Journal on Interdisciplinary Studies in Humanities
Volume 9 Issue 2, 2026
DOI: https://doi.org/10.34050/elsjish.v9i2.52426
Homepage: journal.unhas.ac.id/index.php/jish

Blueprinting Teacher-Made Tests in EFL Classrooms: Challenges Faced by Junior High School Teachers

Wiwiek Wahyuni[1], Muhammad Nur Akbar Rasyid[1], Surya Dita[1]

[1]Universitas Islam Negeri Alauddin Makassar, Indonesia
*Correspondence: wwahyuniambas@gmail.com

ABSTRACT

Blueprinting, or a table of specifications, is essential in question preparation because it aligns test items with learning objectives, content coverage, cognitive levels, item formats, and scoring weight. In practice, however, it is still often treated as an administrative document rather than an assessment-design tool. This study analyzed English teachers' difficulties in applying blueprinting to question preparation at SMP Negeri 1 Limboro, the factors contributing to these difficulties, and the strategies teachers used to overcome them. Using a descriptive qualitative design, data were collected from five English teachers through semi-structured interviews, observation, and document analysis of blueprint forms and teacher-made tests, and were analyzed thematically. The findings show that teachers' difficulties centered on limited conceptual understanding of blueprinting, difficulty formulating measurable indicators, uncertainty in distributing cognitive levels, weak alignment between items and learning objectives, time constraints, and limited training or institutional support. To cope with these difficulties, teachers relied on peer discussion, adaptation of existing templates, online examples, and workshops. Overall, these difficulties point to a broader issue of teacher assessment literacy shaped by workload and school-level assessment culture. Practical training, collaborative item review, and simple blueprint templates are recommended to improve the quality of teacher-made questions.

ARTICLE HISTORY
Published June 30th 2026
Check for updates

KEYWORDS
Assessment Literacy; Blueprinting; EFL Teachers; Junior High School; Question Preparation; Table of Specifications; Teacher-Made Test.

1. Introduction

Assessment is a central component of classroom instruction because it provides evidence of students' achievement and informs teachers' instructional decisions. In English as a Foreign Language (EFL) classroom, assessment is not only used to assign scores, but also to determine whether students have achieved the expected language competencies. Therefore, the quality of teacher-made questions is important (Said et al., 2021; Rahman et al., 2019; Adinda et al., 2025). Poorly planned tests may fail to represent the taught materials, overemphasize low-level recall, or measure skills that are not aligned with the learning objectives. Conversely, well-planned tests can provide more valid information about students' learning and help teachers improve their teaching practices (Brookhart, 2024; Yaumi et al., 2023; Andini et al., 2026).

One important tool in planning classroom assessment is blueprinting. Blueprinting, or a table of specifications, refers to a systematic plan that connects learning objectives, content areas, cognitive levels, item formats, and the number of test items. Fives and DiDonato-Barnes (2013) explain that a table of specifications can guide teachers in constructing classroom tests and improve the validity of teacher-made assessments. In language assessment, blueprinting is also useful for balancing language skills, linguistic components, text types, and communicative functions so that tests do not depend only on teachers' intuition or previously used questions (Idham et al., 2026; Aswad et al., 2019; Karubaba & Rahman, 2025).

Previous research has contributed important insights into teacher assessment literacy, classroom-based assessment, and the use of tables of specifications. Aria, Sukyadi, and Kurniawan (2021) reported that Indonesian EFL secondary teachers generally had positive perceptions of classroom-based assessment, yet they still needed support in connecting assessment principles with classroom practice. Pastore's (2023) systematic review also shows that teachers often struggle when they have to translate assessment theories into actual classroom action. Coombs & DeLuca (2022)

further argues that teachers' assessment competence should be understood as contextual and situated because assessment decisions are shaped by classroom realities, school expectations, and professional support.

However, the existing body of research still leaves several gaps. Conceptually, many studies discuss teacher assessment literacy as a broad construct, but they do not sufficiently explain how teachers apply that literacy in a specific assessment-design activity such as blueprinting. Methodologically, several studies rely on teachers' perceptions or survey data, while fewer studies triangulate interviews with observation and document analysis of actual blueprint forms and teacher-made tests (Fu, 2025; Anggawirya et al., 2021). Contextually, studies on Indonesian EFL assessment often discuss curriculum implementation or HOTS-based assessment in general, but they provide limited attention to how junior high school teachers construct question grids under real school conditions such as limited time, administrative workload, and informal peer review.

