CHAPTER III

RESEARCH METHOD

This part discusses the methodological aspects of the study; they are (A) Research Design, (B) Source of Information, (C) Research Instrument, (D) Data Collection Procedure, and (E) Data Analyses Method. The theories are opined by some researchers who had conducted similar studies.

A. Research Design

Research design is a method utilized by a researcher to conduct and lead the research process. The design is determined based on the study objectives and hypotheses (Creswell, 2009).

This study focused on students' perception towards the use of Edubox as a testing method and its impact to their English learning motivation. Mix-method in data collection and data analyses to acquire richer and more meaningful result was utilized. Johnson, Onwuegbuzie, & Turner. (2007)

> "Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the purposes of breadth and depth of understanding and corroboration". (p. 123).

Furthermore, Creswell (2009) explains that a mix-method can offer the best fathom for quantitative and qualitative research or strengthen both to gain best fathom.

Quantitative method was applied for statistical data which represents numbers such as average, percentage or quotas. As Creswell (2003, p. 18) states that quantitative research "employs strategies of inquiry such as experimental and surveys, and collects data on predetermined instruments that yield statistical data". In this study, the quantitative method was chosen to answer the following research questions: (1) How students' perception and motivation mediated by Edubox contributes to their English learning performance and to what extent; (2) How the students' prior experience to online testing affects their English learning performance when tested using Edubox. Questionnaires and English test using Edubox were used as the instruments to collect quantitative data.

Whereas, qualitative method was employed for non-statistical data such as perception or students' attitudes. As Crossman (2020) opined that qualitative research is contrived to utter the significance of behavior or results usually measured by quantitative research. In other words, the researcher used qualitative method to obtain deeper and more detailed understanding about the population being studied (Glen, 2016). There are various factors that can lead to different perception about something and motivation to conduct an action; both variables are subjective and complex in nature. Therefore, qualitative data collection and analysis method would benefit this study tremendously. As Given (2008) explained that qualitative method is the best tool to find out how people enact a decision and hold the robust foundation to comprehend the government and social program. In addition, Seki (2004) stated that data and knowledge obtained by qualitative research can equip quantitative research ones.

Furthermore, perception will influence the way someone behaves. As Mohanta (2015, p 2) argues that

> People's behavior is based on their perception of what reality is, not on reality itself. There are some factors influencing perception: characteristics of the perceiver, the target, and the situation. The perceptual process is influenced by the perceiver's: past experience, needs or motives, personality, values and attitudes. The perceptual process is influenced by the situation's: physical context, social context and organizational context. The perceptual process is influenced by the target; the perceived person, object, or event, such as: contrast, intensity, figure-ground separation, size, motion, repetition or novelty.

This Study used the qualitative method to answer the third research problem; How students' perceptions of Edubox as an assessment mode affect their English learning motivation and the fourth one; What factors that affect students' attitudes to assessment mode. The observation and focused-group discussion were applied as research instruments.

B. Source of Information

This research utilized primary and secondary data. Ajayi (2017) in his paper explained that primary data is authentic and specific because the data is directly collected by the researcher. It is collected through observations, surveys, questionnaires, case studies, and interviews based on what he or she needs for his or her study. Whereas, secondary data is not original. It is unique because it is collected from the existing data, such as government publications, websites, books, journal articles, and internal records.

The primary data in this study was collected from the students as the informants by means of distributing questionnaires; conducting direct observation,

field notes, video recording; and administering focused-group discussion and English test using Edubox.

Secondary data was collected from some available documents, such as the result of English test using paper-based testing (PBT), English text books, syllabus, etc. The results obtained from English test using paper-based testing (PBT) was compared to the ones received by utilizing Edubox- a computer-based testing method. It complemented the primary data as well. The English text book and syllabus was reviewed by experts to ensure test validity and reliability.

Total research population in this study was 843 students divided into 320 7th graders, 296 8th graders, and 227 9th graders. The desired sample size was 30% of the total population. The study was undertaken at the school where the researcher conducted the study.

Table 3.1 Population

No	Grade	Numbers of grades	Numbers of Students
1	7th	10	320
2	8th	9	296
3	9th	7	227
	Total numbers		843

Source: School administration Office

Stratified sampling method was used in this study because research participants were from three different levels (7th, 8th and 9th graders). All levels had equal chance to be the sample (Hayes, 2020). As seen on the following table:

Table 3.2 Sample distribution

Research Instruments	Participants' Levels		
	7th	8th	9th
English Test	96	89	68
Questionnaires	96	89	68
Focused- Group Discussion	10	10	10
Observation	32	32	32

The English test scores were gained from the evaluation given to 30 % of students from each level; and the questionnaires were distributed to the same sample. The return rate of the questionnaires was 100%. The researcher decided to set the above sample size (30%) to ensure that it was large enough to give more meaningful data. Arikunto (2005) argues that if a researcher applies the questionnaires to collect data and he or she has several hundred population, he or she can determine more or less 25 – 30% from the total population. However, the researcher can take the total population if they are only 100 – 150 people.

