Leptosphaeria maculans causing blackleg disease in canola (Brassica napus L.), Theoretical and Applied Genetics. 125, no. 2, pp. 405-418

Journal of Engineering and Applied Sciences, Vol. 7, Issue (2) November 2020

Eye Fixation Operational Definition: Effect on Fixation Duration when Using I-DT

Amin G Alhashim*
School of Industrial and Systems Engineering, University of Oklahoma,
Norman, OK, USA, alhashim@ou.edu
Abdulrahman Khamaj
Industrial Engineering Department, Jazan University, Jazan, Saudi Arabia,
abdulrahman@jazanu.edu.sa

Abstract

Many eye movement metrics such as the eye fixation duration metric depend, directly or indirectly, on eye fixations. However, eye fixations require an operational definition in order to be correctly classified by the eye movement event detection algorithms. This paper examines the effect of 24 different operational definitions of fixation on the results generated by Identification based on Dispersion Threshold (I-DT), the most popular eye movement event detection algorithm; and hence, the fixation duration metric. The 24 operational definitions are a combination of four different values (0.5º, 1.0º, 1.3º, and 2.0º) of the dispersion threshold and six different value (80ms, 100ms, 150ms, 200ms, 300ms, and 400ms) of the duration threshold of the I-DT algorithm. The preliminary results show that there is no statistically significant difference between the various operational definitions of fixation on the fixation duration metric.

Keywords:
Eye Tracking, Eye Movement Classification Algorithms, Eye Movement Event Detection Algorithms, I-DT, Eye Fixation, Eye Fixation Operational Defintion, Fixation Duration

1. Introduction

Eye tracking technology has been around since 1879 [10] and it went through different phases where it shifted from being intrusive and inconvenient to becoming unobtrusive and barely noticeable. Poole & Ball [35] generally define eye tracking as “a technique whereby an individual’s eye movements are measured, so that the researcher knows both where a person is looking at any given time and the sequence in which the person’s eyes are shifting from one location to another.” Eye fixation occurs when a person directs their visual gaze towards a particular location [two-dimensional coordinate points (horizontal and vertical)] on a display. However, an agreed-upon operational definition of fixation is still lacking ([33], [15], [23]). This operational definition of fixation is the essence of the eye movement event detection algorithms where it is being used to discriminate between two main events in the raw eye tracking data: fixations and saccades. The lack of such a de facto standard for the operational definition of fixation and the fact that such a definition greatly impacts the higher-level analysis, makes the measures resulting from the eye movement event detection algorithms very difficult to compare ([11], [23]).

This paper represents a preliminary study conducted to investigate the extent to which variants of the operational definition of fixation can impact the fixation duration metric, a widely used metric in many fields such as scene perception [8], educational research [24], and human behavioral research [12].

The authors in [5] & [17] mentioned a number of studies that utilized the eye fixation duration as a metric in their work. Different variants of the operational definition of fixation will be used on the best and most used eye movement event detection algorithm: Identification based on Dispersion Threshold (I-DT). I-DT is proposed by [23] which was adopted from [29].

Figure 1 shows the pseudo-code of the algorithm adopted from [23]. The algorithm starts by initializing a window based on the first raw fixation point. The window is expanded by adding more raw fixation points until the distance (i.e. the dispersion) between the farthest two points in the window is greater than the (maximum) dispersion threshold. When the newly added raw fixation points causes the dispersion of the window to go beyond the dispersion threshold, a fixation point will be created and centered based on average of all the points in the window. The process will repeat till no more raw fixation points exit.

Fig. 1. The pseudo-code of the I-DT algorithm proposed by [23]

```
INPUT: dispersion threshold, duration threshold

WHILE there are still points
  initialize window over first points to cover the duration threshold

  IF dispersion of window points <= threshold
    add additional points to the window until dispersion > threshold
      note a fixation at the centroid of the window points
    remove window points from points
  ELSE
    remove first point from points
  END IF
END WHILE

RETURN fixations
```

The fixation duration metric can be used to reveal various aspects of the task under study. Eye fixation duration refers to the amount of time a person fixates their eyes on a particular object in an area of interest [34]. Several authors believe that fixation duration determines the extent to which the person’s cognitive processing is easy or difficult [37] & [33]. Specifically, the authors in [17] reported a number of plausible interpretation for the longer fixation duration when debugging a computer program such as the difficulty to understand, complexity, importance, and notability. In addition, the authors in [31] stated that “longer fixations are a sign of increased difficulty in extracting and processing information due to higher information density, ambiguity, or complexity.” Similarly, the author in [36] claimed that long eye fixation duration indicates that a person is struggling and/or confused when cognitively processing an element on a display. For example, in a reading task, a longer fixation duration could mean that the text under study is conceptually difficult [21]. Similarly, in a visual search task, a longer fixation duration means, in general, that the objects presenting on the scene are not consistent [5] & [7].

Given such interpretations for the fixation duration and the fact that this metric is built on top of the fixations being generated by the eye movement event detection algorithms from the raw eye tracking data, the operational definition of fixation could play a significant role in these interpretations. This means if different operational definitions of fixation lead to significantly different fixation durations, the interpretations of such fixation durations will be different. This will lead to different conclusions and invalid results especially if the difference between the interpretation of success and failure is subtle [8] & [29].

The remaining of this paper proceeds as follows: Section 2 presents the details of the method used to investigate the research question. Section 3 details the process used to analyze the collected data from cleaning to perpetration to event detection and metric calculation. Section 4 shows the results obtained after analyzing the data and Section 5 discusses the implication of the obtained results and outlines some of the future directions.

2.Method

2.1 Participant

A 35-year old male graduate student from the University of Oklahoma took part in this study. The participant had a normal uncorrected vision.

2.2 Apparatus & software system

Tobii Pro TX300 screen-based eye tracker [25], ancestor of Tobii Pro Spectrum [27], has been used to track and collect the raw gaze data of the participant at a sampling rate of 120 Hz. Each tuple of the collected raw gaze data consists of seven main parts: timestamp, eye position, relative eye position, 3D gaze point, 2D gaze point, validity code, and pupil diameter. The timestamp holds one place in the tuple while each eye holds 13 places: three for eye position, three for relative eye position, three for 3D gaze point, two for 2D gaze point, one for pupil diameter, and one for validity code. Therefore, each packet received from the eye tracker contains a total of 27 pieces of information (see Figure 2). More information about the meaning of each part of the raw eye tracking data packet can be retrieved from Tobii Analytics SDK Developer's Guide [26].

Fig. 2. The structure of the data provided by Tobii Pro TX300 via the Tobii Analytics SDK.

Time Stamp	Left Eye													Right Eye												
	3D Eye Position			Relative 3 D Eye Position			2D Gaze Point		3D Gaze Point			Pupil Diameter	Validity Code	3D Eye Position			Relative 3 D Eye Position			2D Gaze Point		3D Gaze Point			Pupil Diameter	Validity Code
	X	Y	Z	X	Y	Z	X	Y	X	Y	Z			X	Y	Z	X	Y	Z	X	Y	X	Y	Z		

The stimulus was displayed on the monitor that comes attached to eye tracker. The monitor is 23” TFT with a resolution of 1920 x 1080 pixel and an aspect ratio of 16:9.

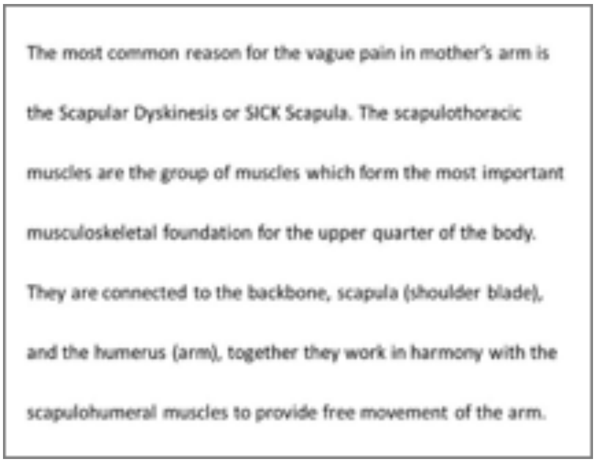
The raw eye tracking data packets recorded by the eye tracker has been read using a custom MATLAB script that utilizes Tobii Analytics SDK [27]

The software has been implemented to collect and store the raw eye tracking data into a CSV file for each session of the experiment.

2.3 Stimulus

A medical text, 69 words, excerpted and modified from [1] was used as a stimulus (Figure 3). The used text was purposely chosen as it contained several long medical terms. We assumed that those medical terms would have a low level of recognition and hence the participant would fixate longer on them [11], [20] & [21].

Fig. 3. The stimuli used in the experiment. An excerpt of a medical text with a few long medical terms. The lines of the excerpt have been spaced apart to allow for a more accurate discrimination of eye fixations.



The most common reason for the vague pain in mother’s arm is

the Scapular Dyskinesis or SICK Scapula. The scapulothoracic

muscles are the group of muscles which form the most important

musculoskeletal foundation for the upper quarter of the body.

They are connected to the backbone, scapula (shoulder blade),

and the humerus (arm), together they work in harmony with the

scapulohumeral muscles to provide free movement of the arm.

The stimuli text, 7 lines, was prepared in a way that the lines were spaced apart to allow for more accurate discrimination of fixations. An accurate discrimination of fixations is needed to account for any precision issues of the eye tracker and allow for associating the fixations with the words that appear in the stimuli more confidently. A single space was used to separate the words in each line because the center of the word, more precisely the position before the center of the word, is considered the optimal viewing position [16] according to [18]. Given this case, a single space between the words of each line would be sufficient to accurately discriminate fixations that belong to different words.

A sans-serif font, Calibri, was used for the stimuli text. According to [14], a word displayed in a sans-serif font type is recognized faster that the same word displayed in a serif font type. In our experiment, the use of sans-serif font type would be expected to reduce the fixation duration the participant spends on the non-medical terms (i.e., frequently used words).