The issue is particularly relevant at SMP Negeri 1 Limboro. English teachers at this school are expected to prepare teacher-made tests based on learning objectives and question grids. Preliminary field concerns indicate that blueprinting has not always been applied consistently. Some teachers still rely on textbook questions, previous test items, or online examples, while the alignment between indicators and final items is not always reviewed systematically. This condition shows the need for a context-specific investigation that captures not only what difficulties teachers face, but also why those difficulties emerge and how teachers attempt to overcome them.

Therefore, this study fills the gap by examining blueprinting as a concrete classroom assessment practice in a junior high school EFL context. It contributes to the discussion of assessment literacy by showing how conceptual knowledge, technical skill, and school-level assessment culture interact in the preparation of teacher-made tests. The study is expected to provide practical implications for teachers, schools, and teacher professional development programs in improving the quality of EFL assessment design. The research questions of this research are; (1) What difficulties do English teachers at SMP Negeri 1 Limboro experience in applying blueprinting to question preparation? (2) What factors contribute to teachers' difficulties in applying blueprinting to question preparation? and (3) What strategies do teachers use to overcome difficulties in applying blueprinting to question preparation? This study aimed to identify the types of difficulties experienced by English teachers in applying blueprinting, analyze the factors contributing to the difficulties, and describe the strategies used by teachers to improve the quality of question preparation.

Previous studies are important to position this research within the broader discussion of blueprinting, teacher assessment literacy, alignment, and language assessment. The following studies show that teachers' difficulties in blueprinting are not isolated problems; they are connected to assessment knowledge, assessment culture, and the practical conditions of school-based test development.

Table 1. Summary of Previous Studies on Blueprinting and Teacher Assessment Literacy

Previous Study	Focus	Main Finding	Relevance to This Study
Fives & DiDonato-Barnes (2013)	Explained the use of a table of specifications in classroom test construction.	A TOS helps teachers frame decisions in test construction and improves the validity of classroom tests.	This study uses the TOS concept as the basis for analyzing teachers' blueprinting practice.
Aria, Sukyadi, & Kurniawan (2021)	Investigated Indonesian EFL secondary teachers' self-perceived assessment literacy.	Teachers may understand assessment purposes but still need support in classroom-based assessment practice.	This supports the need to study how assessment literacy affects blueprinting practice.
Weng & Shen (2022)	Reviewed and discussed language assessment literacy of teachers.	Teachers' assessment literacy includes knowledge, beliefs, classroom practice, and contextual influences.	This supports the analysis of blueprinting as part of language assessment literacy.

Pastore (2023)	Reviewed teacher assessment literacy studies from 2013 to 2022.	Teachers often struggle when translating assessment concepts into classroom practice.	This mini research examines that struggle in the specific activity of question blueprinting.
Sumardi & Guci (2023)	Discussed HOTS-based language assessment literacy in English teaching.	Teachers may confuse linguistic difficulty with cognitive complexity when preparing HOTS items.	This helps explain why teachers may misclassify cognitive levels in a blueprint.
Damayanti, Setyaningsih, & Sumardi (2024)	Studied Indonesian secondary EFL teachers' language assessment literacy.	Language assessment literacy involves knowledge, perspectives, and classroom practice.	This study extends the issue by focusing on blueprinting as a concrete assessment task.
Brookhart (2024)	Revisited educational assessment knowledge and skills for teachers.	Assessment competence is contextual and connected to classroom decisions.	This supports the analysis of contextual factors such as workload, peer review, and school support.

As presented in Table 1, previous studies converge on the idea that a table of specifications supports validity and informs teachers' assessment decisions, yet most research addresses assessment literacy at a general level rather than within a specific design task such as blueprinting. This gap justifies the present study's focus on how conceptual, technical, and contextual dimensions interact when teachers construct blueprints in an actual EFL classroom setting.

1.1 Blueprinting in Question Preparation

Blueprinting is a systematic process of planning a test before the test items are written. In educational assessment, a blueprint or table of specifications functions as a map that shows the relationship between instructional objectives, content coverage, cognitive processes, item formats, and scoring weight. Nitko and Brookhart (2014) explain that assessment should begin with clear learning targets because the validity of classroom assessment depends on the match between what is intended to be measured and what is actually measured. In this sense, blueprinting is not only a technical document, but also a validity-oriented planning process.

