CHAPTER I
INTRODUCTION

1.1 Background of the Research

Some of today's teachers in Indonesia seem to lack mastery in composing assessments. Mili (2020, p. 145) noted that many teachers still have difficulty planning and conducting formative assessments. In line with that, the Ministry of Education and Culture (Kemendikbud, 2017), 6.84 percent of teachers in Indonesia's public and private junior high schools are not yet qualified to teach (p. 98). It cannot be ignored since it impacts how they design assessment items, particularly assessments based on higher-order thinking skills. Higher-order thinking is an important transferable skill to develop in the teaching and learning process at school (Suryani, Sapriya, Malihah, & Komalasari., 2019, p. 1). Moreover, higher-order thinking skills are the 21st-century skills requirements as Pratiwi, Dewi & Paramartha (2019) noted that the 21st-century era introduces four skills that must reflect in school (1) Critical Thinking and Problem-Solving; (2) Creativity and Innovation; (3) Communication; and (4) Collaboration (p. 128) these four skills are part of higher-order thinking skills (HOTS). Hence, assessments are required to determine whether they have achieved HOTS during their learning process.

Despite the undeniable excellence of thinking skills in this modern world, it seems that cultivating higher-order thinking skills is regardless of the various educational level (Hurley & Hurley, 2013., cited in Jahedizadeh, Ghanizadeh, & Al-Hoorie., 2020, p. 2). According to Smith and Szymanski (2013, in Jahedizadeh et al., 2020), in the search for more outstanding test scores based on rote learning and recitation, many students exit high school without the critical thinking abilities required for success in further education (p. 2).

Nevertheless, the current assessment system design focuses on measuring students' ability to remember facts using multiple-choice tests instead of adequately measuring students' ability to engage in and complete complex thinking and problem-solving tasks (Winaryati, 2017, p. 62). Consequently, there are gaps

between the knowledge & students' skills acquired in school and the 21st-century knowledge & students' skill required at the workplace. Assessment is an evaluation in the teaching and learning process since evaluation is essential in measuring the teaching and learning progress. However, most English teachers have trouble creating and developing higher-order thinking (Lestari, Bharati, & Rukmini 2018, p. 3). There are two kinds of assessment: formative and summative assessment. Formative assessment is an ongoing assessment; meanwhile, summative assessment is conducted at the end of the learning period (e.g., semester, midterm.). Furthermore, the researcher used formative assessment as the research variable development to assist the teacher in constructing them. Managing the assessment item requires a database that contains item questions; this term is called a question bank.

Based on Wright & Bell in Widana (2014), the question bank is a coordinated, developed, defined, and quantified query composition to provide an operational definition (p. 186). Since the item question is quantified thus, the item question is verified, validated, and feasible for taking the formative assessment. However, not all teachers utilize a question bank in compiling constructive assessment questions as assessment instruments. Meanwhile, the question bank contains a collection of questions and crucial related information; thus, what is used from a question bank are calibrated question items. Therefore, the item tests and the accompanying information are for consideration for the assessment purpose. After the formative assessment is constructed and the item is calibrated and compatible with the assessment purpose, the issue arising is the consideration of media for delivering the question. There are platforms for providing test items, printed media (Paper-pencil based test), or computerized media (Computer-based test).

Developing technology emphasizes that educators must fully employ technology in teaching and learning processes, including assessment. Teachers should consider improving their digital literacy to answer the digital era. Irawati & Fitriati (2018) found that one of the common challenges test-takers encountered was the students' lack of interest in responding to questions. The questions seem monotonous and less interactive, affecting stagnant and low student learning

outcomes. Thus, developing formative assessments using an interactive assessment platform is required. There are specific applications and software assessment tools, such as *Quizizz*, *Kahoot*, and *Quipper* (Amalia, 2020, p. 2). Rahayu & Purnawarman (2019) noted that *Quizizz* is a fantastic online evaluation tool to measure students' knowledge and progress in learning English. Teachers can use *Quizizz* to deliver assessments and provide homework to students as additional practice (p. 103).

However, to answer those issues mentioned above, the researcher proposes a research developing question bank as a formative assessment for the eighth grade of junior high school. The researcher will develop question bank-oriented higher-order thinking skills for formative assessment, and the question bank will be displayed in the *quizizz* application. Moreover, several studies that discuss assessment and question bank development have been conducted. These clusters relate to the research topic: **Question Banks and Computer-Based Test** (Naimah. 2015; Parthasarathy & Ananthasayanam, 2016; Glamočić, Mešić, Neumann, Sušac, Boone, Aviani, Hasović, Erceg, Repnik, and Grubelnik, 2021; Koşan, Koç, Elhan, & Öztuna, 2019; Yasa, Suastika, & Zubaidah., 2020; Dilla. 2019), **Item Bank Development** (Furtuna, 2013; Hancock, Hare, Denny, & Denyer., 2018; Fadilah, Kudsiah, Karlinda, Haifaturrahmah., 2021; Safrina, Yustiana, & Afandi., 2020), **Assessment Development** (Irawati & Fitriati., 2018; Lestari et al., 2018; Cayani., 2021; Fanani., 2018; Widana., 2017; Cagasan, Care, Robertson, & Luo., 2020).

The researcher found in the above mentioned clusters four missed points that should be completed in this study. First, the recent studies did not tie 21st-century skills as a significant research issue. Meanwhile, students require today's skills to answer the challenges in the future. As Murugiah (2020) noted, in sustainable development goal 4.7, human resources should master specific skills in 2030; thus, education sectors should ensure the students obtain the knowledge and skills required to support sustainable development (p. 42). A revolution in educational sectors is essential for sustainable growth to achieve 21st-century collaboration, problem-solving, critical thinking, and communication

competencies. Second, the studies mentioned above did not specifically discuss the characteristics of question bank, item bank, and assessment in the 21st century. On the other hand, test developers and question bank makers required updated sources on recent test development/question bank development. Third, the researcher did not find the cluster above the English subject question bank for eighth junior high school. Fourth, the researcher found certain studies focusing on particular skills and topics. Meanwhile, a good item bank or question bank contains various skills and covers several issues.

The point missing mentioned above are essential to consider in the following studies. The researcher should complete those gaps in this research. The researcher will consider 21st-century skills as the central issue in the studies, particularly higher-order thinking skills as part of 21st-century skills. Moreover, the researcher will investigate and analyze the currently published question banks to discover the weaknesses and shortcomings. The researcher will be developing English for the eighth junior high school question banks based on the item analysis results.