2.4 Procedure

Upon the arrival of the participant to the experiment’s site, a formal consent form was signed and short description about the experiment was given. To begin the experiment (reading task), the participant was seated at an average viewing distance of 68.8 cm (min=68.6 cm, max=71.7 cm) from the monitor. Then, a simple calibration of eye movements was performed. The MATLAB script available with Tobii

Analytics SDK [27] was adopted to do the calibration process that required the participant to fixate on five points in succession. If the calibration process was deemed satisfactory, which was determined based on the calibration visualization (see Figure 4), the reading task would be initiated by the experimenter. Otherwise, the calibration process would be repeated.

No eye tracking data were captured during the calibration process.

After the calibration process, the participant was asked to read the provided text carefully. The participants had been instructed to press the Escape button once he was done reading. The participant's raw eye tracking data were collected after the disappearance of the 5-second timer shown on the whole screen and blurring the stimuli (i.e., the medical text). Once the participant completed the given task and pressed the Escape button, the recording of the raw eye tracking data stopped accordingly. The raw eye tracking produced by the eye tracker device was saved into a CSV file at the end of the session to be processed afterward.

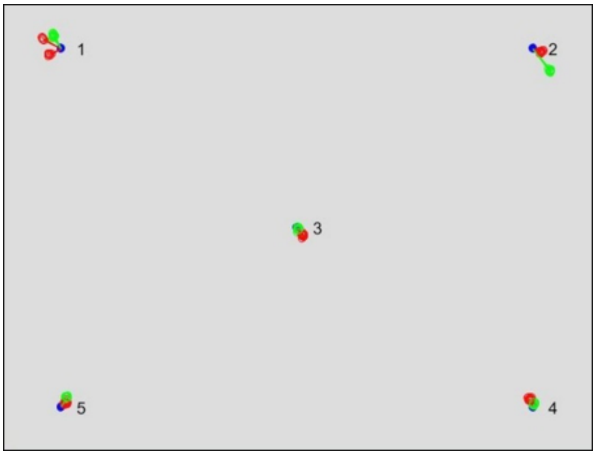


Fig. 4. Visualization of the result of calibration process for the participant. The blue dots represent the calibration points where the participant needs to look at while doing the calibration. The green circles represent the left eye offset from calibration point while the red circles represent the right eye offset.

After finishing the reading task, the participant was asked to fill a brief survey about his English skills. The survey was intended to be used later when running the experiment with a larger sample and to account for any discrepancies that may arise from the differences in English proficiency levels of the participants and hopefully helps in explaining them.

3. Data Analysis

3.1 Data cleaning

Each packet of the collected raw eye tracking data contained a validity code for each eye. The validity code associated with the captured data about a particular eye represented how confident the eye tracker was assigning this particular data to that particular eye. A validity code of zero for both eyes represents the highest confidence level in detecting both eyes while a validity code of four for both eyes means that eye tracker was not able to detect any eye. Any mix of validity codes for the two eyes is either invalid or means that the eye tracker is not confident to which eye the captured data belong. For the complete list of different combinations of validity code, see [26].

Any raw eye tracking data packet with a validity code other than zero for any eye will not be included in the computation process. This is because the average of both eyes will be used when applying the eye movement event detection algorithm. In addition, any eye tracking data packet outside the stimuli boundary box will be excluded as well.

3.2 Data preparation

For the I-DT algorithm to work, the raw eye tracking data need to be provided. The four main components that need to be provided are: the x- and y-coordinates of the gaze on the monitor, the distance of the participant's eye from the monitor, and a reference point in time of when that raw gaze data had been collected. The gaze x- and y-coordinates are the average of the x- and y-coordinates of the 2D gaze point part of the left and right eyes reported in the raw eye tracking data package collected by the eye tracker. Similarly, the distance of the participant's eye from the monitor is the average z-coordinate of the left and right eyes reported in the 3D eye position part (refer to Apparatus & Software System section above for more information). The z-coordinate in the 3D eye position part is reported in cm while the x- and y-coordinates in the 2D gaze point part is reported in what is the Tobii call it 'Active Display Coordinate System (ADCS)'. In this system, the point (0,0) denotes the upper left corner and the point (1,1) denotes the lower right corner of the active display area which is the monitor in our case. All the x- and y-coordinates are converted from the ADCS to the monitor pixel-based coordinates.

A timestamp starting from zero is calculated for each raw eye tracking data packet based on the timestamp provided by the eye tracker. The processed version of the raw eye tracking data that are feed into the I-DT algorithm consists of four columns: timestamp, the gaze x- and y-coordinates on the monitor, and the gaze z-coordinate from the monitor (see Figure 5).

Fig. 5. Structure of data prepared from the raw gaze data provided by the Tobii Pro TX300 via Tobii Analytics SDK. The 'Time Stamp' column represents the zero-based time of when the corresponding raw gaze data packet had been collected. The 'X' and 'Y' columns represents the x- and y-coordinates of the gaze on the monitor. The 'Z' column represents how far (the z-coordinate) the eye was from the monitor.

Time Stamp	X	Y	Z

3.3 Event detection

The operational definition of fixation consists of two thresholds: duration and dispersion. The duration refers to the minimum time and the dispersion refers to the maximum area. Hence, the eyes must stay stationary within a particular area for a minimum time in order for a group of raw eye tracking data to be detected as a fixation by the I-DT algorithm. After detecting a fixation, the I-DT algorithm will continue to include all the following raw eye tracking data until the dispersion threshold is violated. When the violation happens, the duration of the fixation and the average x and y relative positions of all the raw eye gaze data packets that were part of the fixation will be recorded. The process will continue until all the raw eye tracking data are processed. Figure 1 shows the pseudo-code of the I-DT algorithm.

Table 1 presents the common values reported in the literature for the dispersion and duration thresholds as an operational definition for the fixation. This paper will investigate the effect of several combination of values for dispersion and duration thresholds, namely, 0.5º, 1.0º, 1.3º, and 2.0º for the dispersion threshold and 80ms, 100ms, 150ms, 200ms, 300ms, and 400ms for the duration threshold. Hence, a total of 24 sets of dispersion and duration thresholds combinations will be tested.

Table 1. Common values for the dispersion (in degrees) and duration (in milliseconds) thresholds reported in the literature as an operational definition for the fixation.

Source	(Maximum) Dispersion Threshold	(Minimum) Duration Threshold
(Buurman, Roersema, & Gerrissen, 1981)	1.3º	100ms
(Salthouse, Ellis, Diener, & Somberg, 1981)	2.0º	200ms
(Moffitt, 1980)	2.0º	200ms
(Widdel & Kaster, 1981)	2.0º	200ms
(Salvucci & Goldberg, 2000)	0.5º-1.0º	100-200ms
(Jacob & Karn, 2003)	2.0º	100-200ms
(Blascheck et al., 2017)	-	200-300ms
(Nyström & Holmqvist, 2010)	0.5º	80-150ms
(Blignaut, 2009)	0.5º-1.0º	100-400ms

3.4 Fixation duration metric calculation

The array of fixations obtained from the I-DT algorithm for each set of the dispersion and duration thresholds were used to calculate the fixation duration spent on each word in the stimulus. The coordinates of bounding box for each word in the stimulus (Figure 6) as well as the obtained fixations were provided into a custom MATLAB script to calculate the fixation durations as well as depict the fixations on top of the stimulus for each set of the combination of the dispersion and duration thresholds.

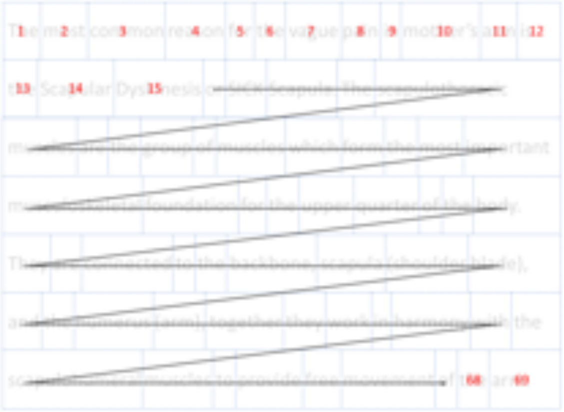


Fig. 6. The stimulus with a bounding box around each word. The bounding box is for illustration only. The coordinate of each bounding box, the x- and y-coordinates of the upper left and lower right corners, is what is provided to the MATLAB script.

4. Results

The participant took 42.6 seconds to finish the reading task and a total of 5019 raw eye tracking data had been collected by the eye tracker. Cleaning the raw eye tracking data resulted in excluding 52 (1.04%) packets of the data because the validity code reported by the eye tracker device is not zero for at least one of the eyes. All the raw eye tracking data were inside the stimuli bounding box and hence no data other than those that have invalid code were excluded from the analysis. Table 2 presents a fragment of the cleaned raw eye tracking data that are ready to be passed to the event detection algorithm. Figure 7 shows a visualization of the same data depicted over the stimulus.

Table 2. The raw eye tracking data after cleaning. The 'TS' column represents the zero-based time stamp for each raw eye tracking data packet. The 'X' and 'Y' columns represent the x- and y-coordinates of the gaze on the monitor in pixels. The 'Z' column represents how far the participants' eyes were from the monitor, the z-coordinate, in cm.

TS	X	Y	Z
0	964	623	686
25	968	691	686
33	967	691	686
42	968	692	686
50	971	689	686
…	…	…	…
20864	527	515	689
20872	533	522	689
20880	526	523	689
20889	528	509	689
20897	526	499	689
…	…	…	…
42519	1430	848	689
42527	1435	843	689
42536	1441	841	689
42544	1437	849	689
42552	1412	849	689

Fig. 7. The cleaned raw eye tracking data depicted over the stimuli. The small blue circles represent raw eye tracking data while the red lines represent the order in which the raw eye tracking data had been collected.