While, FGD was organized to 30 students: consisted of 10 students from each level who were selected by means of purposive sampling method to gain deeper and more detailed information about the students' perspective. They were comprised of high, mediocre and low achievers. Palinkas et al (2015: 533) states: "Purposeful sampling is widely used in qualitative research for the identification and selection of information-rich cases related to the phenomenon of interest". Besides, this type sampling method was used to delineate a specific subgroup in depth, to diminish diversity, to simplify analysis and facilitate group interviewing.

Furthermore, observation was organized to 32 students from each level while they were taking English test using Edubox in mid semester test. The number of students observed was adjusted with the researcher's capability to conduct the activity. The English testing period were organized simultaneously at the school for all levels, hence, the researcher had to divide the appropriate time to visit each class to get real feeling of the condition during the test. Arikunto (2005) claims if the researcher utilizes interview and observation the numbers of participants can be adjusted based on his or her capability.

Thinh (2016) argues that to analyze qualitative data typically entails smaller sample size than quantitative one because it needs substantial enough to gain feedback for most or all perceptions. Obtaining them will cause the achievement of saturation. It occurs when adding more participants to the study yet there is no additional ancillary information or perspectives. Besides, there are no specific rules when deciding a convenient sample size in qualitative research. In determining proper sample size in qualitative studies: Morse (1994) suggests 30 to 50 participants to be interviewed; Creswell (1998) advises 20 to 30 participants; while Patton (1990) explains that in determining qualitative sample size the researcher should consider time allotted, resources available, and study objectives.

In a quantitative study, furthermore, it is suggested to have sample size of thirty at the minimum. This amount would allow us an adequate observation to take the benefit from the central limit theorem, i.e at n = 30 we start to see the bell shape curve if the data is normally distributed (LaMorte, 2016).

C. Research Instrument

Research instruments are tools applied by a researcher to collect data and to facilitate the researcher's work in gaining better result (Arikunto, 2006). Besides, Notoatmodjo (2010) stated that research Instruments are the means utilized to collect data. They consist of questionnaires, observation forms, and other forms to record the information.

The researcher triangulated the information needed by conducting 4 different data collection methods: 1) English test, 2) questionnaires, 3) observation and 4) focused-group discussion. Eladio, in Trigueros (2017, p.1) proposes that there are three methods mostly used in a qualitative research, namely observation, interview and questionnaires. The observation and interview are used in qualitative investigation. Whereas the test and the questionnaires are mostly used in quantitative research (Sharma, 2018. P 2). Each method is expected to give deeper, richer, and more meaningful results. The English test result was used to support the other three data collection techniques. The test measured the students' performance after they carried out the test using Edubox.

To ensure and judge the quality of research instruments in this study, namely the blueprint of English test items; the blueprint of observation and its guidance; the blueprint of questionnaires; the blueprint of FGD questions; two experts: (1) DR. Irianti Usman, CEO of Yes Research and Training Village and lecturer in educational psychology field, (2) DR. Endah Yanuarti, a staff of LPMP *(lembaga penjamin mutu pendidikan*) at LPMP Jawa Barat served as the validators. Their judgments had escorted the researcher in developing good research instruments.

Then, the researcher had adopted their comments or feedbacks to revise the points that needed revisions and refinements.

The researcher held trial session to some students that were not participants to check the readability and clarity of the questionnaires and English test items, and to see if the time portion was appropriate to answer them.

Each Instrument is elaborated as follow;

1. The English test items

a. The purpose of the test

The English test using Edubox was aimed at measuring the students' English learning motivation. As Buck (2001) argued that a test instrument was applied to measure some extent. English test scores taken using Edubox in this research was used as primary resource. Previous English test results documented by English teachers which were conducted by means of PBT method served as secondary data. It was meant to portray the difference of their performance using two different evaluation testing media.

The test items were designed based on basic cognitive competences dictated on *Kurikulum 13* syllabus for each level as seen on table 3.3. below. Detailed information can be gotten from the blue print in appendix 4. The test materials only tested two or three cognitive competences. There were thirty test items that needed to be answered in 80 minutes.