1.2 Identification of the Issues

The researcher found some difficulties in conducting formative assessments. The questions are from the students' handbook, and it does not stimulate students' critical thinking. Hence, based on the study background above, the identified problems are

1. Some of today's teachers in Indonesia lack mastery in composing assessments.
2. Today's teaching and learning process must reflect higher-order thinking skills (HOTS).
3. The current assessment system design is obliged to measure students' 21st-century skills.
4. Not all teachers use question banks to assist them in conducting the formative assessment.
5. Developing technology emphasizes that educators must fully employ technology in teaching and learning processes, including learning assessment.

6. Students lack interest in answering the question due to the assessment delivery not using an attractive platform.

1.3 Delimitations and Focus of the Study

The research was limited to investigating shortcomings of the current product of question bank and designing the preliminary form of question bank as a formative assessment. The question bank is in the form of multiple choices and covers the first basic competency (*3.1* and *4.1*) in the first semester of eighth-grade junior high school. The product covered four topics asking for attention, showing appreciation, checking someone's understanding, and asking &giving an opinion. The question bank is used to assist the teacher in taking formal formative assessments such as quizzes, practices, etc. The researcher displayed question items in the *quizizz* application as the instrument tool. Moreover, the researcher employs three steps in developing a question bank: reviewing, planning, and developing. The product from the question bank development is still a prototype since the product has not been tested in the field and validated by an expert.

1.4 Research Questions

1. What are the shortcomings of question banks as formative assessments in two published products?
2. How is the preliminary form of question bank as a formative assessment for the eighth grade of junior high school?

1.5 Aims of the Research

1. To investigate the shortcomings of question banks as formative assessments in two published products.
2. To design a preliminary form of question bank as a formative assessment for the eighth grade of junior high school.

1.6 Significances of the Research

In every study, the significance of the research seems to be a crucial thing to reveal. In this research, the importance of this study is as follows.

1. Theoretically, this research's result is for another researcher's reference to develop another question bank, assessment, or topic related to the ELT.

2. Practically: This research aims to help teachers take formative assessments as their teaching evaluations straightforwardly. Also, the results of this research are to motivate teachers to develop the question by themselves. Moreover, the result is to help students train their higher-order thinking skills as a student requires. Also, the development of this research is to find their HOTS to evaluate their learning process.

1.7 Theoretical Foundation

1.7.1 Assessment

1.7.1.1 Definition of Assessment

Assessment measures the students' performance, skill, and knowledge. Thus assessment must be conducted as a learning evaluation. Khairil & Mokshein (2018) defined *assessment* as a component of the teaching and learning process that tries to help both the assessor and the person being assessed improve (p. 662). In other words, assessment results can be a barometer for teachers and students to evaluate teaching and learning activities. As Black and Wiliam (2009 in Cagasan et al., 2020) asserted, assessment is all activities implemented by teachers that contain information and reports used as evaluations of teaching and learning activities (p. 1).

Meanwhile, Shaw (2015) said that the act of appraisal or judgment is often referred to as assessment (p. 2). Moreover, in employing an assessment, the teacher should consider the relevance of the employed curriculum. In line with that, Lestari et al. (2018, p. 499) described that the assessment must be relevant and appropriate to the curriculum's learning objectives being employed.

To summarize, assessment is a part of the teaching and learning process. When teaching, the teacher finds out and gathers evidence and data on how the student understands the material. An assessment may include a test, but it may also include techniques such as observation, interviewing, observing behaviours, and monitoring. One sort of evaluation is the test. Not all tests are assessments, but all assessments are tests.

1.7.1.2 Kinds of Assessment

In general, assessment can be distinguished by the assessment period. Formative and summative assessments are the two types of assessments (Khairil and Mokshein, 2018).

1.7.1.2.1 Summative Assessment

Summative assessment is a crucial tool in learning to determine the students' outcomes in the learning process. The researcher provided several definitions based on certain studies. As Mosquera & Macías (2015) said, Summative assessment is a form of evaluation that provides teachers with numerical feedback on their students' progress in the learning process (p. 1). Furthermore, Ahmed, Ali, & Shah (2019) noted that summative assessment is the process of keeping track of pupils' accomplishments up to a certain point on a numerical scale in order to look back and assess how students met their goals (p. 111). To summarize, summative assessment is a type of assessment that informs teachers of their students' success that translates all of the pieces of evidence to a particular point. Hence, the goal of summative assessment is to document or report learners' progress.

1.7.1.2.2 Formative Assessment

The critical point of formative assessment is *"ongoing,"* also known as "ongoing assessment," based on how well students complete learning assignments and include them to improve performance. Furthermore, Yorke (2003) stated that the essential concept of formative assessment contributes to students' learning by giving feedback about performance (p. 478).

In contrast, the formative assessment indicated evidence of teaching, whereas the summative evaluation demonstrated learning over a period. Formative evaluation feedback can help students stay engaged in class, enhance their learning performance, and increase their enthusiasm to study (Khairil and Mokshein, 2018:663).

1.7.1.3 Formative Assessment

Formative assessments are crucially required for both teacher and learner to measure the learning progress. The assessment feedback holds principal information to modify teaching and learning measurement in the following activity. Frey & Fisher (2011. p. 2) stated that feedback could effectively raise student success when used as a formative assessment system. Black & Wiliam (2009) asserted that formative assessment is when inferences about a student's learning activity are elicited, inferred, and used by teachers or students to make decisions for the next moves in the following teaching and learning activity. The inferences might become likely better or perhaps more well-founded than the decisions they would have made in the absence of the findings (p. 7). Leenknecht, Wijnia, Kohlen, Fryer, Rikers, & Loyens (2020) point out that formative assessment is a cyclical program of high and low-stake activities. Students are actively engaged (as assessee and assessor). When assessing assessment as formative, practice the role of teachers', students', and their peers' involvement is addressed, and the assessment's developmental significance is emphasized (p. 2).

Nevertheless, formative assessment is generally classified into two types formal and informal. As Yorke (2003) pointed out, a formative assessment might take the shape of a formal or informal evaluation. Formal formative assessment is anything done in the curriculum context and consists of assessment activities embedded in the curriculum; students are expected to complete a task. The instructor is required to assess and provide feedback to the student. Informal formative assessments occur naturally during events but are not explicitly stated in the curriculum design. Some of these are instantaneous feedback when the student is engaged in a learning activity and comments on drafts of work for inclusion in portfolios (pp. 478-479).