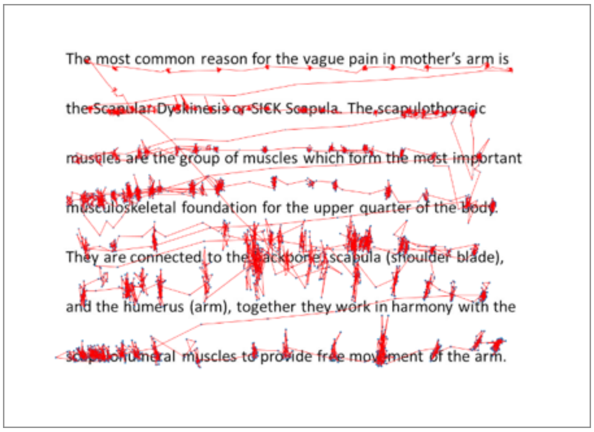


Figure 8 shows the eye fixations obtained from the I-DT algorithm from the raw eye tracking data for each of the 24 different combinations of the dispersion and duration thresholds. The eye fixations are depicted over the stimulus. The circles sizes are proportional to the fixation duration.

5. Discussion & Future Work

The reader may question the validity of the study given that only one participant was recruited. However, the main purpose of this study at this stage is not to deploy inferential statistic, i.e., to check whether there is a statistically significant differences between the different operation definitions of fixation. The goal at this stage is to check, given everything else (the participant, the task, etc.) the same, how the I-DT behave under different operational definitions of the fixation.

By looking at the patterns of the fixations in Figure 8 and the exact fixation durations in Table 3 under the different operational definitions for the fixation, we can clearly notice that the dispersion threshold is the most determinant factor of the fixation duration metric. Regardless of the duration threshold, the I-DT algorithm seems to result in almost exact fixation durations when fixing the dispersion threshold. For example, regardless of the duration threshold, ‘Dyskinesis’ (word #15 in line #2) was determined to be fixated on for a period of 2,217ms when the dispersion threshold was set to 0.5º; 4,374ms in the case of 1.0º dispersion threshold; 5,424ms in the case of 1.3º dispersion threshold; and zero ms in the case of 2.0º dispersion threshold. The rationale behind such behavior of the I-DT algorithm is the nature of the dispersion threshold. The dispersion threshold works as a perimeter within with the I-DT algorithm works. Hence, when the perimeter stays the same, the number of raw eye movements in close proximity that falls withing this parameter will most probably stay the same.

The other eye tracking metrics that can be taken into consideration are overall number of fixations on the whole stimulus; number of fixations on each part of the stimulus; fixation pattern, i.e., the order in which the participant scan the stimulus; time to first fixation, i.e., how long the participant spend till fixating on any part of the stimulus; number of fixated on parts of the stimulus; and others. Although some of these metrics such as the fixation pattern and the number of fixated on words can be answered here, we chose otherwise in order to keep the paper short.

Beside investigate the effect of the different settings of the two thresholds of the I-DT algorithm on other eye tracking metrics, it is worth in the future to investigate the effect of these setting when using different stimuli and/or tasks. In particular, how will the I-DT behave when administering a task that include pictures instead of text? When administering a search task instead of a reading task? Will the I-DT algorithm behave differently when using different values for its two thresholds?

Fig. 8. Fixations obtained from the I-DT algorithm for different combinations of the dispersion and duration thresholds depicted over the stimulus. The change of color from blue to yellow represents the time order of fixations where the dark blue represents the very beginning and the light yellow represent the very end. The size of the circles is proportional to the fixation duration.

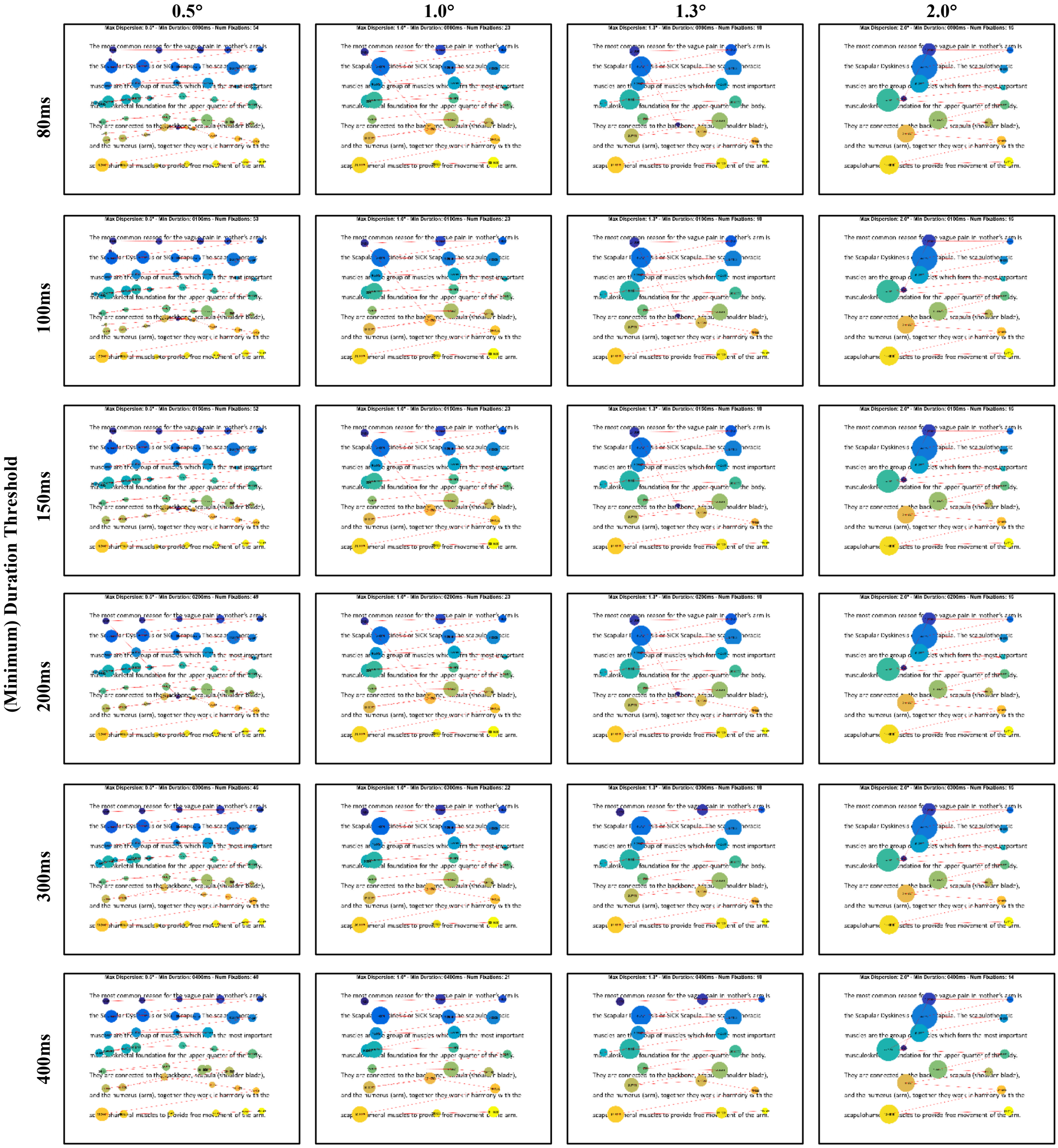


Table 3:The fixation duration (in millisecond) spent on each word in the stimulus reported by the I-DT algorithm for different combinations of values of the dispersion and duration thresholds. The empty cells if the table represent fixation duration of zero milliseconds, i.e., the I-DT algorithm does not report any eye fixation on that particular part of the stimulus when using the corresponding thresholds for dispersion and duration parameters.

Line No	Word ID	Word	0.5°						1.0°						1.3°						2.0°					
			80	100	150	200	300	400	80	100	150	200	300	400	80	100	150	200	300	400	80	100	150	200	300	400
1	1	The																								
	2	most	642	642	642	583	583	583	825	825	825	825	825	825												
	3	common				516	516	516							1358	1358	1358	1358	925	925						
	4	reason	542	542	542																					
	5	for																								
	6	the				492	492	492													2392	2392	2392	2392	2392	2392
	7	vague	608	608	608				1433	1433	1433	1433	1433	1433					1633	1633						
	8	pain																								
	9	in				925	925	925																		
	10	mother’s	575	575	575										1742	1742	1742	1742								
	11	arm	525	525	525	533	533	533	833	833	833	833	833	833					550	550	558	558	558	558	558	558
	12	is																								
2	13	the																								
	14	Scapular	2199	2199	2199	2041	2041	2041																		
	15	Dyskinesis	2217	2217	2217	2217	2217	2217	4374	4374	4374	4374	4374	4374	5424	5424	5424	5424	5424	5424						
	16	or																								
	17	SICK	1133	1133	1133	1133	1133	1133													8598	8598	8598	8598	8598	8598
	18	Scapula.	834	834	834	834	834	834	2616	2616	2616	2616	2616	2616												
	19	The																								
	20	scapulothoracic	3232	3232	3232	3232	3232	3232	2525	2525	2525	2525	2525	2525	4108	4108	4108	4108	4108	4108	933	933	933	933	933	933
3	21	muscles	867	867	867	867	867	867																		
	22	are																								
	23	the							2142	2142	2142	2142	2142	2142												
	24	group	1242	1242	1242	1242	1242	1242							2667	2667	2667	2667	2667	2667						
	25	of																								
	26	muscles	1208	1208	1208	1208	1208	1208													4350	4350	4350	4350	4350	4350
	27	which																								
	28	form	1500	1500	1500	1500	1500	1500	2208	2208	2208	2208	2208	2208												
	29	the													2066	2066	2066	2066	2066	2066						
	30	most	183	183	183																					
	31	important	333	233	233	241	325		366	366	366	366	366								383	383	383	383	383	
4	32	musculoskeletal	4167	4167	4167	4167	4157	4141	5874	5874	5874	5874	5874	5874	6216	6216	6216	6216	6216	6216	7091	7091	7091	7091	7091	7124
	33	foundation	1659	1659	1659	1659	1658	1275													425	425	425	425	425	425
	34	for																								
	35	the																								
	36	upper							1400	1400	1400	1400	1400	1400												
	37	quarter	533	533	533	533	533	533																		
	38	of													2050	2050	2050	2050	2050	2050						
	39	the																								
	40	body.	1150	1150	1150	1150	1150	1150	975	975	975	975	975	975							1192	1192	1192	1192	1192	1192
5	41	They	500	500	500	500	500	500																		
	42	are	350	350																						
	43	connected	333	333	333	333	333		1066	1066	1066	1066	1066	1066	1700	1700	1700	1700	1700	1700						
	44	to																			4166	4166	4166	4166	4166	4166
	45	the																								
	46	backbone,	2190	2190	2190	1999	1625	1541	2033	2033	2033	2033	1783	1783	258	258	258	258			4425	4425	4425	4425	4425	4425
	47	scapula	683	683	683	692	392	2150	3067	3067	3067	3067	3067	3067	5333	5333	5333	5333	5333	5333						
	48	(shoulder	1775	1775	1775	1775	1775	1050	1142	1142	1142	1142	1142	1142							875	875	875	875	875	875
	49	blade),	1050	1050	1050	1050	1050																			
6	50	and																								
	51	the	141	141	350	350	350	900																		
	52	humerus	791	791	900	900	900	808	1975	1975	1975	1975	1975	1975	2741	2741	2741	2741	2741	2741						
	53	(arm),	775	775	808	808																				
	54	together																								
	55	they																								
	56	work					716	517																		
	57	in																								
	58	harmony	275	275	275	275	325	475	1308	1308	1308	1308	1308	1308												
	59	with	500	500	500	500	517	591							800	800	800	800	800	800	1099	1099	1099	1099	1099	1099
	60	the																								