Table 3.3 The Basic Competencies for The English Mid-Test using Edubox

Grade	Basic Competencies
7	*3.1 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi interpersonal lisan dan tulis yang melibatkan tindakan menyapa, berpamitan, mengucapkan terimakasih, dan meminta maaf, serta menanggapinya, sesuai dengan konteks penggunaannya.*
	3.2 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait jati diri, pendek dan sederhana, sesuai dengan konteks penggunaannya. Perhatikan unsur kebahasaan dan kosa kata terkait hubungan keluarga; pronoun (subjective, objective, possessive).
8	*3.1 Menerapkan fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi interpersonal lisan dan tulis yang melibatkan tindakan meminta perhatian, mengecek pemahaman, mengharg ai kinerja, meminta dan mengungkapkan pendapat, serta menanggapinya, sesuai dengan konteks penggunaannya.*
	3.2 Menerapkan fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait kemampuan dan kemauan, melakukan suatu tindakan, sesuai dengan konteks penggunaannya. (Perhatikan unsur kebahasaan can, will).
	3.3 Menerapkan fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait keharusan, larangan, dan himbauan, sesuai dengan konteks penggunaannya. (Perhatikan unsur kebahasaan must, should.)
9	*3.1 Menerapkan fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi interpersonal lisan dan tulis yang melibatkan tindakan menyatakan harapan, doa, dan ucapan selamat atas suatu kebahagiaan dan prestasi, serta menanggapinya, sesuai dengan konteks penggunaannya.*

	3.2 Menerapkan struktur teks dan unsur kebahasaan untuk melaksanakan fungsi sosial dari ungkapan persetujuan, serta responnya, sesuai dengan konteks penggunaannya.
	3.3 Menerapkan struktur teks dan unsur kebahasaan untuk melaksanakan fungsi sosial menyatakan dan menanyakan tentang keharusan dan himbauan melakukan suatu tindakan/kegiatan, sesuai dengan konteks penggunaannya.

b. Validity and Reliability Test

1) Test Validity

Sudjana (2004: 12) stated that validity refers how accurately a method measures what it is intended to measure. For this purpose, the English test items were validated by two experts as aforementioned before the test was held. The first expert suggested that the items should be adjusted with the materials they have learned. While, the second expert commented there were some minor revisions needed to be done toward the test items when it came to presenting daily conversations in a proper English sense of language. The researcher revised the items in accordance with the experts' suggestions. Then, the test items were tried out. The validation form from the experts can be seen in appendix 6.

2) Test Reliability

Sudjana (2004: 16) argued that the reliability of an assessment tool deals with accuracy or constancy of that tool to assess what it is tested. In other words, research reliability is the degree to which research method produces stable and

consistent results. So, whenever that tool is used its result will be relatively the same.

To attest the validity and reliability of the test items, the researcher undertook a tryout to 32 students for each level who were not the sample for this research. The test results were automatically analyzed by Edubox system. This testing media presented scores as well as the validity and the reliability of test items. As seen on the following table:

Table 3.4.1
Reliability Test Results (7th grade)

Item number	Variant each item	Item number	Variant each item	Item number	Variant each item
1	0.26	11	0,22	21	0,22
2	0.25	12	0,26	22	0,26
3	0,24	13	0,22	23	0,24
4	0,26	14	0,26	24	0,19
5	0,25	15	0,09	25	0,26
6	0,09	16	0,26	26	0,19
7	0,24	17	0,26	27	0,11
8	0,18	18	0,23	28	0,26
9	0,21	19	0,26	29	0,03
10	0,25	20	0,21	30	0,03
Numbers of variant of items			6,29		
Variant of Students Scores			40,52		
Score Reliability			0,87		
Category			Very high		

Table 3.4.2
Reliability Test Results (8th grade)

Item number	Variant each item	Item number	Variant each item	Item number	Variant each item
1	0,2	11	0,26	21	0,2
2	0,19	12	0,25	22	0,25
3	0,26	13	0,26	23	0,26
4	0,23	14	0,17	24	0,17
5	0,26	15	0,25	25	0,25
6	0,15	16	0,22	26	0,17
7	0,25	17	0,26	27	0,26
8	0,25	18	0,26	28	0,23
9	0,19	19	0,26	29	0,19
10	0,26	20	0,26	30	0,2
Numbers of variant of items			6,87		
Variant of Students Scores			54		
Score Reliability			0,9		
Category			Very high		

Table 3.4.3
Reliability Test Results (9th grade)

Item number	Variant each item	Item number	Variant each item	Item number	Variant each item
1	0.06	11	0,25	21	0,26
2	0.21	12	0,21	22	0,25
3	0,11	13	0,23	23	0,24
4	0,24	14	0,25	24	0,21
5	0,09	15	0,19	25	0,21
6	0,26	16	0,25	26	0,26
7	0,21	17	0,24	27	0,21
8	0,18	18	0,19	28	0,21
9	0,25	19	0,18	29	0,25
10	0,16	20	0,14	30	0,25
Numbers of variant of items			6,25		
Variant of Students Scores			28,77		
Score Reliability			0,81		
Category			Very high		