1.7.1.4 Formative Assessment in The 21st Century

Assessment (formative or summative) in the 21st century must transform from discrete measures of knowledge to measuring students' abilities to think critically, solve problems, gather information, and make reasoned decisions using

technology (Winaryati, 2017, p. 163). Therefore, assessment should focus on students' operational skills, such as using multiple sources appropriately and efficiently, not on whether students have responded correctly to the teacher's words. Mitana, Mugagga, & Ssempala (2018) stated that in terms of assessment, higher-order thinking skills are classified according to their level of complexity using the taxonomy of thinking. Similarly, evaluation questions can be categorized based on their difficulty level (p. 243).

Winaryati (2017) listed various characteristics of practical formative assessment for measuring 21st-century skills:

a) Focus on the content and skills of the 21st century.
b) Generate thoughts that reveal students' conceptual strategies used to solve problems.
c) Constructing assessment is structured to identify students' knowledge and background to solve each problem in real-time.
d) Primarily performance-based and authentic, asking students to use 21st-century skills.
e) Create data that can be used to inform instructional practice directly.
f) It aims to build capacity - both teachers and students.
g) Make a comprehensive assessment series.
h) Reflecting the multidimensional, integrated, and revealed understanding of learning in performance over time. (pp. 170-172)

1.7.1.5 HOTS-Based Assessment

In line with the definition mentioned above, formative assessment holds information about students' attainments and is used by teachers to evaluate the following teaching and learning; thus, formative assessment should reflect the current curriculum. The 2013 curriculum introduced higher-order thinking skills (HOTS) throughout all levels of education to improve students' comprehension and critical thinking (Lestari et al., 2018, p. 499). Higher-order thinking is an important transferable skill to develop in the teaching and learning process at school (Suryani et al., 2019, p. 1). Critical thinking and creativity, considered higher-order thinking

skills, are essential skills to master to equip students for 21st-century challenges. (Lestari et al., 2018, p. 500).

Moreover, Anderson & Krathwohl (2001) in Setiawan, Sudrajat, Aman, & Kumalasari (2021, p. 545) said that HOTS is a high-level thinking ability is theoretically founded on Bloom's Taxonomy. Theresia (2020, p. 10) noted six dimensions in thinking skills based on bloom taxonomy revised by Anderson & Krathwohl (2001):

1) Remembering (*C1*)
2) Understanding (*C2*)
3) Applying (*C3*)
4) Analyzing (*C4*)
5) Evaluating (*C5*)
6) Creating (*C6*)

Based on Bloom's taxonomy above, six dimensions of thinking skills are separated into two categories according to level thinking. Three dimensions analyzing (*C4*), evaluating (*C5*), and creating (*C6*), are considered higher-order thinking skills (HOTS). Whereas the three other dimensions remembering (*C1*), understanding (*C2*), and applying (*C3*), are categorized as lower-order thinking skills (LOTS) (Theresia, 2020, pp. 14-15., Diana, Latifah, Gunawan, & Anggriani., 2020, p. 1).

Furthermore, the Ministry of Education and Culture, through the education assessment center team (puspendik) (2019, pp. 6-7), introduced the principles of developing higher-order thinking skills assessment instruments:

1) Using stimulus

   The stimulus can be in texts, pictures, scenarios, tables, graphics, discourses, dialogues, videos, or problems. Stimulus serves as a medium for students to think; questions tend to ask or assess memory without a stimulus.

2) Using the new context

   The new context here means the context of the question as a whole, which can be in the form of material or problem formulation, to function as

a tool that measures higher-order thinking; questions should not be answered solely by relying on memory.

3) Distinguishing the level of difficulty and complexity of the thought process

Measuring Higher order thinking skills should distinguish the thought process's difficulty and complexity within the item questions.

1.7.2 Question Bank

1.7.2.1 Definition of Question Bank

The question bank can be both learning media and evaluation. Banking the number of questions is helpful for teachers to take an examination straightforwardly. As Hayati & Mardapi (2014) noted, the question bank is a collection of test items that refer to making questions that can make it easier for teachers to compile and take questions as evaluations every time they take an exam (p. 29). Meanwhile, Lane, Raymond, Haladyna, & Downing (2016) noted that an item bank is a database for storing, accessing, and tracking test items and associated data securely. In line with that Sumardyono & Wiworo (2011), a question bank is a set of calibrated (tested) items both theoretically and empirically and contains essential information so that it can be easily used in the preparation of an assessment instrument (test) (p. 9).

Moreover, the question bank is helpful for the teachers to take an examination. The question bank also can assist students in learning evaluation. Higher education institutions use question banks to enhance access to high-quality material for grading students (Chilivumbo, 2015). Besides question banks used in higher education, question banks are also used at every school institution since question banks are proven to have advantages, as stated beforehand.

Furthermore, in the process of making a question bank, considering the question item based on its characteristics is crucial to do. As Naimah (2015) defined, a question bank is a collection of organized and catalogued questions to achieve a specific amount based on the content and characteristics of the items (p. 7). The word *"organized"* in the statement above is a slightly important point of

question bank since a question bank is a result or product of organizing the number of questions.

To summarize, a question bank is a collection of calibrated items organized and itemized based on a particular characteristic and content to assist teachers and students in learning evaluation and assessment.

1.7.2.2 Characteristics of Question Bank

The researcher found some theories about the characteristics of question banks in the 21st century. The first characteristics are provided by Chilivumbo (2015) stated that a good question bank should:

a) Integrated with the learning-oriented assessment.
b) Facilitate the learning-oriented assessment activities.
c) Create evaluation competence.
d) Aid in student engagement by providing feedback. (p.23)

Those characteristics represent integrating a question bank with the teaching and learning process. Thus test developers should analyze them before assembling the question banks to reach the validity of the test. However, the question bank should integrate with the teaching and learning process. The item question within question banks should represent some points. As king (2005, cited in Chilivumbo, 2015) noted, a good question bank should be able to test knowledge in *factual*, *conceptual*, *procedural*, and *metacognitive areas* (p. 26).

On the other hand, Sumardyono & Wiworo (2011) displays that the characteristics of an item bank used for tailored testing are essential to the efficiency and accuracy of the process. A good item bank has three primary characteristics banks based on Sumardyono & Wiworo (2011):

a) Each item in the question bank is calibrated.
b) Each item is equipped with a variety of helpful information. (e.g., based competencies, answer key, score).
c) The database (questions item) in a question bank is structured. (p. 10)

1.7.2.3 Kinds of Question Banks

A crucial design of question bank consideration is to select a delivery platform. Since technology influences all education sectors, including assessment, the most widely used assessment method is the paper and pencil test (Retnawati, 2015. p. 135). Nowadays, the development of technology and the digitalization of educational terms demand teachers to employ computer-based tests in conducting assessments. Furthermore, access to information inside and outside the school has become easier as the Internet and intranet networks have grown popular. This technology can also be used for other reasons, such as computer-based testing, a type of examination that uses a computer, and the Internet (CBT). There are basic types of item banks as Wendler & Walker (2016) provided two choices platforms in delivering question banks. The most basic choice is between paper-and-pencil testing (PPT) and computer-based testing (CBT) (p. 439). There are some advantages and disadvantages in both platforms in delivering test items:

1.7.2.3.1 Paper and Pencil Testing (PPT)

The advantages and disadvantages of the computer-based test were explained by Wendler & Walker (2016). Paper Pencil Testing (PPT) has various benefits, including

- The accessibility of use.
- Familiarity.
- Inexpensive production costs.
- The capacity to test large groups of people at the same time.