7	61	scapulohumeral	1066	1066	1066	1066	3216	3299	3467	3467	3467	3467	3467	3467	3717	3717	3717	3717	3717	3717	4816	4816	4816	4816	4816	4816
	62	muscles	2441	2441	2441	2441	858																			
	63	to																								
	64	provide	858	858	858	858	375	475	833	833	833	833	833	833												
	65	free																								
	66	movement	375	375	375	375	458	642							1691	1691	1691	1691	1691	1691						
	67	of																								
	68	the	458	458	458	458	642	583	1650	1650	1650	1650	1650	1650												
	69	arm.	642	642	642	642	583	541							541	541	541	541	541	541	1133	1133	1133	1133	1133	1133
		Total	**40552**	**40452**	**40453**	**40095**	**39761**	**38494**	**42112**	**42112**	**42112**	**42112**	**41862**	**41496**	**42412**	**42412**	**42412**	**42412**	**42162**	**42162**	**42436**	**42436**	**42436**	**42436**	**42436**	**42086**

References

[1] Agarwal(pt), T. 2017. Musculoskeletal Disorder Commonly Found in Homemakers - Dr. Tanushree Agarwal(pt). Retrieved 5 February 2018, from https://www.icliniq.com/articles/physiotherapy-physio-health/musculoskeletal-disorder-commonly-found-in-homemakers

[2] Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. 2017. Visualization of Eye Tracking Data: A Taxonomy and Survey. Computer Graphics Forum, 0, 1–25. https://doi.org/10.1111/cgf.13079

[3] Blignaut, P. 2009. Fixation identification: The optimum threshold for a dispersion algorithm. Attention, Perception, & Psychophysics, 71, 881–895. https://doi.org/10.3758/APP.71.4.881

[4] Buurman, R. Den, Roersema, T., & Gerrissen, J. F. 1981. Eye Movements and the Perceptual Span in Reading. Reading Research Quarterly, 16, 227. https://doi.org/10.2307/747557

[5] Duchowski, A. T. 2002. A Breadth-First Survey of Eye-Tracking Applications. Behavior Research Methods, Instruments, & Computers, 34, 455–470.

[6] Guo, K., Mahmoodi, S., Robertson, R. G., & Young, M. P. 2006. Longer fixation duration while viewing face images. Experimental Brain Research, 171, 91–98. https://doi.org/10.1007/s00221-005-0248-y

[7] Henderson, J. M. 1992. Object identification in context: The visual processing of natural scenes. Canadian Journal of Psychology/Revue Canadienne de Psychologie, 46, 319–341. https://doi.org/10.1037/h0084325

[8] Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press.

[9] Jacob, R. J. K., & Karn, K. S. 2003. Eye Tracking in Human–Computer Interaction and Usability Research: Ready to Deliver the Promises. In Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research (pp. 573–605).

[10] Javal, E., Ciuffreda, K. J., & Bassil, N. 1990. Essay on the Physiology of Reading. Ophthalmic & Physiological Optics :

The Journal of the British College of Ophthalmic Opticians (Optometrists),
10, 381–4.

[11] Just, M. A., & Carpenter, P. A. 1980. A theory of reading: From eye fixations to comprehension. Psychological Review, 87(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

[12] Koć-Januchta, M., Höffler, T., Thoma, G.-B., Prechtl, H., & Leutner, D. 2017. Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study. Computers in Human Behavior, 68, 170–179. https://doi.org/10.1016/J.CHB.2016.11.028

[13] Moffitt, K. 1980. Evaluation of the fixation duration in visual search. Perception & Psychophysics, 27, 370–372. https://doi.org/10.3758/BF03206127

[14] Moret-Tatay, C., & Perea, M. 2011. Do serifs provide an advantage in the recognition of written words? Journal of Cognitive Psychology, 23, 619–624. https://doi.org/10.1080/20445911.2011.546781

[15] Nyström, M., & Holmqvist, K. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. Behavior Research Methods, 42, 188–204. https://doi.org/10.3758/BRM.42.1.188

[16] O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. 1984. Convenient fixation location within isolated words of different length and structure. Journal of Experimental Psychology: Human Perception and Performance, 10, 250–257. https://doi.org/10.1037/0096-1523.10.2.250

[17] Obaidellah, U., Al Haek, M., & Cheng, P. C.-H. 2018. A Survey on the Usage of Eye-Tracking in Computer Programming. ACM Computing Surveys, 51, 1–58. https://doi.org/10.1145/3145904

[18] Perea, M., & Gomez, P. 2012. Subtle increases in interletter spacing facilitate the encoding of words during normal reading. PLoS One, 7, e47568. https://doi.org/10.1371/journal.pone.0047568

[19] Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372–422. https://doi.org/10.1037/0033-2909.124.3.372

[20] Rayner, K. 2009. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology, 62, 1457–1506. https://doi.org/10.1080/17470210902816461

[21] Rayner, K., & Duffy, S. A. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition, 14, 191–201. https://doi.org/10.3758/BF03197692

[22] Salthouse, T. A., Ellis, C. L., Diener, D. C., & Somberg, B. L. (1981). Stimulus processing during eye fixations. Journal of Experimental Psychology: Human Perception and Performance, 7, 611–623. https://doi.org/10.1037/0096-1523.7.3.611

[23] Salvucci, D. D., & Goldberg, J. H.2000. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the symposium on Eye tracking research & applications - ETRA ’00 (pp. 71–78). New York, New York, USA: ACM Press. https://doi.org/10.1145/355017.355028

[24] Shayan, S., Abrahamson, D., Bakker, A., Duijzer, C. A. C. G., & van der Schaaf, M. 2017. Eye-Tracking the Emergence of Attentional Anchors in a Mathematics Learning Tablet Activity. In C. Was, F. Sansosti, & M. Bradley (Eds.), Eye-Tracking Technology Applications in Educational Research (pp. 166–194). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-5225-1005-5.ch009

[25] Tobii Technology Inc. 2010. Tobii TX300 Eye Tracker: Product Description. Retrieved 22 June 2017, from https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-tx300-product-description.pdf

[26] Tobii Technology Inc. 2013. Tobii Analytics SDK Developer’s Guide. Retrieved 22 June 2017, from http://www.acuity-ets.com/downloads/Tobii Analytics SDK Developers Guide.pdf

[27] Tobii Technology Inc. 2015. Tobii Pro SDK: Develop Eye Tracking Applications for Research. Retrieved 22 June 2017, from https://www.tobiipro.com/product-listing/tobii-pro-sdk/#Download

[28] Tobii Technology Inc. 2017. Tobii Pro Spectrum Product Description. Retrieved 5 February 2018, from https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-spectrum-product-description.pdf/?v=2.0

[29] Widdel, H. 1984. Operational Problems in Analysing Eye Movements. Advances in Psychology, 22, 21–29. https://doi.org/10.1016/S0166-4115(08)61814-2

[30] Widdel, H., & Kaster, J. 1981. Eye Movement Measurement in the Assessment and Training of Visual Performance. In J. Moraal & K.-F. Kraiss (Eds.), Manned Systems Design (pp. 251–270). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4613-3306-7_11

[31] Bojko, A., & Schumacher, R. M. 2008. Eye tracking and usability testing in form layout evaluation. In Proceedings of the 38th International Symposium of Business Forms Management Association (pp. 1 – 13). Las Vegas, USA.

[32] Graf, W., & Krueger, H. 1989. Ergonomic evaluation of user-interfaces by means of eye-movement data. In Proceedings of the third international conference on human-computer interaction (pp. 659-

665). Boston, USA.

[33] Jacob, R. J., & Karn, K. S. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. Mind, 2, 4.

[34] Just, M. A., & Carpenter, P. A. 1976. Eye fixations and cognitive processes. Cognitive psychology, 8, 441-480.

[35] Poole, A., & Ball, L. J. 2006. Eye tracking in HCI and usability research. In Encyclopaedia of human-computer interaction (pp. 211-219). Pennsylvania: Idea Group Inc.

[36] Cooke, L. 2006. Is eye tracking the next step in usability testing? In IEEE International Professional Communication Conference (pp. 236-242). Washington DC, USA.

[37] Graf, W., & Krueger, H. 1989. Ergonomic evaluation of user-interfaces by means of eye-movement data. In Proceedings of the third international conference on human-computer interaction (pp. 659-665). Boston, USA.

Journal of Engineering and Applied Sciences, Vol. 7, Issue (2) November 2020

Evaluating and Measuring the Impact of E-Learning System Adopted in Saudi Electronic University

Thamer Alhussain
E-commerce Department, Saudi Electronic University, talhussain@seu.edu.sa

Abstract

This research aims to investigate and quantify the impact that the Blackboard system has as an e-learning system, with the specific context of its use in Saudi Electronic University (SEU) using the IS Impact Measurement model. This study will help to provide rich insights and increased understanding on how the creation of effective and successful adoption of such Learning Management System (LMS) can be achieved. Thus, this research will help to improve the quality of existing learning environment in Saudi Electronic University (SEU) and other similar context. The significance of this research is that it addresses gaps in current knowledge for a special need of SEU to measure and evaluate its use of LMS after about eight years of implementation in order to increase knowledge on how to best apply such LMS on a blended learning environment.