A good blueprint helps teachers make deliberate decisions about the number and type of items needed to represent the curriculum. If a unit emphasizes reading comprehension and text analysis, for instance, the blueprint should allocate a reasonable number of items to those competencies. If the curriculum requires higher-order thinking skills, the blueprint should include items that measure analyzing, evaluating, or creating, not only remembering and understanding (Yosepha et al., 2023). This makes blueprinting useful for content representativeness and prevents teachers from overtesting minor materials while ignoring important competencies.

In EFL question preparation, blueprinting is especially important because teachers must consider language skills, vocabulary, grammar, text type, communicative purpose, and cognitive level at the same time. Brown and Abeywickrama (2019) argue that classroom language assessment should connect test tasks with the ability or skill intended to be measured. Therefore, an English blueprint should not only mention the number of questions, but also clarify whether the items measure literal comprehension, inference, vocabulary in context, grammar use, writing organization, or other language competencies.

1.2 Teacher Assessment Literacy

Teacher assessment literacy refers to teachers' knowledge, skills, and dispositions in designing, implementing, interpreting, and using assessment appropriately. Popham (2018) states that teachers need assessment literacy because many classroom decisions are based on assessment evidence. Brookhart (2024) also emphasizes that educational assessment knowledge includes selecting appropriate assessment methods, developing high-quality tasks, interpreting evidence, and using results to support learning. These competencies are directly related to blueprinting because teachers must understand what to assess, how to assess it, and how to interpret the result.

In language education, teacher assessment literacy becomes more specific because teachers need to understand language constructs, communicative competence, item formats, scoring criteria, and the relationship between test tasks

and language use. Xu and Brown (2016) argue that assessment literacy in practice is shaped by teachers' knowledge base, conceptions of assessment, institutional contexts, and professional identity. This means that a teacher may understand the definition of blueprinting but still experience difficulty applying it when time, examples, or school support are limited.

Several Indonesian studies support this issue. Aria et al. (2021) show that Indonesian EFL teachers need continuing support in classroom-based assessment practice. Damayanti et al. (2024) also report that language assessment literacy includes the interaction between knowledge, perspectives, and actual classroom practice. Therefore, teachers' difficulty in preparing blueprints can be understood as part of a broader assessment literacy issue, not merely as a problem of filling in a table.

1.3 Alignment, Cognitive Level, and Validity

Alignment refers to the consistency among learning objectives, indicators, teaching materials, cognitive levels, and assessment items. In blueprinting, alignment is important because the blueprint should guide teachers to write questions that measure the intended learning outcomes. Revised taxonomy is often used to classify learning objectives and assessment tasks into cognitive levels such as remembering, understanding, applying, analyzing, evaluating, and creating. This taxonomy helps teachers decide the expected level of thinking for each item.

However, alignment is not achieved simply by writing action verbs in the blueprint. The cognitive level of an item is determined by the mental process required to answer the question. For example, a question using the word analyze may still measure lower-order thinking if students only need to copy explicit information from a text. Sumardi and Guci (2023) note that HOTS-based language assessment is challenging because teachers must distinguish between linguistic difficulty and cognitive complexity. A long English text may look difficult, but if the answer can be found directly, the item may only measure understanding.

Alignment is also closely related to validity. The Standards for Educational and Psychological Testing define validity as the degree to which evidence and theory support interpretations of test scores for intended uses (Association et al., 2014). In classroom assessment, validity means that the test should measure the intended learning outcomes and should be used for appropriate decisions. A blueprint can strengthen validity because it clarifies what will be assessed before the items are written. However, if the blueprint is poorly constructed or not followed during item writing, its value becomes limited.

1.4 Difficulties in Applying Blueprinting

Teachers' difficulties in applying blueprinting may arise from conceptual, technical, and contextual sources. Conceptual difficulties occur when teachers know that a blueprint is required but do not fully understand its role in ensuring content validity, cognitive distribution, and item alignment. Pastore (2023) emphasizes that teachers often face difficulty when moving from assessment concepts to classroom implementation. In blueprinting, this difficulty can appear when teachers fill out the form but do not use it as a guide during item construction.