Meanwhile, Paper Pencil Testing (PPT) disadvantages are:

- Interface drawbacks, such as graphics not allowing dynamic presentation of test content.
- User population, paper-based testing does not allow more various tests to be placed.
- Standardization of the test environment, i.e., tests are not presented in the same way and format for a given time and user population.
- Paper-pencil test, giving feedback takes time due to manual correction.

- Paper-pencil tests sometimes deal with human error in the preparation of questions.

1.7.2.3.2 Computer-Based Test (CBT)

The advantages and disadvantages of the computer-based test were displayed by Noyes & Garland (2008). The benefits of online assessments or computer-based tests are

- The wealth of the interface, for example, graphics, allows for a dynamic presentation of the test content.
- For the user population, computer-based testing via the internet allows for a more diverse test to be placed.
- The standardization of the test environment, that is, the test is displayed in the same way and the same format for a specified time and the user population.
- Online scoring leads to speedier feedback.
- Enhanced (i.e., Fewer human mistakes). (pp. 1362-1369)

On the other hand, the disadvantages of using online assessment or computer-based tests (CBT) are

- Lack of a controlled environment, with responses coming in at various times and locations, sometimes not even by the assigned individual; multiple submissions may also be an issue.
- Computer hardware and software are subject to freezing and crashing; time can be wasted in the test set when systems must be restarted or updated.
- The computer screen may be more exhausting for lengthier tests.
- With computer and paper presentations, equivalency is challenging to achieve. (p. 1369)

1.7.2.4 Kinds of Question Item

1.7.2.4.1 Constructed Response(CR)

On the other hand, some examinations include questions that require the test taker to generate the answer rather than selecting it from a list. Constructed response

items generally demand the examinee to respond to a stimulus in writing, usually a question or a statement (Downing, 2009. p. 2). The needed response can range from as simple as penning a single word to as complex as designing a laboratory experiment to test a scientific hypothesis. These test types are called the constructed response (CR) (Livingston, 2009. p. 1). Constructed response items are more appropriate when a written answer is required, effective solutions are requested, detailed process information is required through synthesizing, organizing, and sequencing data or justification or explanations are necessary (Rodriguez. 2016, pp. 260-261).

1.7.2.4.2 Selected Response (SR)

However, the selected response (SR) items contrast with the constructed response (CR). The constructed response requires the test taker's response to a stimulus in writing, which is usually a question or a statement. Meanwhile, selected-response (SR) items are employed in the selected item. Downing (2009) noted that the selected-response (SR) items require examinees to choose a correct or best answer from a fixed listing of possible answers to a question or other stimuli (p. 2). Furthermore, the numerous forms for selected-response items are well known. However, the item writer also has access to several less well-known formats, many of which have not been empirically verified (Rodriguez. 2016, p. 264). However, all SR items are suggested for classroom usage since learners should be exposed to a range of item kinds, mainly since each type may engage various cognitive skills. There are many a current taxonomy of SR item formats based on Rodriguez (2016, pp. 264-266):

1) Multiple Choice (MC).
2) Fill in the Blank MC format.
3) True-False.
4) Matching.

1.7.2.5 Multiple Choice Question

Multiple-choice questions (MCQ) are a typical approach for assessing student learning. They can be utilized as a tool for assessment (Santos, Hu, and

Jordan., 2014., p. 71). Furthermore, the multiple-choice question (MCQ) format is an efficient and objective technique that often comprises a significant component of the overall assessment in an important junior science course. (Hancock et al., 2018. p. 1). Multiple-choice questions have been used in many assessments since they have many advantages. Moreover, Perie & Huff (2015, p. 122) demonstrated that multiple-choice questions could be used for large-scale educational evaluations to examine students' problem-solving and critical thinking skills.

To sum up, the multiple-choice questions item is one in which the respondent is given a restricted number of options from which to choose one or more like a typical approach for assessing student learning and can be utilized as a tool for assessment than can be used for large-scale educational evaluations in which can examine students' problem solving and critical thinking skills.

1.7.2.6 Characteristics of Question Item in the Question Bank (Multiple Choice Format)

Although the multiple-choice question is commonly used, not all published multiple-choice items are standard. Rodriguez (2016, p. 264) noted characteristics of multiple-choice items: a) provide a stem that contains completed questions or partial statements, which have often been included in item writing guidelines. b) Provide one or more correct response options. c) Provide one or more plausible incorrect response options (distractor). d) Provide instructional statements. Furthermore, statistical requirements must be specified before determining acceptable test content. Based on Wendler & Walker. (2016) noted that these standards define the test's technical requirements and ensure equivalence between different versions. The overall test difficulty should be determined, and the target distribution of item difficulties and item discrimination indices (p. 437).

1.7.2.6.1 Item Difficulties

Based on Hartati & Yogi (2019), a good test should offer a range of difficulty levels, including easy, moderate, and challenging. (p. 61). Thus, item difficulty should be denoted by the item developer to assist the teacher in taking the test item. Moreover, an item's difficulty index (DIF I) is the proportion or

percentage of examinees that appropriately answered a test item. (Sharma, 2021, p. 18).

However, too easy problems do not stimulate students to increase their efforts to solve them. On the other hand, too difficult questions will cause students to despair and not be enthusiastic to try again because they are beyond their reach. Furthermore, more items at the level of difficulty corresponding to the position on the reporting scale where decisions are made are included in the desired distribution of item difficulties, representing the test's objective. (Wendler & Walker. 2016. p. 437). In line with multiple question choices in an item bank, Smith & Dickinson (2017) stated that questions in published multiple-choice question banks are commonly classified into three levels of difficulty: hard, medium, and easy (p. 10).

1.7.2.6.2 Item Discrimination

Item discrimination is the quality of an item based on which the superior (the group of high achievers) and inferior (the group of low achievers) are distinguished. (Sharma, 2021, p. 18). Moreover, item discrimination indices must also be addressed at the design stage. Item discrimination indices reflect the correlation between item scores and total test scores, allowing examinees of varying skill levels to be distinguished. (Wendler & Walker. 2016. p. 437).