Keywords:
E-learning; Blackboard; Learning Management System

1. Introduction

Information and communication technologies have been experiencing rapid growth, which has in turn led to the emergence of a unique set of opportunities for education, specifically when it comes to e-learning processes. Leaning Management Systems (LMSs) have been deployed by several institutes to augment learning. LMS uses contemporary technology to enhance accessibility and availability to learning around the world. According to Paulson [1] the spread of e-learning is the result of LMS. Virtual Learning Environment (VLEs) allow educational institutions “to develop electronic learning materials for students, to offer these courses electronically to students, to test and evaluate the students electronically, and to generate electronically student databases in which student results and progress can be charted” [1, p 2]. Using LMS allows learners to grasp their content easily, while tracking their courses. The instructors are also able to navigate things in a simpler manner because they are able to track and evaluate every student.

Given the value that this system brings to the table, evaluating its success and influence has become more and more important, especially to help augment the quality of the educational processes at play. Several studies, including Aceto et al. [2], Wang et al. [3], and Alkhalaf [4] look at why such systems are needed.

This research has been developed within the scope of information systems, gauging the impact and success of Blackboard systems being used at SEU. In particular, this work will employ the IS-Impact Measurement Model [5] to evaluate the affect that the system in question has so as to augment the existing learning environment's quality at SEU, which is the sole institute to use such a style of learning at the university level, within Saudi Arabia. Moreover, this research will use a single case study to satisfy its research aim. This case design is suitable in this research for testing a theory, anomaly or a special case [6, pp 38- 40]. According to Walsham [7] and Irani et al. [8] the use of a case design will allow researchers to investigate the phenomena in a better fashion for the purpose of developing a deeper description and understanding. This research extended the previous work [9; 10] that only discussed individual impact and information quality of Blackboard system. This paper further contributes with measuring the impact of "system quality" and "organisational impact" along with the perspectives from both students and teachers. It provides a comprehensive assessment of the impact of Blackboard system on the educational process in SEU. This paper is structured as follows. It begins with a brief background relating to SEU. Next, overview about learning management systems is presented. The paper then discusses the research methodology and finally concludes by presenting the survey results and outcomes of the study.

2. Saudi Electronic University (SEU)

King Abdullah Bin Abdul-Aziz issued a royal decree on August 10, 2011, so that the university in question could be launched as a governmental educational institute. It is focused on blended learning and is the only one of its kind in the country to allow for both graduate and undergraduate programs. The university has a vision to be one that performs with excellence when it comes to adopting ICTs to build knowledge society. Its mission is to offer "high quality programs for all segments of society through the use of blended learning. It promotes knowledge production and community services to achieve development goals by optimal utilization of technology and engagement in local and global partnerships" [11].

Saudi Electronic University adopts a blended learning pattern as a recent learning method used in world's universities, it uses a combination of e-learning and direct traditional edification. The blended learning method here uses a mixture so that it can provide the optimum advantage from technology to students and teachers, helping each reach their learning goals effectively.

3. Literature review

Universities which have adopted Learning Management System (LMS) normally refer it as a Virtual Learning Environment [12;1]. Learning Management System (LMS) has been given different meanings by different researchers. Ayub et al. [13] explained it as a web-based technology that served the purpose of creating, dissemination and evaluating a given learning process. It is essentially a software that has been created to inform the process of learning, alongside providing resources with which the learner can augment their understanding of a subject or topic. It can also be seen as a set of tools and framework that allow for hassle-free creation of content online, and at the same time guiding learning [12]. Wahlstedt and Honkaranta [14] confirm that it is an advanced form of traditional learning, and contains instructional and evaluation devices, and learning contents. Fairly unique with learning management system is the fact that it can be useful when it comes to playing, disseminating, and managing learning, therefore integrating multiple tasks that were previously handed over to various stakeholders. Management under this system includes delivery, exams, tracking progress, statistical evaluation and virtual lessons [12; 15; 16]. Learning Management System can be defined as "web-based software platforms that provide an interactive online learning environment and automate the administration, organization, delivery, and reporting of educational content and learner outcomes." [16]. This is the reason it is a crucial tool when it comes to institution management, because it helps combine multiple factors together.

This is an essential platform through which teachers and their students can connect and simultaneously exchange and share materials. Therefore, it can be said that internet-based tech solutions are beneficial for both the teachers and the students since they made it possible for the two to engage through useful interactive features such as file sharing platforms, forums and discussion boards [12]. The Learning Management System can be used by the instructors to distribute courses and at the same time aiding in instructor-learner interaction [17]. The management function of Learning management System is particularly of great importance because it requires less effort and it also saves time that could otherwise been wasted by the instructor without changing the entire instructional process. Communication tools, virtual classes, and discussion forums are the key characteristics of Learning Management System [12]. These features make an interactive learning environment possible.

"Learning Management System has tremendous effect on e-learning. According to Paulsen [1], the presence of a Learning Management System will determine how e-learning will succeed. With a Learning Management System in place, an institution can easily develop web content, teach electronically, evaluate learners electronically and generate learners' databases for which the learners can access their results [1]."

Although it provides much support when it comes to e-learning, there has been a gap noticed in terms of the situation and advanced instructional tools, including multimedia, which are thought to help augment the level of instruction being handed out [17]. Multimedia tools, in several cases, are not deployed. And if they are deployed, the instructors are unable to use them to their full potential. For instance, several institutes are using LMS to help augment e-learning, however, teachers place limits on their own selves when they upload course material and do not ever get around to using other features such as forums that can help increase interaction and learning through discussions [17]. In some cases, users have reported being discouraged because they fail to acquire immediate feedback from tools such as the email [17]. Even though the features exist within the system, their use is limited because of the actions of the stakeholders. The system can act as a bridge to fix the issues that are present. However, this is only possible if the system is built to be adaptive and customizable [17]. Creating such a system will make it possible to ensure that teachers and students with varying levels of digital literacy can use it with ease.

The Blackboard is an LMS and VLE. It is one of the famous E-learning systems that use the electronic educational technology in learning and teaching with features for online collaboration and interaction. Colleges and universities use the Blackboard System to deliver online courses and augment on-campus courses. Also allow instructors to create, deliver, and manage web-based components for e-courses. In addition, allow students to download and upload files, attend virtual classrooms, participate in discussion forums and to submit assessments. Also students are able to view their grades and send e-mails to instructors and classmates within courses, and these services are available for instructors and students anytime and anywhere by using the Internet.

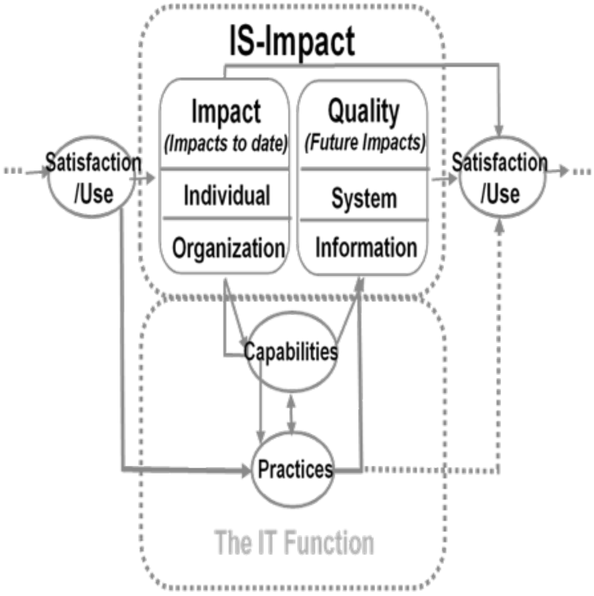
Saudi Electronic University [11] uses different products of Blackboard system including:

- Blackboard Learn – Provides a learning platform that includes course delivery, community engagement, and content management functionality as well as mobile applications
- Blackboard Collaborate –integrates -in-class web conferencing into the learning environment
- Blackboard Ally - a product that focuses on making digital course content more accessible.
- Analytics for Learn - a comprehensive learning analytics solution.

4.Methodology

This work will be employing a positivist research paradigm, which looks to verify hypotheses and validate theories for the purpose of evaluating the real world as it is [5; 19]. This research will analyze and gauge Blackboard use by testing IS-Impact Measurement model developed by Gable et al. [5; Figure 1]. This model was chosen for its ability to take stock of ICTs comprehensively, through use of 37 measures in four significant aspects, i.e., “System Quality,” “Information Quality,” “Individual Impact," and "Organisational Impact."

Fig. 1. IS-Impact Measurement model (Gable et al. 2008)



Much has been done with regards to research on information system success measurement since its emergence into the academic disciplines in the 1990's. Despite having been researched extensively by use of different approaches, scholars have failed to come up with a unified decision on a common measure if information system success. As a result of these diverse studies, scholars have had differing opinions thus coming up with different models. However, the most discussed IS success model are the DeLone and McLean [20] and the IS-Impact Measurement model by Gable and others [5; 21]. This model emerged upon the review of communications' research done by Shannon and Weaver [22] and the information influence theory work done by Mason [23]. According to Gable et al. [5; 21], it was difficult to use these studies in the IS success measurement owing to the fact that researchers had lumped together different factors of success thus making it difficult to use [20]. The model was borne of the synthesis and harmonization of these earlier separated measures [24]. This model includes six main IS success constructs and they include: system quality, use, information quality, user satisfaction, single person and organizational impact [24; 25].

The model was then subjected to review by Seddon [26] to give an advanced version of the D&M model. In his review Seddon [26] removed the interpretation process of the D&M model and fragmented the remainder into two different models; the partial behavioral model of IS use and the IS success model [3]. The updated IS model done by Seddon has six dimensional components, as previously discussed [3].