Technical difficulties include formulating measurable indicators, selecting item formats, determining cognitive levels, distributing questions proportionally, and checking the match between indicators and final items. Fives and DiDonato-Barnes (2013) show that a table of specifications can support test construction, but teachers still need practical skill to apply it. In EFL contexts, the task becomes more complex because teachers must balance language skills and cognitive demand.

Contextual difficulties include limited time, workload, lack of training, limited access to examples, and weak peer review. Brookhart (2024) argues that assessment knowledge and skills are situated in classroom and institutional contexts. Therefore, teachers' difficulties cannot be solved only by asking teachers to be more careful. Schools need to create assessment support systems, including collaborative planning, item review, and practical training.

1.5 Conceptual Framework

The conceptual framework of this study was developed from the idea that blueprinting is influenced by teacher assessment literacy, alignment principles, and school-based assessment culture. Fives and DiDonato-Barnes (2013) position the table of specifications as a tool for improving classroom test validity. Xu and Brown (2016) explain that teachers' assessment literacy in practice is shaped by knowledge, conceptions, identity, and context. Brookhart (2024) further emphasizes that teachers' assessment decisions are contextual and connected to classroom realities. Based on

these references, this study views teachers' difficulties in blueprinting as the interaction of cognitive, technical, and contextual dimensions.

The cognitive dimension refers to teachers' understanding of assessment concepts, learning objectives, indicators, cognitive levels, and validity. The technical dimension refers to teachers' ability to construct blueprint tables, distribute items, write aligned questions, and review item quality. The contextual dimension refers to time, workload, training opportunities, school policy, peer collaboration, and availability of examples. These dimensions influence the quality of question preparation and the alignment between curriculum objectives and teacher-made test items.

Table 2. Dimensions of Teachers' Difficulties in Applying Blueprinting

Dimension	Examples of Difficulty	Possible Impact on Question Preparation
Cognitive	Limited understanding of blueprinting, indicators, cognitive level, and assessment validity	Questions may not match learning objectives or intended thinking skills.
Technical	Difficulty constructing blueprint tables, distributing items, and matching items with indicators	Blueprint exists as a form but does not guide test construction.
Contextual	Limited time, heavy workload, lack of training, and weak peer review	Teachers rely on previous tests or online examples without systematic planning.

As shown in Table 2, the three dimensions are interrelated: limited cognitive understanding constrains teachers' technical execution, while unfavorable contextual conditions such as time pressure and weak peer review further limit teachers' opportunities to refine their blueprinting practice. This framework guided the subsequent data collection and analysis in identifying, explaining, and categorizing teachers' difficulties, their contributing factors, and their coping strategies.

2. Methodology

This study used a descriptive qualitative design. The design was selected because the study aimed to explore teachers' experiences, difficulties, and strategies in applying blueprinting to question preparation at SMP Negeri 1 Limboro. A qualitative design is appropriate when the researcher wants to understand a phenomenon in its natural context and interpret participants' perspectives in depth (Creswell, 2018). In this study, the focus was not to measure the frequency of difficulties statistically, but to describe the types, causes, and meanings of difficulties experienced by teachers.

The study was conducted at SMP Negeri 1 Limboro. The site was selected because English teachers at the school prepare teacher-made tests and are expected to use blueprinting or question grids before constructing examination questions. Site selection in qualitative research should be based on relevance to the phenomenon being studied and access to information-rich cases.

The participants were five English teachers who had experience preparing classroom tests, mid-semester tests, or final-semester questions. The participants were selected purposively because they were directly involved in question preparation and had relevant experience with blueprinting. Purposive sampling is suitable in qualitative research when the researcher needs participants who can provide rich information about a specific issue (Creswell & Poth, 2018).

The data were collected through semi-structured interviews, observation, and document analysis. Interviews were used to explore teachers' understanding of blueprinting, their difficulties, the causes of the difficulties, and their strategies for solving problems. Semi-structured interviews are useful because they provide guiding questions while still allowing participants to explain their experiences in detail (Creswell & Poth, 2018).

Observation was used to examine how teachers prepared blueprints and questions in practice, especially how they identified competencies, wrote indicators, selected cognitive levels, and developed items. Document analysis was conducted by reviewing sample blueprint forms, question grids, and teacher-made questions to see whether the items were aligned with learning objectives, indicators, and cognitive levels. Bowen (2009) explains that document analysis can support qualitative studies by providing contextual and corroborative evidence.