Table 3.4.4
Validity Test Results (7th grade)

Item Number	r(x,y)	Category	Conclusion	Item Number	r(x,y)	Category	Conclusion
1	0,6	Fair	Valid	16	0,39	Low	Valid
2	0,47	Fair	Valid	17	0,72	High	Valid
3	0,5	Fair	Valid	18	0,38	Low	Valid
4	0,42	Fair	Valid	19	0,6	Fair	Valid
5	0,65	High	Valid	20	0,47	Fair	Valid
6	0,33	Low	Not valid	21	0,54	Fair	Valid
7	0,35	Low	Not valid	22	0,14	Very low	Not valid
8	0,43	Fair	Valid	23	0,59	Fair	Valid
9	0,36	Low	Valid	24	0,33	Low	Not valid
10	0,58	Fair	Valid	25	0,47	Fair	Valid
11	0,3	Low	Not valid	26	0,6	Fair	Valid
12	0,76	High	Valid	27	0,17	Very low	Not valid
13	0,63	High	Valid	28	0,71	High	Valid
14	0,57	Fair	Valid	29	0,32	Low	Not valid
15	0,24	Low	Not valid	30	0,23	Low	Not valid

Table 3.4.5
Validity Test Result (8th grade)

Item Number	r(x,y)	Category	Conclusion	Item Number	r(x,y)	Category	Conclusion
1	0,63	High	Valid	16	0,02	Very low	Not valid
2	0,66	High	Valid	17	0,45	High	Valid
3	0,13	Very low	Not valid	18	0,72	High	Valid
4	0,6	Fair	Valid	19	0,49	Fair	Valid
5	0,5	Fair	Valid	20	0,47	Fair	Valid
6	0,46	Fair	Valid	21	0,65	High	Valid
7	0,63	High	Valid	22	0,51	Very low	Not valid
8	0,44	Fair	Valid	23	0,43	Fair	Valid
9	0,61	High	Valid	24	0,36	Low	Valid
10	0,55	Fair	Valid	25	0,26	Low	Valid
11	0,76	High	Valid	26	0,51	Fair	Valid
12	0,65	High	Valid	27	0,71	High	Valid
13	0,86	Very	Valid	28	0,54	Fair	Valid
14	0,6	Fair	Valid	29	0,13	Very low	Not valid
15	0,62	High	Valid	30	0,68	High	Valid

Table 3.4.6

Validity Test Result (9h grade)

Item Number	r(x,y)	Category	Conclusion	Item Number	r(x,y)	Category	Conclusion
1	0,34	Low	Not valid	16	0,43	Fair	Valid
2	0,6	Fair	Valid	17	0,38	Low	Valid
3	0,26	Low	Not valid	18	0,37	Low	Valid
4	0,17	Very low	Not valid	19	-0,01	Very low	Not valid
5	0,47	Fair	Valid	20	0,41	Fair	Valid
6	0,63	High	Valid	21	0,38	Low	Valid
7	0,59	Fair	Valid	22	0,29	Low	Not valid
8	0,58	Fair	Valid	23	-0,46	Very low	Not Valid
9	0,42	Fair	Valid	24	0,56	Fair	Valid
10	0,48	Fair	Valid	25	0,42	Fair	Valid
11	0,69	High	Valid	26	0,5	Fair	Valid
12	0,57	Fair	Valid	27	0,59	Fair	Valid
13	0,42	Low	Not valid	28	0,46	Fair	Valid
14	0,3	Low	Not valid	29	0,43	Fair	Valid
15	0,18	Very low	Not valid	30	0,39	Low	Valid

The tables show that there are some invalid items. For example, in table 3.4.6 the items number 1, 3, 4, 13,14,15,19,22, and 23 were not valid. The detailed description can be seen in appendix 5. The researcher then removed and replaced those items with the ones which are valid and reliable before implementing the actual test.

2. The Questionnaires

a. The purpose of the questionnaires

The questionnaires were utilized to probe data which was corresponding with the research problems. In other words, the questions were developed based on the research questions. According to Bremer (2011) the data can be gained from a great amount of people primary using the questionnaires. The answers are measurable and comparatively simple to elaborate. However, the questionnaires may present

restricted supplementary concept. Since the answers have been provided before so the result of questionnaires may not describe whole responses that the participants really have.

b. The description of questionnaires

Arikunto (2006) explained that the questionnaires items were arranged after determining the purpose as aforementioned then identify variables and sub variables that are related to dimensions included in the research problems. As shown in the blueprint of questionnaires in appendix 7, the first dimension was related to the students' perception and motivation mediated by Edubox that may contribute to their English learning performance; the second aspect described the students' prior experience to online testing that may affect their English learning performance when tested using Edubox; the third element dealt with the students' perceptions towards Edubox as an assessment mode that may affect English learning motivation; the fourth feature delineated the factors that affect students' attitudes to assessment mode.