Despite the fact that these three IS models; DeLone & McLean [20], Seddon model and Update DeLone & McLean model have contributed immensely to the IS success measurement research, they have failed to address several issues. One major setback of these models that has not been addressed is failure to address the constructs as either formative or reflective [21]. A good example of such as a confusion was highlighted by Petter et al. [27] and they argue that constructs were under threat of being mis-specified and validated as reflective while close scrutiny could easily reveal that they are formative. In addition to this, the D&M IS success model has widely been criticized for failure to offer a good and clear explanation on its theoretical and epistemological basis [21]. It is for these reasons that the IS-impact model has always been suggested as the best option to address these weaknesses.

The IS-Impact measurement model is "more comprehensive and valid model for use. According to [24], this model has been tested statistically though surveys and has proven to be valid and it employs the perceptual measures. These tests depicted the validity and reliability of this model. Despite borrowing heavily from the DeLone & McLean model by adopting its constructs, it has succeeded in employing them for a different purpose [21]. The model and approach employs perpetual measures, aiming to offer a common instrument answerable by all relevant stakeholder groups, thereby enabling combining or comparison of stakeholder perspectives [25]."

"Moreover, a study conducted by Alotaibi [28] validated the IS-Impact Model and emphasize on the completeness and validity of IS-Impact Model as a Hierarchical Multi-dimensional Formative Measurement Model in the Saudi Arabian context. Accordingly, this model has been adopted in this research owing to its strengths in comparison to other models. It is quite clear that this model has eliminated all the weaknesses of other models by including and reviewing their constructs. Furthermore, since a single case design is more suitable for such research aims to test a theory, anomaly or a special case [6], this research will use a single case design to delve more deeply into the phenomena in order to insure that a rich description and understanding be provided." It will use a case study to help achieve the aim of this research which is to evaluate and measure the impact of Blackboard system adopted in Saudi Electronic University for the purpose of improving the quality of existing learning environment. As mentioned by Benbasat et al. [29; p 370], "A case study examines a phenomenon in its natural setting, employing multiple methods of data collection to gather information from one or a few entities (people, groups, or organizations)."

As can be observed in Figure 1, within the scope of the framework in question, the impact and success of a system can be gauged through:

- quality of information produced (information quality),
- impact on individual users (individual impact),
- performance of system from a technical perspective (system quality),
- impact on relevant organisation (organisational impact).

More specifically, a questionnaire is used to collect data for this study. It is a decent method through which data can be acquired from a significantly large group. It also helps bring together the answers to the research questions under study, and also acts as an efficient method to look into people's opinions and attitudes, in specific regarding the issue under study. The questionnaire will include 37 measures in the context of four aspects mentioned earlier in the IS-Impact Measurement Model. The questionnaire will contain two sections.

One of the sections will look at demographic information, while the other will focus on the 37 measures, previously discussed, alongside other dependent variables that can help look at validity. A Likert scale has been used to help rank the participants' responses.

5. Student questionnaire

Based on the tables containing the results of survey on: a) impacts of the Blackboard on the students individuals; b) Blackboard System information quality; c) Blackboard system quality; d) Students' satisfaction on the use of the blackboard.

Table 1. Demographics of the students survey participants

Qualifications	Frequency	Percent	Valid Percent	Cumulative Percent
High school	398	89.0	89.4	89.4
Bachelor degree	47	10.5	10.6	100.0
Total	447	100.0		

5.1. Impacts of the Blackboard on the students individuals

In this section we were able to asses on how the blackboard role on the individual performance. The respondents were expected to reflect on how the black board has helped them to improve on the "ability to interpret information accurately, understand the information and work related activities in their university, the decision making effectiveness and the overall productivity. The students answered the questions on a scale of 1-5 where 1 represents strongly disagree, 2- disagree, 3-Neutral, 4- agree, 5- strongly agree."

Table 2. Impacts of the Blackboard on the students individuals

N	Items	Strong disagree		Disagree		Neutral		Agree		Strongly agree	
		F	%	f	%	f	%	F	%	f	%
1	Blackboard impact on learning	10	2.2	32	7.2	78	17.4	237	53.0	90	20.1
2	Blackboard enhances awareness and job related information	8	1.8	36	8.1	74	16.6	246	55.0	83	18.6
3	Blackboard enhances effectiveness in the educational process.	11	2.5	46	10.3	92	20.6	217	48.5	81	18.1
4	Blackboard increases productivity	15	3.4	55	12.3	125	28.0	185	41.4	67	15.0

From the table 1, it is clear that from the 398 high school students and 47 bachelor degree students; table 2 shows that around two percent said they strongly agreed that the Blackboard augmented their learning, while 7.2% disagreed. Around 17% of the sample was neutral. Around 53% of the sample said that the board helped them, while 20% strongly agreed that it augmented their learning. The data showed that 73% of those queried were of the view that the blackboard augmented their learning, while 9.2% said that it did not; 17% stuck to neutrality. " 9.1% of the students say that Blackboard has not enhanced their awareness and recall of relative information, 73% agree that the Blackboard has enhanced their awareness and recall of relative information while 16 % are neutral on the same question.12% of the respondents say that the blackboard has not enhanced their effectiveness in the educational process, 66% say that the blackboard has enhanced their effectiveness in their educational process while 20% of the students that responded remained neutral on this matter. On the last question on this section 15% of the students disagree that the Blackboard has increased their productivity 28% of the students neither agree nor disagree while a total of 56% of the students believe that the black board has increases their productivity."

The data can be generalized to the students in the university. The standard deviations that were recorded were not particularly large, while the chi-square was noted to be significant. Therefore we conclude that the Blackboard has helped the students' ability to interpret information accurately and their overall productivity [30].

5.2. Blackboard system information quality

Table 3. Blackboard Information Quality

N	Items	Strong disagree		Disagree		Neutral		Agree		Strong Agree	
		F	%	f	%	f	%	F	%	f	%
5	Information available from Blackboard is important	7	1.6	19	4.3	60	13.4	236	52.8	125	28.0
6	Blackboard provides output that seems to be exactly what is needed	14	3.1	52	11.6	112	25.1	191	42.7	78	17.4
7	Information needed from Blackboard is always available	9	2.0	66	14.8	102	22.8	205	45.9	65	14.5
8	Information from Blackboard is in a form that is readily usable	19	4.3	41	9.2	107	23.9	212	47.4	68	15.2
9	Information from Blackboard is easy to understand	20	4.5	46	10.3	110	24.6	219	49.0	52	11.6
10	Information from Blackboard appears readable, clear and well formatted	17	3.8	41	9.2	101	22.6	218	48.8	70	15.7
11	Though data from Blackboard may be accurate, outputs sometimes are not	13	2.9	97	21.7	175	39.1	140	31.3	22	4.9
12	Information from Blackboard is concise	14	3.1	48	10.7	130	29.1	231	51.7	24	5.4
13	Information from Blackboard is always timely	20	4.5	55	12.3	112	25.1	203	45.4	57	12.8
14	Information from Blackboard is unavailable elsewhere	12	2.7	45	10.1	211	47.2	145	32.4	34	7.6

Under this category, the researcher is concerned with the timeliness, relevant and the accuracy of the information generated by the blackboard. Therefore, as well seen in table 3, we seek the answers of the ten questions all wish are intended to give us the insight to this matter.

The results show that 80% of the students were of the view that the information acquired through the blackboard is of importance. Around 5.3% thought it was not important, while 13% gave a neutral response. Over 59% thought the output being produced by the blackboard was precisely what was required, while 14% negated this notion, and 25 remained neutral.

Over 16% of the respondents reported that the system was not always timely, while 57% responded the opposite. Around 25% of the sample responded with neutrality. From the sample, 12% believed that the data found in the system was available elsewhere, 39.6% thought that it was not available anywhere else, and 47% chose to remain neutral. Despite the large number of neutral responses, the results can be generalized to the entire school since the answers present only a small standard variation, and the chi-square statistic was also within range. We can, therefore, conclude that the data is relevant, accurate and timely.

5.3. Blackboard system quality

Table 4. Blackboard system quality

N	Items	Strong disagree		Disagree		Neutral		Agree		Strong agree	
		f	%	f	%	f	%	f	%	f	%
15	Data from Blackboard often needs correction	23	5.1	179	40.0	142	31.8	73	16.3	30	6.7
16	Data from Blackboard is current enough	26	5.8	135	30.2	113	25.3	155	34.7	18	4.0
17	Blackboard is missing key data	12	2.7	131	29.3	167	37.4	105	23.5	32	7.2
18	Blackboard is easy to use	15	3.4	49	11.0	89	19.9	210	47.0	84	18.8
19	Blackboard is easy to learn	12	2.7	38	8.5	81	18.1	236	52.8	80	17.9
20	It is often difficult to get access to information that is in the Blackboard system	29	6.5	176	39.4	127	28.4	85	19.0	30	6.7
21	Blackboard meets university requirements	21	4.7	46	10.3	87	19.5	222	49.7	71	15.9
22	Blackboard includes necessary features and functions	8	1.8	24	5.4	76	17.0	256	57.3	83	18.6
23	Blackboard always does what it should	10	2.2	54	12.1	141	31.5	205	45.9	37	8.3
24	The Blackboard user interface can be easily adapted to one's personal approach	13	2.9	59	13.2	154	34.5	181	40.5	40	8.9
25	The Blackboard system is always up-and-running as necessary	44	9.8	126	28.2	83	18.6	150	33.6	44	9.8
26	The Blackboard system responds quickly enough	20	4.5	82	18.3	93	20.8	203	45.4	49	11.0

N	Items	Strong disagree		Disagree		Neutral		Agree		Strong agree	
		f	%	f	%	f	%	f	%	f	%
27	Blackboard requires only the minimum number of fields and screens to achieve a task	11	2.5	56	12.5	171	38.3	181	40.5	28	6.3
28	All data within Blackboard is fully integrated and consistent	12	2.7	48	10.7	148	33.1	208	46.5	31	6.9
29	Blackboard can be easily modified, corrected or improved.	13	2.9	60	13.4	193	43.2	152	34.0	29	6.5

From the data collected from the respondents represented in table 4, it is clear that 45% of the students believe that the Data from Blackboard often need not to be corrected, while 22.7% of the students believe that the Data from Blackboard often needs to be corrected and 315 of the students remain neutral. This clearly indicates that most of the students are satisfied with the accuracy of the data from the blackboard. However, believe that Data from Blackboard is current not enough 35% and 38% believe that Data from Blackboard is current enough. Note that these percentages are almost equal and therefore reflect that the blackboard needs to be improved in such aware that it includes more data.