Nevertheless, to make question items reflect a wide range of difficulty and discrimination items, a term distractor makes the test taker perplexed and hard to choose between the options. Haladyna (2016, p. 401) stated that a proper set of discriminating distractors is required for a highly discriminating item. The discrimination of a test item can be improved by removing or enhancing ineffective distractors. Item developers frequently implement difficulty by identifying three or four plausible distractors inside a similar item (Rodriguez, 2016, p. 268). As mentioned earlier, distractors or incorrect options should be provided in an item since distractors play a significant role in determining the item's properties. For instance, a distractor is a fundamental component of the multiple-choice items' format that includes incorrect (oils or misleads) options (Smith & Dickinson, 2017, p. 10). Furthermore, the distractor should be well-constructed; besides implemented

implausibility, the distractor should reflect common errors or misconceptions, providing diagnostic information within a single item that is practically identical (Rodriguez, 2016, pp. 267-268)

1.7.2.6.3 Item Writing Guidelines

In developing a question bank, the item developer should follow the guideline as item indicators. As Wise & Plake (2016, p. 28) pointed out, the standard in constructing item assessment offered specific guidance based on item specifications. Indonesian ministry of education through Puspendik (2019, pp. 13-14, and Widana (2014, p. 192) listed sixteen points, and Rahmadhani, Budiono, & Suparno (2014, p. 426) nineteen items writing guidelines that separated into three aspects: content, construction, and language. Moreover, Haladyna (2004, pp. 99-100) provided twenty-nine points divided into four elements of item writing guidelines: content guidelines, Style and format concern, writing of the items, and writing options. Meanwhile, Yanti, Paramartha & Wahyuni (2020, p. 40) listed eighteen points of item writing guidelines without dividing them into some aspects.

In conclusion, there are twenty-six points of item writing guidelines as shown in *appendix 1*, divided into three aspects, content & material, construction, and language & grammatical, that have been merged and eliminated some points from those theories. Moreover, in aspect content and material, there are seven indicators in terms of basic competencies, level of cognitive, and other indicators related to content and materials. Whereas in aspect constructions, there are fifteen indicators in terms of item structure, whether in stimulus, stems or options. Meanwhile, language and grammatical aspects have four indicators: grammar, vocabulary, phrases, punctuation and capitalization.

1.7.3 Junior High School

1.7.3.1 Definition of Junior High School

Students' leveling education is primarily used in every educational system globally, particularly in Indonesia. Junior High School, abbreviated as SMP, is a basic level of formal education in Indonesia after graduating from elementary school (or its equivalent). Junior high school is taken in 3 years, starting from grade

7 to grade 9 (Suradi, 2017, p. 525). Junior high school graduates can continue their education to senior high school or vocational high school (or equivalent). Junior high school students are generally aged 13-15 years. Meanwhile, Lewis (2007, in Arda & Doyran, 2017) points out that eleven to nineteen students are called teenage or young adolescent learners. Thus, the eighth grade of junior high school is categorized as teenage or young adolescent learners (p. 182).

However, junior high school level is categorized as basic education levels as Sa'ud & Sumantri (2007, p. 5) stated that primary education is general education which should be taken for nine years and is held for six years at Elementary School/Madrasah Ibtidaiyah (SD/MI) and three years at Junior High School/Madrasah Tsanawiyah (SMP/Mts) or an equivalent educational unit. Moreover, Rozano (2010) states that primary education is the initial education for children, whether formal or informal, starting from 3 years to at least 12-15 years of age (p. 52). In other words, the junior high school level is a base level to prepare students for future life and higher education.

1.7.3.2 Characteristics of Junior High School

Students in junior high school have distinct characteristics that distinguish them from younger and older students. Since junior high school students are experiencing the transition phase from children to teenagers, they should comprehend their characteristics to reach teaching and learning outcomes. As Gibbons (2002, p. 7) noted, teenagers or adolescents are transitioning from childhood to adulthood, demanding them be more self-sufficient, responsible, competent, and optimistic. Meanwhile, Arda & Doyran (2017) demonstrated adolescence as a beautiful set of people interested in learning, energetic, curious, ambitious, and friendly (p. 182).

Hence, there are challenges for teenagers learning in instructional activities; thus, the teacher has a great job managing them. In line with that, Arda & Doyran (2017) teenagers are unsatisfied with practically everything; finding fascinating and appealing topics and activities is complicated. Consequently, the instructor has a fantastic job bringing spectacular activities into the classroom (p. 183).

Furthermore, Hidayah, N. (2007) noted that one of the fundamental challenges in teens' search for distinct identity is generally acknowledged. They enjoy challenges, peer approval, and being forced among peers and friends. The development of specific psychological and physiological features in children aged 11 to 14 years necessitates a set of educational settings in the classroom (p. 17).

To sum up, with the unstable character of junior high school students. The teachers should assist students in engaging and motivating students in learning; thus, teachers should know their character for the treatment preparations since junior high school should prepare through learning activities to face 21st-century challenges in the future. The researcher offers question banks development to examine students' higher-order thinking skills. Besides that, the teacher knows the students' thinking skills through question bank developments.

1.7.3.3 English At 8th Junior High School

English has grown recently, and people understand the need and importance of learning English. Some have been learning English since kindergarten, while others have only recently begun. People occasionally require additional time to learn English, as though studying English for years in school is insufficient. Indonesia does not have a native language towards English; thus, Indonesian students are categorized as English as a foreign language (EFL). Furthermore, Indonesian education recently employed the 2013 curriculum (K-13). English is one of the subjects taught in the 2013 curriculum, whether it is mandated or only for local content. Moreover, there are sixteen basic competencies in English in the second semester of the eighth grade of junior high school that should be reflected in teaching & learning, materials, and assessment. In other words, question banks as formative assessments demanded to examine all the skills in the 2013 curriculum implemented on the syllabus.

1.7.4 Question Bank as Formative Assessment for 8th Junior High school

Despite generating a question bank is very challenging and requires complex steps. Nevertheless, banking the number of questions is beneficial for the teacher since the teachers can take assessments easily by compiling and taking item

tests from the item banks. Hayati and Mardapi (2014) said that the question bank is a collection of test items that refer to making questions that can make it easier for teachers to compile and take questions as evaluations every time they take an exam (p. 29). Therefore, the teachers can implement formative assessments with questions from the item bank.