The questionnaires in this study applied Likert scale comprised of five responses, namely strongly agree, agree, neutral, disagree and strongly disagree. It is commonly used as qualitative instruments. Likert scale measures the different level of what participants felt or thought by choosing the level of agreement; strongly agree, agree, neither disagree nor agree. According to Bissonnette (2007) Likert-Scale is an instrument to assign the opinion, behavior and perception of participants. The questionnaires items were delivered in Bahasa Indonesia to ensure

validity, reliability, clear, and honest responses from the students: 7th graders, 8th graders and 9th graders.

c. Validity and Reliability

1) Questionnaires Validity

The blueprint of questionnaires and the questionnaires items were validated by the experts as aforementioned. They commented that the questionnaires items had been constructed in accordance with relevant data expected in the research questions. (See in appendix 9) Then, the questionnaires items had been tried out before implementing to the participants. Bolarinwa (2015. P 195-201) argued that "A drafted questionnaires should always be ready for establishing validity. Validity is the amount of systematic or built-in error in questionnaires". According to him, there are some methods used to test the validity of questionnaires such as face and content validity; criterion-related, and construct validity. He explained that face and content validity involved a panel of experts in exploring theoretical construct used in the questionnaires. These methods denoted how well the idea of a theoretical construct represented in the questionnaires. Whereas criterion-related and construct validity utilized a field test to verify how well a given measure relates to one or more external criterion, based on empirical constructs.

2) Questionnaires Reliability

Bolarinwa (2015. P 195-201) argued that "Reliability is an extent to which questionnaires, test, observation or any measurement procedure produces the same results on repeated trials". Morse (2001) stated that to verify reliability of the

questionnaires, it mostly utilizes a pilot test. So, the researcher conducted a tryout to 30 students not included in the sample. The data collected was then analyzed by SPSS 23 version.

d. Validity and Reliability of Questionnaires Test Results

1) Validity Test

Correlational technique used to examine each item in this study was Pearson product moment. If the value of a correlational coefficient of each item higher than r =0,254, it is concluded that construct is valid as shown in the table 3.5.1 to 3.5.4.

Table 3.5.1
Validity Test Results Related to The First Research Question

Items	Rcount	Rtable	Conclusion
Item 1	0,406	0,254	Valid
Item 2	0,547	0,254	Valid
Item 3	0,619	0,254	Valid
Item 4	0,672	0,254	Valid
Item 5	0,643	0,254	Valid
Item 6	0,530	0,254	Valid
Item 7	0,448	0,254	Valid
Item 8	0,543	0,254	Valid
Item 9	0,729	0,254	Valid
Item 10	0,700	0,254	Valid

Table 3.5.2
Validity Test Result Related to The Second Research Question

Items	Rcount	Rtable	Conclusion
Item 11	0,434	0,254	Valid
Item 12	0,495	0,254	Valid
Item 13	0,522	0,254	Valid
Item 14	0,388	0,254	Valid
Item 15	0,671	0,254	Valid
Item 16	0,807	0,254	Valid
Item 17	0,576	0,254	Valid
Item 18	0,675	0,254	Valid
Item 19	0,473	0,254	Valid
Item 20	0,554	0,254	Valid

Table 3.5.3
Validity Test Results Related to The Third Research Question

Items	Rcount	Rtable	Conclusion
Item 21	0,720	0,254	Valid
Item 22	0,481	0,254	Valid
Item 23	0,698	0,254	Valid
Item 24	0,648	0,254	Valid
Item 25	0,452	0,254	Valid
Item 26	0,576	0,254	Valid
Item 27	0,648	0,254	Valid

Table 3.5.4
Validity Test Results Related to The Fourth Research Question

Items	Rcount	Rtable	Conclusion
Item 28	0,747	0,254	Valid
Item 29	0,751	0,254	Valid
Item 30	0,741	0,254	Valid
Item 31	0,698	0,254	Valid
Item 32	5,544	0,254	Valid
Item 33	0,806	0,254	Valid

The validity test results of all items in the questionnaires indicated that they were valid and could be used as the data instrument tool for this study (r count of each item was higher than 0,254).