On the other hand a substantial number of students feel that the Blackboard is easy to use, easy to learn and it is not difficult to get access to information that is in the Blackboard system since from the table C, the data indicates that most of the students believe so as compared to the smaller number of the students with a contrary opinion. 64% of the students believe that Blackboard meets university requirements while 14% of the students believe that the Blackboard does not meets university requirements. Blackboard can be easily modified, corrected or improved since with the standard deviation of 0.873 and a chi-square statistic of 279.65 clearly indicate that 44%of the students believe that the blackboard has made it easy to correct and modify the data available from the blackboard.

Therefore, we can conclude that the Blackboard system is a multifaceted system designed to capture the system performance. The Blackboard system is considered to be of high quality, consistent and easy to maintain since the standard deviation of the answers of the questions is small while the chi-square statistic is significant [31].

5.4. Students' satisfaction on the use of Blackboard system

Table 5. Satisfaction on the use of Blackboard system

N	Items	Strong disagree		Disagree		Neutral		Agree		Strong agree	
		F	%	f	%	f	%	F	%	f	%
30	Overall, the Blackboard System Quality is satisfactory	9	2.0	47	10.5	102	22.8	242	54.1	47	10.5
31	Overall, the Blackboard Information Quality is satisfactory	12	2.7	47	10.5	101	22.6	238	53.2	49	11.0
32	Blackboard is enjoyable to use	26	5.8	75	16.8	116	26.0	168	37.6	62	13.9
33	Overall, Blackboard is satisfactory	11	2.5	34	7.6	96	21.5	257	57.5	49	11.0
34	Overall, Blackboard system related knowledge has been managed satisfactorily.	33	7.4	57	12.8	107	23.9	215	48.1	35	7.8
OVERALL											
35	The impact of Blackboard on the teaching has been positive.	13	2.9	40	8.9	78	17.4	220	49.2	96	21.5
36	The impact of Blackboard on me has been positive.	14	3.1	37	8.3	82	18.3	226	50.6	88	19.7

Under this category, the researcher is concerned with the satisfaction level of the students on the use of the blackboard by evaluating the various consequences of the blackboard. Table 5 clearly demonstrates the following facts: For example, 64% of all students are satisfied with the Overall quality Blackboard System Quality while only 12% are not satisfied with the Overall quality Blackboard System Quality. 64 % of the students are satisfied with the overall quality of the information available from the Blackboard while 21% of the students are not satisfied with the overall quality of the information available from the Blackboard. 68% of the students believe that Blackboard is enjoyable to use. 56% of the students believe that the blackboard is satisfactory overall and finally 69% of the students believe that the overall system of the Blackboard system related to knowledge has been manage satisfactorily. From the data it is clear that the blackboard system is satisfactory.

6. Teachers questionnaire

The questions for teachers on the use of blackboard were grouped within the following categories for ease of understanding: A) Impacts of the Blackboard on teachers' individuals; B) Blackboard System Information Quality; C) Blackboard System Quality; D) Teachers Satisfaction on the use of Blackboard System.

Table 6. Demographics of the survey participants

Qualifications	Frequency	Percent	Valid Percent	Cumulative Percent
Bachelor	1	1.9	1.9	1.9
Master	28	53.8	53.8	55.8
PhD	23	44.2	44.2	100.0
Total	52	100.0	100.0	

6.1. Impacts of the Blackboard on the teachers' individuals

Here we assessed the impact of the blackboard on the individual performance. The respondents were expected to reflect on how the black board has helped them to improve on the ability to interpret information accurately, understand the information and work related activities at their university, the decision-making effectiveness and the overall productivity. The teachers answered the questions on a scale of 1-5 where 1 represents strongly disagree, 2- disagree, 3-Neutral, 4- agree, 5- strongly agree.

Table 7. Impacts of the Blackboard on the teacher individuals

N	Items	Strong agree		Agree		Neutral		Disagree		Strong disagree	
		f	%	f	%	f	%	f	%	f	%
1	Blackboard impact on learning	19	35.8	26	49.1	8	15.1	0	0	0	0
2	Blackboard enhances awareness and job related information	13	24.5	27	50.9	12	22.6	0	0	1	1.9
3	Blackboard enhances effectiveness in the educational process.	20	37.7	28	52.8	3	5.7	2	3.8	0	0
4	Blackboard increases productivity	20	37.7	25	47.2	7	13.2	1	1.9	0	0

As represented in table 6, researcher surveyed 52 higher education faculties teaching different higher education levels. While table 7 demonstrates survey results as follows:

Whether the blackboard helped "teachers learn much, 0% responded that they strongly disagree that the use of the blackboard has helped them to learn much, as well 0% disagreed on the same, while 15.1% remained neutral. On the other hand, 49.1% of the respondents agreed that the presence of the black board has helped them to learn much and 35.8% of the respondents strongly agree that the presence of the blackboard has helped them to learn much. It is evident from the data collected that 84.9% of the respondents say that the presence of the blackboard has helped them to learn much, 0% say that the presence of the blackboard has not helped them to learn much while 15.1% remain neutral on the same."

Regarding enhancement of the teachers cognizance and memory of job information, 1.9% of the teachers say that Blackboard did not augment either, 75.4% reported that it did, and 22.6% remained neutral. But what concerns the role of Blackboard on effectiveness in the educational process we can see that 3.8% of the respondents were of the view that it did not augment their efficiency in terms of the educational process, 90.5% say that it did, while 5.7% remained neutral. For the last question, on whether the Blackboard has increased their productivity, 1.9% disagree that it increased their productivity, 13.3% remained neutral, while a total of 84.8% of the teachers believe that their productivity saw a rise because of it.

The data can be generalized since teachers both in higher education degree levels since the standard deviations are insignificant, whereas chi-square statistic are substantial. Therefore, we conclude that the Blackboard has helped the teachers' ability to interpret information accurately, and their overall productivity.

6.2. Blackboard system information quality

Table 7. Impacts of the Blackboard on the teacher individuals

N	Items	Strong agree		Agree		Neutral		Disagree		Strong disagree	
		f	%	f	%	f	%	f	%	f	%
5	Information available from Blackboard is important	16	30.2	32	60.4	5	9.4	0	0	0	0
6	Blackboard provides output that seems to be exactly what is needed	9	17.0	29	54.7	13	24.5	2	3.8	0	0
7	Information needed from Blackboard is always available	10	18.9	26	49.1	14	26.4	2	3.8	1	1.9
8	Information from Blackboard is in a form that is readily usable	9	17.0	27	50.9	15	28.3	1	1.9	1	1.9
9	Information from Blackboard is easy to understand	14	26.4	26	49.1	10	18.9	3	5.7	0	0
10	Information from Blackboard appears readable, clear and well formatted	16	30.2	25	47.2	10	18.9	1	1.9	1	1
11	Though data from Blackboard may be accurate, outputs sometimes are not	4	7.5	4	7.5	21	39.6	5	9.4	0	0
12	Information from Blackboard is concise	5	9.4	36	67.9	10	18.9	2	3.8	0	0
13	Information from Blackboard is always timely	27	27	27	50.9	10	18.9	4	7.5	0	0
14	Information from Blackboard is unavailable elsewhere	5	9.4	14	14	19	35.8	15	28.3	0	0

Under this category, Table 8 demonstrate the survey results of the researcher's concern with the timeliness, relevant "and the accuracy of the information generated by the blackboard. Therefore, we seek the answers for the ten questions all wish are intended to give us the insight to this matter. 90.6% of the teachers believe that the information available from the blackboard is important, 0% of the teachers believe that the information available from the blackboard in not important while 9.4% are neutral. 71.7% of the respondents believe that the Blackboard provides output that seems to be exactly what is needed while 3.8% of the respondents believe that the Blackboard provides output that does not seem to be exactly what is needed while 24.5 remain neutral. 7.5% of the teachers believe that the Information from Blackboard not always timely, while 77.9% of the students believe that the Information from Blackboard always timely, while 18.9% of the students are neutral on the same." Around 28.3% believed that the information can be found elsewhere, 23.4% believed otherwise, while 35.8 % were neutral. Although in this case, the teachers that felt that they should remain neutral presented in much larger numbers. The outcome of our study can be generalized to the population of teachers in universities at different educational levels, given that all the responses that we have gathered only showed a small standard variation. This shows that the standard error will be insignificant and the chi-square statistic is within range. Hence, we can conclude that the blackboard's information is accurate, timely and relevant.