The data collection was conducted in several stages. First, the researcher coordinated with the English teachers at SMP Negeri 1 Limboro and explained the purpose of the study. Second, the researcher collected available blueprint documents and teacher-made questions. Third, the researcher observed the process of question preparation, particularly how teachers identified competencies, wrote indicators, selected cognitive levels, and developed items. Fourth, the researcher interviewed the teachers individually using a semi-structured interview guide. Finally, all interview notes, observation notes, and documents were organized for analysis. These procedures followed qualitative research principles that emphasize natural setting, multiple data sources, and systematic data organization.

The data were analyzed using thematic analysis. Braun and Clarke (2006) describe thematic analysis as a method for identifying, analyzing, and reporting patterns or themes within qualitative data. In this study, the analysis began with reading the interview transcripts, observation notes, and documents repeatedly to understand the overall data. After that, the researcher coded important statements related to teachers' difficulties, causes, and strategies. Similar codes were grouped into broader categories. These categories were then developed into themes. The themes were reviewed by comparing them with the data sources to ensure that the interpretation was supported by evidence. Finally, the findings were written and connected with relevant theories and previous studies.

To ensure trustworthiness, this study used credibility, transferability, dependability, and confirmability. These criteria are commonly used in qualitative research to strengthen the quality of findings. Credibility was strengthened through triangulation by comparing interview data, observation notes, and blueprint documents. Transferability was supported by providing a clear description of the research context, participants, and data collection process. Dependability was maintained through an audit trail, including documentation of data collection procedures, coding, and theme development. Confirmability was supported by checking whether the findings were grounded in participants' responses and document evidence rather than the researcher's personal assumptions.

3. Results and Discussion

3.1 Results

The findings are presented based on the three research questions. Overall, the data showed that teachers recognized the importance of blueprinting, but they still faced conceptual, technical, and contextual difficulties in applying it consistently. The findings were organized into three main categories: the difficulties faced by teachers, the factors contributing to the difficulties, and the strategies used by teachers to overcome the difficulties. Interview excerpts are presented using teacher codes T1 to T5 to protect participants' identities.

Table 3. Summary of Findings Based on the Research Questions

Research Question	Main Theme	Brief Finding
RQ1	Conceptual understanding	Teachers understood blueprinting as a required document but not always as a tool for validity and alignment.
RQ1	Indicator formulation	Teachers found it difficult to write specific and measurable indicators based on learning objectives.
RQ1	Cognitive-level distribution	Teachers were uncertain in classifying items into LOTS and HOTS levels.
RQ1	Blueprint-item alignment	Some final test items did not closely match the indicators written in the blueprint.
RQ2	Time and workload	Teachers had limited time to prepare detailed blueprints before tests.
RQ2	Limited training	Teachers lacked practical training and examples of good blueprints.

RQ3	Practical strategies	Teachers used peer discussion, online examples, templates, and workshops to overcome difficulties.

As summarized in Table 3, teachers' difficulties were concentrated mainly in the first research question, spanning conceptual understanding, indicator formulation, cognitive-level distribution, and blueprint-item alignment, while the contributing factors centered on time, workload, and limited training. In response, teachers relied on informal but resourceful strategies such as peer discussion, workshops, and template adaptation. Each of these themes is discussed in detail in the following sections.

3.1.1 Difficulties Faced by English Teachers in Applying Blueprinting to Question Preparation

a. Limited Conceptual Understanding of Blueprinting

The first difficulty was related to teachers' limited conceptual understanding of blueprinting. Most teachers were familiar with the term blueprint or question grid, but their understanding tended to be procedural. They knew that a blueprint should contain competencies, indicators, cognitive levels, and item numbers, yet they did not always understand how these components function together. Some teachers considered blueprinting as an administrative requirement rather than a planning tool to improve test quality.

One teacher stated, "I know that the blueprint must be submitted before the test, but sometimes I only see it as a format to complete, not as a guide when writing each question" (T1, interview). Similar views were expressed by other teachers, who reported that they routinely completed the blueprint's components without revisiting it to verify whether the finished questions actually reflected what had been planned.