2) Reliability Test Results

In this study, Alpha Cronbach method was used to fulfill the needs for ensuring reliability of the items in the questionnaires. The items are reliable if the reliability coefficient is positive and higher than 0,70 as seen on the table 3.5.5 below:

Table 3.5.5
The Recapitulation of the Questionnaires Reliability Results

Questionnaires	Reliability Coefficient	Critical Score	Conclusion
RQ1	0,790	0,700	Reliable
RQ2	0,741	0,700	Reliable
RQ3	0,705	0,700	Reliable
RQ4	0,788	0,700	Reliable
Total	0,813	0,700	Reliable

The results of reliability test depicted that each item had the reliability coefficient higher than 0,700. So, the questionnaires were the reliable to measure the research variables.

3. Observation

a. The purpose of observation

The observation was conducted to give more comprehensive data concerning factual situation when the participants were taking the test using Edubox. According to Arikunto (2006:124) observation is a way to collect information that should be carried out through observing the object or the place being investigated directly.

b. The description of observation

Kawulich (2012. pp.150-160) states, "Observation is an excellent tool for collecting data in a variety of situations, and it requires a good memory and extensive note taking. In addition, it was defined that "An observation is the systematic description of the events, behaviors, and artifacts of a social setting" (Marshall & Rossman, 1989, p. 79). Furthermore, Guba and Lincoln in Moleong (2011) explained that observation has a very prominent role in qualitative research because observation based on the factual situation that can be observed by the researcher; the researcher can see and observe every single moment involving participants then the researcher takes some notes as factual data; the researcher can get a great number of information from the participants as the main source; and the

researcher can conceive the complicated and complex situation that being encountered by the participants.

In implementing the observation, the researcher constructed the observation guidance or the observation procedures. It helped the researcher to implement it conveniently. There were ten observation aspects. They were students' expressions while they were doing English test using Edubox; Students' expressions when they had some problems while they were taking English test; Students' expressions about supervisor who was present during the test; students' attitudes during English test using Edubox; the steps or the procedure of doing English test using Edubox; time allocation; internet connection; electrical obstacles; devices incompatibility; and server troubles. In implementing the observation, the researcher observed some points as observer is a part of system that is being observed.

c. Validity and Reliability

Johnson (1997, P.282) declares that a good research is a "plausible, credible, trustworthy, and, therefore, defensible and posits a number of strategies researchers can use to promote validity". In addition, validity in observation and interview is a must because the subject of measurement is human. So, to claim psychological aspects and variables from human is needed. Valid and reliable interview serve as the means to gain accurate conclusion in a research and to reflect the research subject accurately (Azwar, 2012).

d. Observation Validity and Reliability

To ensure the reliability of the observations and validate the procedures, the researcher conducted three steps while designing and refining the instrument. First, the researcher arranged the instruments which had been previously validated by the experts including: the blue print of observation and the observation guidance. (See in appendix 10 and 11). One of them explained that the researcher had conducted sound observation activities and that important points had been taken to enhance the data analyses processes.

The constructs were developed based on the research questions about students' perceptions of Edubox as an assessment mode and how it affected their English learning motivation. The procedure focused on the extent to which students are actively engaged in the process of doing testing using Edubox. The next step was, the researcher collaborated with three teachers to supervise the room where the test was taken. The researcher set a video recorder on a handphone that was kept using tripod in each selected room, room 1, 11 and 12. The supervisor of each room was asked to view a short videotape of an English test session using Edubox. Then, they completed an observation form while supervising the students who were undertaking the English test using Edubox. Whilst, the researcher as the main observer took some notes back and forth from three selected rooms. Afterwards, observers met to discuss findings on the overall number of what the students did during the test using Edubox, as well as details regarding each observation. In doing so, the researcher was able to compare and contrast results from the observations in order to ensure accuracy in the findings across all observers. Finally, the researcher conducted focused-group discussion with 30 students from 7^{th}, 8^{th} and 9^{th} grades,

in order to ensure that the observation results were accurate interpretations of students' experiences during the English test using Edubox in the classroom.

4. Focused-Group Discussion (FGD)

a. The purpose of focused-group discussion

Focused-group discussion assisted the researcher to obtain deeper and more elaborated students' opinions, beliefs, attitudes & perceptions about Edubox as a testing media. According to Anderson (1990) focused-group is an informal yet well-organized discussion administered in which one idea will be equipped with another idea inflicting a series of particular responses from it informatively. Its purpose is to discuss a particular topic deeply in a nice environment to gain various opinions, attitudes, feeling or perception from a group of persons sharing their general experience to the point investigated. The result of focused- group is a typical design of qualitative output leading comprehension about how people respond to an event or creation.

b. The description of FGD

There were 15 questions asked during the FGD activity. Those questions were derived from two constructs which were developed based on the research questions. Each construct has one dimension. The first dimension dealt with students' perception about Edubox as an assessment mode. The second one was related to some factors that might affect students' attitudes. Each dimension was divided into variables and sub-variables which were split into several indicators reflecting the

items of FGD. The whole format of FGD blueprint can be seen in more detail in Appendix 13.