Formative assessment can be conducted accidentally to measure the performances of the teaching and learning process and evaluate them for the next meeting. In the formal formative assessment, the teacher takes quizzes regularly. Thus, the teacher requires question banks to assist them in arranging item tests.

1.8 Previous Studies

This chapter displayed various previous studies that are similar to this study. Three studies identified similarities with this study: Fadilah et al (2021) Developing thematic question bank based on higher-order thinking skills (HOTS) at elementary school, Safrina et al (2020) Developing card-formed question bank hots oriented in Elementary school, Dilla (2019) Development of an android application-based question bank in the subject of vegetable product processing production to find out the learning outcomes of students at SMK PP Negeri Lembang.

The first study is conducted by Fadilah, Kudsiah, Karlinda, and Haifaturrahmah. (2021) Developing thematic question bank based on higher-order thinking skills (HOTS) at elementary school. This research attempts to find out how to develop a question bank based on HOTS at elementary schools and find out the result of question bank development based on HOTS. The study used the Thiagarajan research model as the research method, and the steps are: define, design, develop, and disseminate. The result found that there are 90 questions from 100 developed questions that are feasible to be included in the question bank, divided into three packages; each package consists of 30 questions. However, the researcher found that Fadilah et al did not mention and described how theory and item writing indicators were gained. Also, the mentioned study tends to only create new products without correlating with published products. Fadilah et al. study aims to develop and analyze the result of a thematic question bank at elementary school.

Meanwhile, this study investigates the weaknesses/shortcomings of current question banks and develops them for 8th grade of junior grade high school. Moreover, Fadilah et al. study used the Thiagarajan research model, whereas this study used Gall, Gall, and Borg (2003) and only used three steps reviewing, planning, and developing.

The second study was conducted by Safrina, Yustiana, & Afandi (2020) Developing card-formed question bank hots oriented in Elementary school. The research mentioned above aims to develop and determine the feasibility of question banks in the form of card higher-order think skills-oriented. The mentioned study used the Puspendik procedure that engaged eight steps: (1) determination of test objectives; (2) grid writing; (3) writing questions; (4) qualitative analysis; (5) assembling questions; (6) test questions: (7) quantitative analysis; (8) question bank. The results show that (1) validation test of 3 experts shows "Strongly Valid" with 0.83 average scores; (2) validation test of the product shows "Strongly Valid" with 0.81 average scores; (3) validation test of 47 questions shows "Strongly Valid" with 0.81 average scores; (4) validation test of the questionnaire for teachers' responses shows "Strongly Valid" with 1.00 average score. However, the mentioned study still uses traditional media to display question banks; meanwhile, the 21st-century era massively engages technology development in all sectors, including assessment in educational sectors. Moreover, the mentioned study did not explain and describe each cognitive level indicator. Also, the mentioned study did not mention and described how theory and item writing indicators were gained.

The third study is conducted by Dilla (2019) Development of an android application-based question bank in the subject of vegetable product processing production to find out the learning outcomes of students at SMK PP Negeri Lembang. The research mentioned above aims to determine the feasibility of an android application-based question bank in vegetable product processing. The research used the ADDIE research model of five steps analysis, design, development, implementation, and evaluation. However, the research mentioned above only focuses on how the feasibility of the question bank platform is displayed, without describing how to develop the item question itself and ensuring

the test item is proper with hot skills as the 21st-century demands.The research mentioned above focused on finding out the feasibility of android application-based question banks in vegetable product processing. Meanwhile, this research will investigate the weaknesses and shortcomings of current question banks and develop them for the 8th grade of junior high school. Moreover, the research mentioned above used the ADDIE research model that consists of five steps analysis, design, development, implementation, and evaluation. Meanwhile, this research employs three steps from the Borg & Gall (1983) model: research and information collection, planning, and developing.

1.9 The Frame of Thought

The researcher framed the research idea to simplfy in understanding the research. The following figure is the frame of thought of this study:

Figure 1.1 Frame of Thought
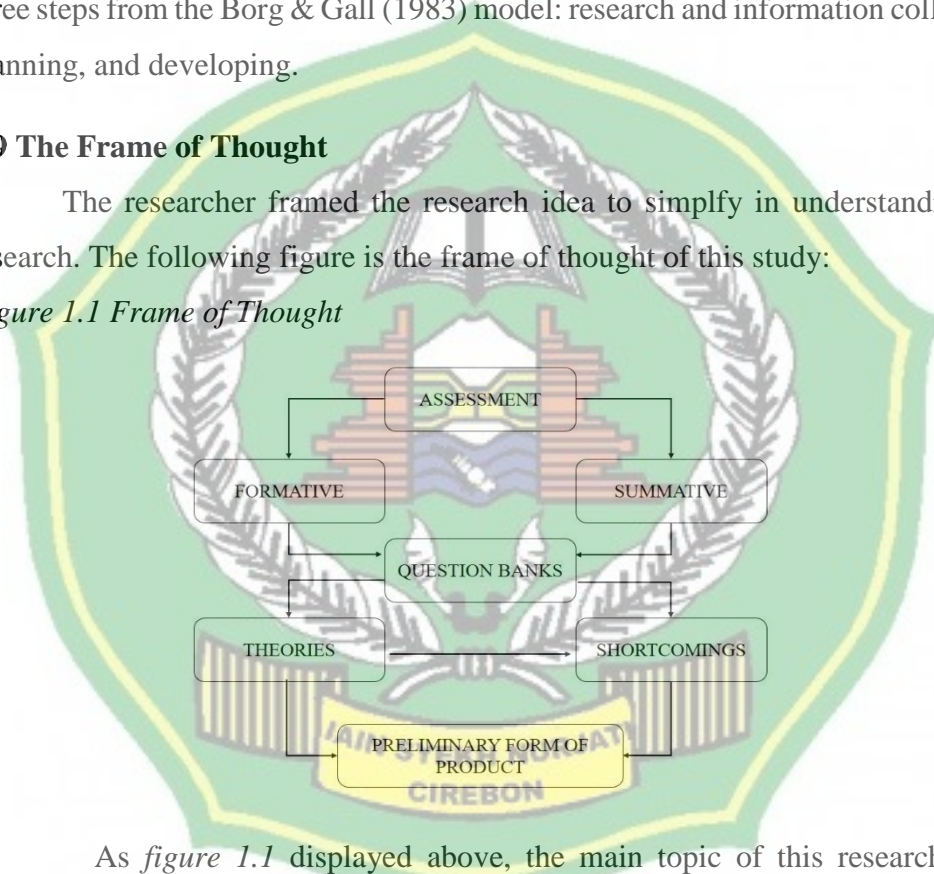
As *figure 1.1* displayed above, the main topic of this research is an assessment which is generally divided into two types formative and summative. Assessment is an evaluation of the teaching and learning process to measure students' performances in the learning process. This study covers formative assessment that measures students' progress in the learning process so that the teacher decides the following steps in the teaching process based on the formative assessment results. However, as pointed out earlier, there are still teachers who have difficulty in composing, planning and conducting assessments; thus, it can impact teaching and learning. Therefore, this study attempted to make a prototype of a

question bank that will be used to assist teachers in conducting a formative assessment.