This finding was supported by document analysis. Several blueprint forms were filled in, but the connection between indicators and questions was not always clear. In some cases, the same indicator was used for several items even though the items measured different skills. In other cases, the cognitive level written in the blueprint did not match the actual demand of the question.

b. Difficulty Formulating Measurable Indicators

The second difficulty was the formulation of indicators. Teachers often found it challenging to transform broad learning objectives into specific indicators that could be measured through test items. Some indicators were too general, such as 'students understand descriptive text,' without specifying what students had to identify, infer, compare, or produce. As a result, the indicators did not provide clear direction for item construction.

A teacher explained, "The learning objective is sometimes broad, so I am confused about how to change it into a specific indicator. I often write 'understand the text,' but later I realize the question can measure vocabulary, main idea, or detailed information" (T3, interview). Other teachers echoed this experience, noting that producing an indicator specific enough to be measured in a single test item was consistently the most demanding part of blueprint preparation.

c. Confusion in Determining Cognitive Levels

The third difficulty was determining cognitive levels. Teachers understood the general distinction between lower-order thinking skills and higher-order thinking skills, but they were not always confident in classifying questions according to the revised Bloom's taxonomy. For example, some questions labeled as analyzing only required students to find explicit information in a text. Similarly, some questions labeled as evaluating did not require students to make judgments based on criteria.

One teacher stated, "Sometimes I choose C4 because the question looks difficult, especially when the text is long. But after discussion, I realize the students only need to find the answer directly in the text" (T4, interview). This pattern recurred across interviews, as teachers frequently admitted relying on the operational verb of a question rather than critically examining the mental process it actually required.

d. Weak Alignment Between Blueprint and Test Items

The fourth finding was weak alignment between the blueprint and the final test items. Although teachers prepared blueprints before constructing questions, some items in the final test did not correspond closely to the indicators. In some cases, the blueprint indicated that students should analyze text structure, but the question only asked about word meaning.

In other cases, the blueprint planned a balanced distribution of item types, but the final test was dominated by multiple-choice questions with similar patterns.

A teacher explained, “After the questions are finished, I sometimes find that some questions do not match the indicator. It happens because I write the questions quickly and use examples from different sources” (T5, interview). Other teachers reported comparable experiences, where blueprints intended to measure inference or text-structure analysis ended up producing items whose answers could be located directly in the text.

e. Time Constraints and Workload

The fifth difficulty was time constraint. Teachers reported that they often had to prepare questions while also handling teaching, administrative tasks, student mentoring, and school programs. Because of this workload, blueprinting was sometimes done quickly. Teachers tended to reuse previous blueprints or adapt questions from textbooks and online sources to save time. Although this strategy helped them complete the task, it did not always ensure alignment with current learning objectives.

One teacher said, “The deadline for submitting questions is sometimes close, while we also have teaching, administration, and other school activities. So, the blueprint is often completed quickly” (T1, interview). This time pressure was a recurring theme, leading several teachers to adapt previous years’ formats rather than developing a blueprint specifically aligned with the current material.

f. Limited Training and Institutional Support

The sixth difficulty was limited training. Teachers stated that they had attended general assessment workshops, but the training often focused on curriculum policy rather than practical blueprinting. They needed concrete examples, step-by-step exercises, and feedback on their own blueprint products. Without practical training, teachers relied on existing formats without fully understanding how to improve them.

A teacher stated, “We have joined assessment training, but the explanation is usually general. What we need is practice, for example making indicators, deciding cognitive levels, and checking questions together” (T2, interview). Teachers consistently noted that peer discussion was helpful but remained informal, without a regular schedule for reviewing blueprints together.

3.1.2 Factors Contributing to Teachers' Difficulties in Applying Blueprinting to Question Preparation

The second research question focused on the factors that contributed to teachers' difficulties in applying blueprinting to question preparation. The data showed that the difficulties did not come from a single cause. Instead, they were influenced by a combination of teachers' assessment literacy, limited practical experience, time pressure, availability of examples, and the absence of systematic school-level review.

a. Limited Assessment Literacy in Practical Test Design

The first contributing factor was limited assessment literacy, especially in translating assessment concepts into practical test design. Teachers were familiar with the general components of a blueprint, but they were not always confident in using the blueprint to control content coverage, cognitive level, item format, and alignment. This condition made blueprinting appear as a document to be completed rather than as a tool for designing valid questions.