Stewart and Shamdasani (1990) argued that focused-group discussion (FGD) is a great method where people having similar backgrounds and experiences criticize a specific topic of interest. It is led by a group facilitator who presents topics for discussion in a lively and natural discussion amongst themselves.

In this study, the researcher was a group facilitator and led three groups of FGD. Each group was attended by 10 students from each level (7^{th} grade, 8^{th} grade and 9^{th} grade) subsequently in a day. It was divided into three sessions. The questions asked about the implementation of Edubox as an assessment tool. They were divided into three parts; before implementing, whilst implementing, and after implementing the Edubox. Each discussion session lasted for about twenty until thirty. Most students expressed their negative feelings towards Edubox.

c. Validity and Reliability

1) Validity of FGD

To ensure the validity of FGD instruments, the blueprint and guidance were developed based on the research problems then the researcher established face and content validity tests by involving two experts as aforementioned. The experts suggested to create more specific questions.

"For focused-groups, validity could mean whether it is reasonably certain that people are talking about what you think they are talking about. Focused-group tend to be strong on validity" (Flom, 2017. Par 4).

2) Reliability of FGD

To verify the reliability of FGD instrument, the researcher conducted triangulation by utilizing recording feature in mobile phone while the FGD was being carried out. The results were then compared with the ones the researcher obtained from the observation.

According to Neuman (2007) interview method becomes a characteristic of qualitative research. Reliability in qualitative research emphasizes on constancy of the subject's responses. It is more flexible and expanding. If the results obtained is different, it does not mean that the reliability of instrument is poor, yet it could be concluded that the result has gained more complete or richer information.

Furthermore, Flom (2017, par 3) described, "reliability in FGD could concern whether another focused-group of similar but different people would give similar answers. Focused-groups often have problems with reliability. These can be lessened if the moderator is highly trained and if questions are relatively specific".

The students' responses from the three groups invited to the discussion were almost similar. Thus, this indicated that the FGD instrument was reliable. Chapter 4 will give more detailed elaboration concerning the students' responses.

D. Procedures of Data Collection

Prior to conducting the research, the researcher requested permit letter to the principal of the school where the researcher conducted the study. After the approval was granted, the next stage was sending consent letter to be filled out by the research's participants' parents due to the fact that the students were considered to

be minors (below 17 years old). Both letters (school permit and participant consent letter) was attached on the appendix 1 and 2.

As soon as the permission from parents was received, the process continued by selecting sample for the research. Stratified sampling method was utilized to meet the needs for research participants. The next step was distributing questionnaires to 253 students from different levels (96 7th graders; 86 8th graders; 68 9th graders).

Focused-group discussion was then held with each grade on Thursday, 10th October 2019 discussing the topics stated on the FGD guidance attended by the total of 30 students (each grade was represented by 10 students selected from higher, mediocre and lower achievers) to gain more diverse perspective. The discussion was recorded using smartphone voice recorder, Redmi 5. It was divided into three sessions. The first session was administered to 10 eighth graders; the second session was attended by 10 ninth graders; and the third session was conducted to 10 seventh graders. Each session took about 30 minutes to complete.

The research process was carried on by executing observation toward 96 students from different levels (32 7th grade; 32 8th grade; and 32 9th grade) from the total population of 842 students. It was undertaken on Wednesday, 18th of September 2019 during mid-semester English test schedule 2019/2020 academic year. It was conducted in two sessions following the school mid-test time allocation. The first session started at 9.20 to 10.40 a.m. and the second one was at 13.00 to 14.20 p.m. 8th and 9th graders were observed simultaneously at 9.20 – 10.40 a.m. 7th graders were observed at 13.00 to 14.20 p.m. The researcher utilized two video recorders on two smartphones which were held by tripods and conducted note

taking to complement observational data gotten from the video recorder. The researcher as the first observer went back and forth to take notes while the second observers watched over the students who were taking the English test. The results of observation were used as the triangulation technique to complement the FGD results.

Data collection techniques were then concluded by comparing the results of English test using Edubox with the ones the students got from their Paper-Based Testing (PBT) evaluation. This effort was aimed at measuring the students' English learning motivation.