Furthermore, the developed product in this research requires two aspects as the basis for developing question banks. The first aspect is theories; the researcher reviews literature and documents relevant to the study to determine the characteristics of question banks. After the characteristics were determined, the researcher selected, adapted and merged the theories into an item analysis checklist containing twenty-six indicators divided into three aspects: content & materials, constructions, and language & grammatical. The item analysis checklist was used to investigate the current question banks to find the second aspect's data.

The second aspect in this study that is the basis for developing question banks is the shortcomings of the previous products. The shortcomings of the current product of the question bank were obtained by analyzing the product using an item indicator analysis checklist. The researcher analyzed two current products of question banks, and the results are ten shortcomings were identified in two products. Moreover, the result of reviewing relevant literature in the form of theories and the result of product analysis in the form of shortcomings were extracted and merged to construct the product specification. Product specification includes a product identification table, item indicators table, and item cards. After identifying the specification product, the researcher drafted the preliminary form of question banks.

1.10 Research Method

1.10.1 Research Design and Steps of The Research

This research was conducted using the research and developments (R & D) model and qualitative approach. The researcher employs R & D models since an R & D model is used for data collecting based on product. Based on Gall, Gall, & Borg, (2003) asserted that educational R&D is an industry-based development technique in which findings are utilized to develop innovative products and methods, which are then field-tested, assessed, and developed until they meet particular efficiency, quality, or other requirements (p. 569). This research is a

sequential process that aims to generate a suitable product for the research and development process. Research and development is an organized process involving creating and refining educational programs and resources through formative and summative assessment (Gall, Gall, & Borg, 2014., p. 390).

Moreover, Lestari et al. (2018) defined the R & D model as a descriptive procedural model that emphasizes the processes of product creation, planning, action, and product evaluation (p. 500). The R & D models are sequential processes that create and refine instructional programs and resources through formative and summative assessment. R & D emphasizes product creation, planning, action, and evaluation as a development approach to product assembly.

In research and development models, the research activity involves specific steps. The researcher employed Borg & Gall (1983), cited in Gustiani's (2019) theory as the foundation of research methodology in this research Borg & Gall's study presents ten steps toward R&D models:

1. Research and information collection.
2. Planning.
3. Developing Preliminary Form of Product.
4. Preliminary Field Testing
5. Revising Main Product
6. Main Field Testing
7. Revising Operational Product
8. Operational Field Testing
9. Revising Final Product
10. Disseminating and Implementing

Nevertheless, the researcher did not conduct all ten steps; the researcher only employed three steps as Hidayah, Ramli, & Hanafi (2018) adapted three steps of Borg & Gall's (1983) theory: (1) research and information collection, (2) planning, (3) developing a preliminary form of the product. The following steps are adapted to develop question banks, as displayed in *the figure.1.2* below:

Figure 1.2 Steps of R & D

1.10.1.1 Research and Informations Collections

The researcher reviewed relevant literature and documents; these steps aim to determine the indicators and characteristics of a question bank as a formative assessment. This step results in the form of an item indicator analysis checklist used to analyze the previous product of question banks. Moreover, the researcher analyzed the documents in the form of journal articles, books, website articles, syllabus, English coursebooks, lesson plans, and question banks. The researcher used a table of selection to review the relevant literature, and the result was retrieved and stored in a table of theories and transferred into table indicator checklist analysis. The researcher analyzed two current products of question banks using twenty-six indicators that transformed into a table indicator checklist analysis (see *appendix. 1*). The result data of qualitative analysis is stored in a table of the summary result of item analysis (see *appendix.2*).

1.10.1.2 Planning

Based on previous steps of research and information collection results, the researcher planned the specification of question banks. In other words, the result of qualitative item analysis and reviewing relevant literature transferred them into the table of product specifications, item indicators, and item cards. The product specifications include item indicators and item cards describing the scope and content of what will be tested and providing details about the questions required by the test (Widana, 2014. p. 189). The result of the previous phase is considered in generating product specifications. The purpose of compiling the item specifications, item indicator, and item card table is to guide the researcher in writing the items. The product specifications consist of product identity and a detailed comparison of two existing products. Meanwhile, item indicators contain information on items developed, such as basic competencies, content materials, cognitive levels, and indicators. (See. *Appendix 3, Appendix 4, Appendix 5*).

1.10.1.3 Developing Preliminary Form of Product

After reviewing and planning, the next step is developing the product. This phase involves constructing and assembling the test items by considering the table test specification created beforehand. Moreover, this phase involves translating indicators of the items and level of cognition into questions that follow the details in the table of item specifications. Thus, each question needs to be clarified regarding what is being asked and what answers are required. In other words, each item's quality will determine the question bank's overall quality. Furthermore, after the test item was constructed, the next activity displayed the item test. In this study, the researcher displayed the test item on *quizizz*.

1.10.2 Sources and types of data

Data can provide an overview of the situation of the problem. Nasution (2016, p. 62) asserted that data is a raw material that still needs to be processed to produce qualitative and quantitative information that shows facts. However, the data source in this research is documentation such as related literature, syllabus, English coursebook, and question banks. Moreover, the researcher obtained the literature through *Google*, *Google Scholar*, *Portal Garuda*, *Eric*, *Libgen*, *Z Library*, *Wiley Online Library*, and so on. Meanwhile, the data types will be in the form of characteristic theories used to create product specifications, including item indicators and item cards.