One teacher stated, "I know the parts of a blueprint, such as indicators, materials, cognitive level, and item number. However, when I have to connect all of them in one test, I still need guidance because I am afraid the questions do not really match the indicators" (T1, interview).

b. Limited Practical Training and Guided Feedback

The second factor was the lack of practical training and guided feedback. Although teachers had received information about assessment and curriculum implementation, the training was often general. Teachers needed more direct practice in writing indicators, classifying cognitive levels, developing item formats, and reviewing the match between blueprints and final test items.

A teacher reported, “In workshops, we usually receive explanations about assessment policy, but we rarely practice making a complete blueprint from the learning objective until the question item” (T2, interview). Teachers similarly indicated that receiving structured feedback on their own blueprint drafts, rather than only general workshop content, would help them identify and strengthen the components that remained weak.

c. Time Constraints and Administrative Workload

The third factor was time constraint. Teachers prepared questions while also handling teaching duties, administration, student mentoring, and school activities. As a result, they often had limited time to prepare, review, and revise blueprints carefully. This situation encouraged teachers to reuse previous formats or adapt existing questions without conducting a detailed alignment check.

d. Limited Access to Standardized Examples and Review Instruments

The fourth factor was limited access to standardized examples and review instruments. Teachers could find many examples of blueprints from the internet, but the quality and suitability of those examples varied. Without a clear checklist, teachers sometimes found it difficult to decide whether an online example was appropriate for their curriculum, students' level, and learning objectives.

e. Absence of Systematic Peer Review and Institutional Support

The fifth factor was the absence of systematic peer review and institutional support. Teachers sometimes discussed blueprints informally with colleagues, but the school did not always provide a regular mechanism for collaborative item review. Consequently, blueprinting quality depended heavily on individual initiative rather than a consistent assessment culture at the school level.

3.1.3 Teachers' Strategies to Overcome Difficulties in Applying Blueprinting to Question Preparation

Despite the difficulties, teachers used several strategies to improve their blueprinting practice. These strategies showed that teachers attempted to solve practical problems through collaboration, adaptation, professional learning, and the use of curriculum resources.

a. Discussing Indicators and Question Formats with Other English Teachers

First, teachers discussed indicators and question formats with other English teachers. Peer discussion helped them check whether the indicators were measurable, whether the question format was suitable, and whether the item matched the intended cognitive level. One teacher explained, "When I am not sure whether the indicator is specific enough, I usually discuss it with another English teacher. We compare the indicator with the question and decide whether the item is suitable" (T3, interview).

b. Searching Online Examples of Blueprints and Adapting Them to Lessons

Second, teachers searched for online examples of blueprints and adapted them to their lessons. Online examples helped teachers understand possible formats and wording, especially when they needed a reference for organizing competencies, indicators, materials, and item distribution. A teacher stated, "I sometimes look for blueprint examples from the internet to see the format and the way the indicators are written. However, I do not copy them directly because the materials and students' level are different" (T4, interview).

c. Reusing Previous Blueprint Formats and Modifying the Content According to the Current Topic

Third, teachers reused previous blueprint formats and modified the content according to the current topic. This strategy helped teachers save time and maintain a familiar structure. One teacher explained, "I usually use last year's blueprint as a reference because the format is already available. After that, I change the topic, learning objectives, indicators, and item numbers based on the current material" (T1, interview).

d. Participating in Workshops When Available

Fourth, teachers participated in workshops when available. Workshops provided teachers with opportunities to refresh their understanding of assessment, learn about curriculum expectations, and discuss examples of question preparation. A teacher stated, "When there is a workshop about assessment, I try to join because it helps me understand the newest guidance. However, I prefer workshops that give us practice in making indicators and questions" (T2, interview).

e. Using Curriculum Documents and Textbooks to Identify Learning Objectives and Content Coverage

Fifth, some teachers used curriculum documents and textbooks to identify learning objectives and content coverage. This strategy helped teachers decide which competencies, materials, language skills, and text types should be represented in the test. A teacher explained, "Before writing the blueprint, I check the curriculum document and the textbook to see the

learning objectives and the materials that have been taught. From there, I decide which indicators should appear in the test" (T3, interview).

These strategies indicate teachers' professional agency, even though the support system still needs to be strengthened. Peer discussion, online references, previous formats, workshops, and curriculum documents helped teachers overcome immediate problems, but these strategies would be more effective if supported by regular collaborative review, a common blueprint checklist, and practical assessment training at the school level.