E. Data Analysis Method

This study applied two data analyses methods, qualitative data analysis for Focused-group discussion (FGD) and quantitative data analysis for English test and questionnaires.

a. Qualitative data analysis

According to Spradley (1979, p. 92), analysis of any kind involves a way of thinking. He noted, —It refers to systematic examination of something to determine its parts, the relationship among parts and the relationship to the whole. This kind of method allows better interpretation of the data. Bogdan and Biklen (1998, p. 157) similarly noted that: —Analysis involves working with data, organizing them, breaking them into manageable units, synthesizing them, searching for patterns,

discovering what is important and what is to be learned, and deciding what you will tell others.

For this particular study, Domain Analysis Method was selected to analyze the data gotten from focused-group discussion (FGD) activity. Spradley (1979) elaborated the function of this data analyses means. He noted that it is about the process of identifying, collecting, organizing, and representing the relevant information in a domain. According to Spradley, the first and second elements in the structure of a domain are cover terms (larger categories) and included terms (instances of the cover term). The next feature is semantic relationships. Spradley explains that, —When two included terms are linked together, we refer to this link as a semantic relationship (1979, p. 100). Some of the most useful semantic relationships he proposes are:

1. Strict inclusion where X is a kind of Y.
2. Cause -effect where X is a result of Y.
3. Rationale where X is a reason for doing Y.
4. Location for action where X is the place for doing Y.
5. Sequence where X is a step stage in Y.
6. Attribution where X is an attribute (characteristics) of Y.
7. Means-end where X is a way to do Y.
8. Spatial where X is a place in Y, X is a part of Y.
9. Function where X is used for Y. (Spradley, 1979, p. 111)

Spradley (1979, p. 95) summarized that in general, the intention of doing ethnographic analyses is, —to find out how do words and behavior and objects become meaningful and how do we find out what things mean?

This type of analysis may include examples such as the following:

- Where X is a kind of Y
  - Edubox is a kind of local assessment network (LAN).
- Cause-effect where X is a result of Y
  - Students' complain is the result of network problem.
- Where X is a reason for doing Y
  - Testing with Edubox is a reason for faster test result feedback

Students' perception towards CBT- Edubox and its impact to their English learning motivation are the reason for this thesis research.

In brief, according to Spradley, domain analyses consist of six interrelated steps:

- Selecting a single semantic relationship.
- Preparing a domain analysis worksheet.
- Selecting a sample of informant statements.
- Searching for possible cover terms and included terms that appropriately fit the semantic relationship.
- Formulating structural questions for each domain.

For example:

— Are there different kinds of students' perceptions about Edubox?

- Making a list of all hypothesized domain. (Spradley, 1979, p. 117-118)

Below is the example of analyses conducted using domain analyses:

Table 3.6
Semantic Domain Analyses

Domain	Semantic relations	Structural question
Edubox	X is a kind of Y	Are there different kinds of students' perception about Edubox?
Students' complain	X is part of Y	What are all the instances of students complain about Edubox? and so forth

Field notes were taken during the observation period as the other endeavor to triangulate the information gotten from the questionnaires and FGD. This was expected to supply additional data concerning factual situation during the Edubox testing period from the researcher's perspective. The information about the results of FGD and observation will be discussed in more detail in the discussion of the results section.

b. Quantitative data analysis

Questionnaires and English test results were analyzed by means of inferential statistics, t-test and correlational coefficient conducted by SPSS consultant. t-test, in statistics, is a method of testing hypotheses about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown. While The 'correlation coefficient' was coined by Pearson in 1896. Accordingly, this statistical method is over a century old, and is still going strong in May 18th, 2009.

It is the most widely used of many chi-squared tests (e.g., Yates, likelihood ratio, portmanteau test in time series, etc.) – statistical procedures whose results are

evaluated by reference to the chi-squared distribution. Its properties were first investigated by Pearson in 1900.

t-test was used to determine whether there was a significant difference between English test scores using Edubox and the ones using PBT. Inferential statistics was expected to allow assumption making about the dependent variables in this research namely perception and motivation

Cronbach Correlational coefficient test was employed to show the direction and strength of the relationship between students' perception toward Edubox and their English learning motivation is. In this study, there is a strong negative relationship between students' perception toward computer-based testing Edubox and their learning motivation. It will be elaborated more in the next chapter.

The main result of a correlation is called the correlation coefficient (or "r"). It ranges from -1.0 to +1.0. The closer r is to +1 or -1, the more closely the two variables are related. If r is close to 0, it means there is no relationship between the variables.

A positive correlation coefficient means that as the value of one variable increases, the value of the other variable increases; as one decreases the other decreases. A negative correlation coefficient indicates that as one variable increases, the other decreases, and vice-versa. The detailed information about the results of the questionnaires and the English test will be discussed in the discussion chapter.