1.10.3 Data collection techniques and instruments

Data collection techniques are essential in research that the researcher must arrange structurally. The researcher used documentation to collect the data that covers two activities:

1) Reviewing relevant literature

In this activity, the researcher reviewed some relevant literature in the form of journal articles, books, and website articles to find the characteristics and theories of question banks. The instrument for reviewing relevant literature is table selection, and the table of selection contains criteria accepted literature for the research sources. There are two criteria, inclusion and exclusion; as Nashruddin &

Mustaqimah (2020, p. 87) pointed out, anything that should be included in the journals is called inclusion. In contrast, anything that is not included in the journal's content is referred to as exclusion. The following table is the list of criteria for inclusion and exclusion:

Table 1.1 Table of Selection

Inclusion	Exclusion
1. Journal articles should be published between 2005 and 2022	1. Journal articles published under 2005
2. Books should be published between 2000 and 2022	2. Book published under 2000
3. Website article is trusted with identified authors	3. Website with no identified authors
4. Journal indexed nationally and internationally	4. Journal articles which do not involve
5. Having journal operation (DOI, ISSN)	5. Without journal operation (DOI, ISSN)
6. Having book operation (ISBN)	6. Without book operation (ISBN)

2) Analyzing the current product of question banks

In this study, the researcher analyzed two products of question banks; the researcher used a table of item indicator analysis checklist obtained from reviewing relevant literature. An indicator checklist contains some characteristics of proper question banks in the 21st century (see *appendix 1*). An indicator checklist was generated from the related theories combined and adapted by the researcher. Meanwhile, item specification tables are an instrument for creating and developing a new item bank. Moreover, item specification tables contain item indicators and item formats for creating new item tests (see *appendix. 3*). Therefore, item-specific tables can guide the researcher in making a question bank. Furthermore, in generating specific tables, the researcher considers the result of qualitative item analysis from analyzing the products of question banks.

1.10.4 Data Analysis Techniques

The researcher implemented critical literature review and content analysis as a data analysis tehniques.

1.10.4.1 Critical literature review

The researcher used a critical literature review; Nashruddin & Mustaqimah (2020, p. 82) defined a critical literature review as an approach utilizing a systematic review strategy. In addition, Jesson & Laccey (as cited in Nashruddin & Mustaqimah, 2020) demonstrated critical literature review as a method used to investigate the strengths and shortcomings of previous work since the last author displayed current concept in their subject field that used to the next author to evaluate their results (p. 82). Furthermore, the researcher employs a critical literature review to review relevant documents that the data result in the form of characteristics of the product. The researcher adopted six steps in analyzing the data from Efron & Ravid (2019, cited in Nashruddin & Mustaqimah, 2020, pp. 83-84):

1) Choosing a review topic and formulating a research question
2) Locating and organizing research sources
3) Selecting, analyzing, and keeping notes of sources
4) Structuring and organizing the literature review
5) Synthesizing and interpreting the literature
6) Putting it all together

The result of this analysis is in these steps in the form of characteristics of the product before transferring them into the table of item indicator analysis checklist. (See *appendix. 1*).
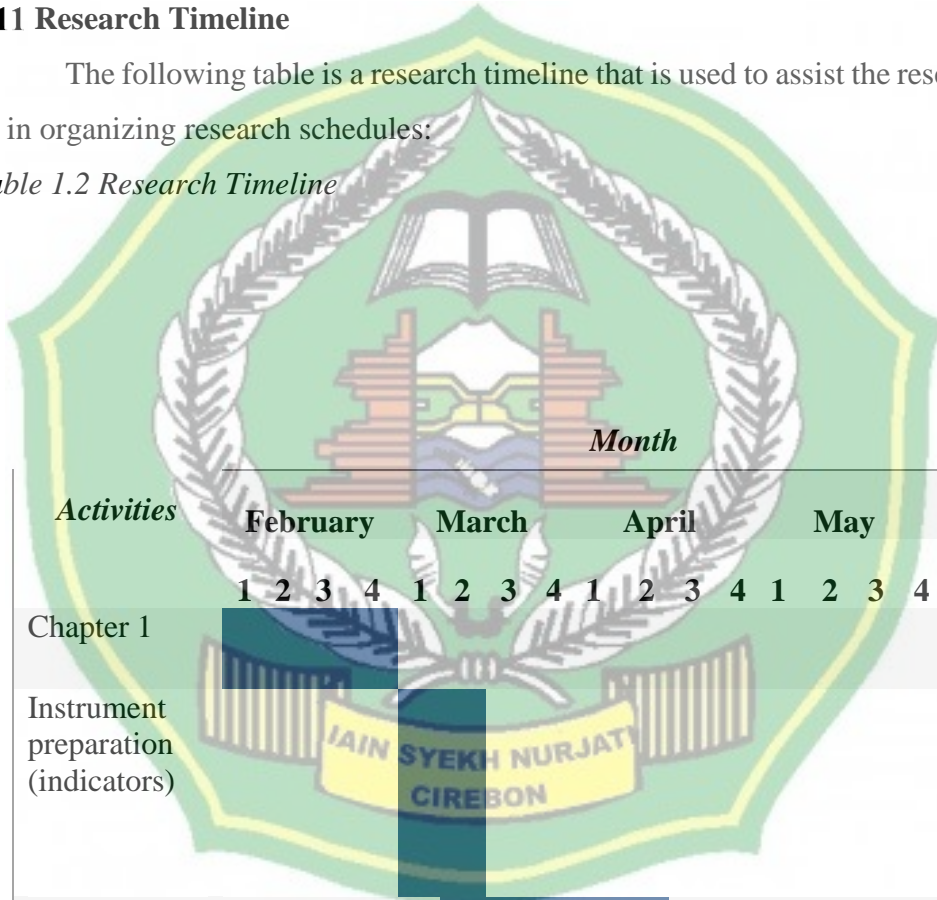
1.10.4.2 Qualitative Item Analysis

The researcher conducts the qualitative item analysis to investigate the shortcomings of question banks. Widana (2014) noted that qualitative item analysis aims to determine whether a question is expected to function correctly, determine the homogeneity of the items in a test, and assess the items in terms of material, test construction, and language (p. 191). Meanwhile, Rahmadhani et al (2014) called qualitative item analysis logic validity which is used to analyze items in terms of technics, content, and editorial. Therefore, qualitative item analysis is used to investigate and evaluate items viewed from content & materials, construction, and language & grammatical; the result of qualitative item analysis is for developing the better quality of items.

Moreover, the instrument for qualitatively analyzing items is the items indicator analysis checklist constructed from critical literature review results (see *appendix 1*). The researcher analyzes the result of item analysis qualitatively by displaying them in the table of summary of the item analysis result. Furthermore, the researcher compares the data from products one and two and confirms the result into theories. Afterwards, the researcher transferred them into product specifications, item indicators, and item cards.

1.11 Research Timeline

The following table is a research timeline that is used to assist the researcher in organizing research schedules:

Table 1.2 Research Timeline

NO	Activities	February				March				April				May				June	
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
1	Chapter 1																		
3	Instrument preparation (indicators)																		
4	Analysis the shortcomings																		
5	Interpretation of findings and writing thesis																		

| 6 | Developing the product | | | | | | | | ■ | |
| 7 | Writing conclusion and finalization of research | | | | | | | | | ■ |