6.3. Blackboard system quality

Table 9. Blackboard system quality

N	Items	Strong agree		Agree		Neutral		Disagree		Strong disagree	
		f	%	f	%	f	%	f	%	f	%
15	Data from Blackboard often needs correction	3	3	15	28.3	17	32.1	17	32.1	1	1.9
16	Data from Blackboard is current enough	7	13.2	29	54.7	11	20.8	5	9.4	1	1.9
17	Blackboard is missing key data	2	3.8	8	15.1	27	50.9	16	30.2	0	0
18	Blackboard is easy to use	19	35.8	26	49.1	6	11.3	2	3.8	0	0
19	Blackboard is easy to learn	22	41.5	29	54.7	1	1.9	0	0	1	1.9
20	It is often hard to acquire access to information within the system	2	3.8	11	20.8	17	32.1	22	41.5	1	1.9
21	Blackboard meets university requirements	11	20.8	30	56.6	10	18.9	1	1.9	1	1.9
22	Blackboard includes necessary features and functions	11	20.8	31	58.5	8	15.1	3	5.7	0	0
23	Blackboard always does what it should	6	11.3	27	50.9	14	26.4	5	9.4	1	1.9

N	Items	Strong agree		Agree		Neutral		Disagree		Strong disagree	
		f	%	f	%	f	%	f	%	f	%
24	The Blackboard user interface can be easily adapted to one's personal approach	8	15.1	24	45.3	15	28.3	6	11.3	0	0
25	The Blackboard system is always up-and-running as necessary	5	9.4	28	52.8	12	22.6	7	13.2	1	1.9
26	The Blackboard system responds quickly enough	8	15.1	27	50.9	9	17.0	9	17.0	0	0
27	Blackboard requires only the minimum number of fields and screens to achieve a task	4	7.5	31	58.5	15	28.3	3	5.7	0	0
28	All data within Blackboard is fully intertwined and reliable	6	11.3	26	49.1	16	30.2	5	9.4	0	0
29	Blackboard can be easily altered, fixed or augmented.	2	3.8	8	15.1	27	50.9	16	30.2	0	0

From table 9 containing the data collected from the respondents, it is clear that 34% of the teachers believe that the Data from Blackboard often need not to be corrected, while 31.3% of the teachers believe that the Data from Blackboard often needs to be corrected and 32.1 of the teachers remain neutral. Note that these percentages are almost equal and therefore reflect that the accuracy of the data from the blackboard needs to be improved. However, believe that Data from Blackboard is current not enough 1.3% and 67.9% believe that Data from Blackboard is current enough. This clearly indicates that most of the teachers are satisfied with the blackboard currency. On the other hand a substantial number of teachers feel that the Blackboard is offer ease of use and learning, and does not pose difficulty when it comes to accessing data and information within the system, since results indicate that most of the teachers believe so as compared to the smaller number of the teachers with a contrary opinion. 77.4% of the teachers believe that Blackboard meets university requirements while 3.8% of the teachers believe that the Blackboard does not meets university requirements. Blackboard can be corrected, altered, or augmented with easy given that the standard deviation of 0.911 and a chi-square statistic of 41.62 clearly indicate that 18.9% of the teachers believe that the blackboard has made it easy to correct and modify the data available from the blackboard.

6.4. Teachers' satisfaction on the use of Blackboard system.

Table 10. Satisfaction on the use of Blackboard system

N	Items	Strong agree		Agree		Neutral		Disa-gree		Strong disagree	
		f	%	f	%	f	%	f	%	f	%
30	Overall, the Blackboard System Quality is satisfactory	7	13.2	33	62.3	12	22.6	1	1.9	0	0
31	Overall, the Blackboard Information Quality is satisfactory	9	17.0	35	66.0	7	13.2	1	1.9	1	1.9
32	Blackboard is enjoyable to use	14	26.4	29	54.7	7	13.2	3	5.7	0	0
33	Overall, Blackboard is satisfactory	10	18.9	34	64.2	8	15.1	1	1.9	0	0
34	Overall, Blackboard system related knowledge has been managed satisfactorily.	8	15.1	35	66.0	8	15.1	2	3.8	0	0
OVERALL											
35	….the impact of Blackboard on the teaching has been positive.	14	26.4	32	60.4	6	11.3	1	1.9	0	0
36	….the impact of Blackboard on me has been positive.	19	35.8	29	54.7	5	9.4	0	0	0	0

Under this category, table 10 shows the survey's results of researcher's concern with the satisfaction level of the teachers on the use of the blackboard by evaluating the various consequences of the blackboard. For example, 75.5% of all surveyed reported that they found the overall quality of the blackboard system to be satisfactory, while only 1.9% reported the opposite. Around 84 % of the students reported they were satisfied with the overall information quality, while 3.8% reported the opposite. 81.1% of the teachers believe that Blackboard is enjoyable to use. 83.1% of the teachers believe that the blackboard is satisfactory overall and finally 81.1% of the teachers believe that the overall system of the Blackboard system related to knowledge has been manage satisfactorily. From the data, it is clear that the blackboard system is satisfactory.

7. Conclusion

This paper outlined and discussed the results of an evaluation for the impact of Blackboard system adopted in SEU in order to improve the quality of existing learning environment and other similar context. Importantly, results supported a number of findings reported in the related literatures. The paper indicates that impact of Blackboard system positively affects the teaching and learning process for both students and teachers.

The analysis of the results points out that the Blackboard has equally helped the students and teachers' ability to interpret information accurately and their overall productivity; the information produced by the Blackboard is timely, accurate and relevant; and from the data, it is clear that the blackboard system is satisfactory. Further, it can be noted that the Blackboard system is a multifaceted system developed so that the performance of the system can be captured. In addition, it is considered to be of high quality, consistent and easy to maintain since the standard deviation of the answers of the questions is small while the chi-square statistic is significant.

While this research has evaluated the use of Blackboard system in SEU, further research is recommended to evaluate the implementation of Blackboard in other universities so that results can be compared and generalized.

Acknowledgements

The author would like to thank the Deanship of Scientific Research at Saudi Electronic University for funding this research under the number (7643-HS-2019-1-1-S).

References

[1] Paulsen, M. F., 2003. Experiences with Learning Management Systems in 113 European Institutions, Educational Technology & Society 6 (4), pp. 134-148.

[2] Aceto, S., Delrio, C., Dondi, C., Fischer, T., Kastis, N., Klein, R., 2007. e-Learning for Innovation: Executive Summary of the Helios Yearly Report 2007. MENON Network EEIG, Brussels.

[3] Wang, Y., Wang, H., Shee, D., 2007. Measuring e-Learning Systems Success in an Organizational Context: Scale Development and Validation, Computers in Human Behavior 23 (4), pp 1792-1808.

[4] Alkhalaf, S., 2013. Creating Effective e-Learning Systems for Higher Education in Saudi Arabia, PhD thesis, Griffith University, Australia.

[5] Gable, G., Sedera, D., Chan, T., 2008. Reconceptualizing Information System Success: The IS-Impact Measurement Model, Journal of the Association for Information Systems 9 (7), pp 377-408.

[6] Yin, R. K., 2009. Case Study Research: Design and Methods (4th ed.). SAGE Publications, Thousand Oaks, CA.

[7] Walsham, G. 1995, Interpretive case studies in IS research: nature and method. European Journal of Information Systems, vol. 4, pp. 74-81.

[8] Iivari, J, Hirschheim, RA and Klein, HK 1998, A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies. Information Systems Research, vol. 9, no. 2, p. 164-193.

[9] Alhussain, T., 2017. Measuring the Impact of the Blackboard System on Blended Learning Students, International Journal of Advanced Computer Science and Applications (IJACSA) 8 (3), pp 297-301.

[10] Alhussain, T., 2017. Assessing Information Quality of Blackboard System, International Journal of Computer (IJC) 25

(1), pp 1-7.

[11] Saudi Electronic University 2020, about the university, Accessed: January 15, 2020: https://www.seu.edu.sa/en/about/.

[12] Adzharuddin, A., Ling, H. L., 2013. Learning management systems among university students: does it work?, International Journal for e-Education, e-Business, e-Management, and e-Learning 3 (3), pp 248-252.

[13] Ayub, A. F. M., Rohani, A. T, Wan, M. W. J., Wan, Z. W. A., Luan, W. S., 2010. Factors Influencing Students' Use [of] a Learning Management System Portal: Perspective from Higher Education Students, International Journal of Education and Information Technologies 4 (2), pp 100-108.

[14] Wahlstedt, A., Honkaranta, A., 2007. Bridging the Gap Between Advanced Distributed Teaching and Use of Learning Management Systems in the University Context. Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007).

[15] Snoussi, T. 2019. Learning Management System in Education: Opportunities and Challenges, International Journal of Innovative Technology and Exploring Engineering 8 (12S), pp 664- 667.

[16] Turnbull, D., Ritesh C., and Jo L. 2019. "Learning Management Systems: An Overview." In Encyclopedia of Education and Information Technologies, edited by A. Tatnall. Cham: Springer Nature.

[17] Almarashdeh, A., Sahari, N., Zin, M., Alsmadi, M., 2010. The Success of Learning Management System Among Distance Learners in Malaysia Universities, Journal of Theoretical and Applied Information Technology 21 (2), pp 80-91.

[18] Guba, G., Lincoln, Y. S., 1994. Competing Paradigms in Qualitative Research, in "Handbook of Qualitative Research" N. K. Denzin, Y. S. Lincoln, Editors. SAGE Publications, Thousand Oaks, CA.

[19] Walliman, N., 2006. Social Research Methods. SAGE Publications, London.

[20] DeLone, W., McLean, E., 1992. Information Systems Success: The Quest for the Dependent Variable, Information Systems Research 3 (1), pp 60-95.

[21] Gable, G., Sedera, D. & Chan, T. 2008. Reconceptualizing information system success: the IS-impact measurement model. Journal of the Association for Information Systems, 9(7), 377-408.

[22] Shanon, C. E. and W. Weaver 1963. Mathematical Theory of Communication, Urbana, IL, University of Illinois Press.

[23] Mason, R. O. (1978), 'Measuring Information Output: A Communication Systems Approach', Information and Management, 1(4), 219-234.

[24] Rabaai, A. A., Gable, G., 2009. Extending the IS-Impact Model into the Higher Education Sector, Research in Progress, Queensland University of Technology, Brisbane.

[25] Elias, N. F., Cao, L., 2009. Validating the IS-Impact Model: Two Exploratory Case Studies in China and Malaysia. Pacific Asia Conference on Information Systems (PACIS) 2009 Proceedings.

[26] Seddon, P. B. 1997, 'A Respecification and Extension of the Delone and McLean Model of IS Success', Information Systems Research, 8(3), 240-253.

[27] Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," MIS Quarterly (31:4), pp. 623-656.

[28] Alotaibi, N., 2012. Extending and Validating the IS-Impact Model in Saudi Arabia: Accounting for Computer Network Quality, PhD thesis, Queensland University of Technology, Australia.

[29] Benbasat, I., Goldstein, D. K., Mead, M., 1987. The Case Research Strategy in Studies of Information Systems, MIS Quarterly 11 (3), pp 369-386.

[30] James, F., 2006. Statistical Methods in Experimental Physics (Vol. 7, No. 4). World Scientific, Singapore.

[31] Fornell, C., Larcker, D. F., 2011. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics, Journal of Marketing Research 18 (3),

Journal of Engineering and Applied Sciences (JEAS)