3.2 Discussion

The findings indicate that teachers' difficulties in applying blueprinting are closely related to assessment literacy, technical competence, and school context. Blueprinting requires teachers to understand learning objectives, indicators, cognitive levels, item formats, validity, and fairness. When teachers' assessment literacy is limited, they may complete the blueprint form without using it meaningfully. This supports Popham's (2018) and Brookhart's (2024) view that assessment literacy is fundamental because classroom assessment decisions directly affect students' learning and the interpretation of achievement.

A critical point emerging from this study is that blueprinting difficulties are not caused only by teachers' lack of knowledge. They also reflect the way assessment work is organized in schools. In the Indonesian school context, teachers often work with multiple responsibilities: teaching, preparing administrative documents, mentoring students, attending school programs, and completing curriculum-related reports. Under these conditions, blueprinting may become a compliance document rather than a reflective assessment-design process. This explains why teachers sometimes reuse previous formats or adapt online examples even when they understand that each test should be aligned with current learning objectives.

The findings also show that the culture of assessment in the school still needs to move from individual completion to collaborative quality assurance. Teachers sometimes discussed indicators and questions with colleagues, but the discussion was informal and depended on personal initiative. This indicates that the problem is not the absence of professional willingness; rather, it is the absence of a structured mechanism that allows teachers to review blueprints and test items together before tests are administered. In this sense, school-level assessment culture influences whether blueprinting functions as a tool for validity or merely as an administrative attachment.

In relation to the first research question, the teachers' main difficulties appeared in the process of translating learning objectives into measurable indicators, determining cognitive levels, and maintaining alignment between the blueprint and the final test items. These findings show that blueprinting is not merely a clerical activity. It requires teachers to make assessment decisions that connect curriculum targets, classroom instruction, and evidence of student learning. Without clear indicators, teachers may produce questions that measure different skills from those stated in the blueprint.

The confusion in determining cognitive levels reflects a common problem in applying the revised Bloom's taxonomy. Teachers often focus on verbs without analyzing the actual thinking process required by the question. This can lead to over labeling simple recall questions as HOTS items. In EFL assessment, this problem becomes more complex because a question may appear difficult due to vocabulary or text length, but cognitively it may still require only locating explicit information. Therefore, teachers need practical training to distinguish linguistic difficulty from cognitive complexity.

The weak alignment between blueprint and test items also has implications for validity. According to assessment standards, validity concerns the interpretation and use of test scores (AERA, APA, & NCME, 2014). If a test item does not measure the intended indicator, the score may not accurately represent students' competence. Thus, blueprinting should be followed by item review and revision so that each question can be traced back to the learning objective and indicator stated in the blueprint.

In relation to the second research question, the factors contributing to the difficulties were both internal and external. Internal factors included limited assessment literacy, uncertainty in applying cognitive taxonomy, and lack of confidence in reviewing item alignment. External factors included time constraints, administrative workload, limited practical training, lack of standardized examples, and the absence of systematic peer review. These factors show that teachers' difficulties cannot be solved only by asking teachers to be more careful. Assessment practice is situated in classroom and institutional realities (Xu & Brown, 2016; Brookhart, 2024).

In relation to the third research question, the strategies used by teachers showed professional agency. Teachers did not ignore the difficulties; instead, they attempted to overcome them by discussing with colleagues, searching online

examples, adapting previous formats, joining workshops, and referring to curriculum documents and textbooks. However, these strategies were mostly informal and individual. Peer discussion, for example, would be more effective if supported by a regular review schedule and a common checklist. Online examples were helpful, but teachers still needed the ability to evaluate whether the examples were valid and suitable for their students.

Overall, this study extends previous research by showing that teacher assessment literacy becomes visible in the micro-practice of blueprinting. The findings suggest that assessment literacy should not only be developed through theoretical explanation, but also through guided practice, document-based review, and collaborative discussion of real teacher-made tests. For SMP Negeri 1 Limboro and similar school contexts, improving blueprinting practice requires a school policy that provides time allocation, peer review routines, simple templates, and feedback-based professional development.

4. Conclusion

This study concludes that the central challenge in blueprinting teacher-made tests in EFL classrooms is the gap between teachers' awareness of blueprinting and their ability to use it as an assessment-design tool. Although the teachers were familiar with the blueprint format, they still struggled to formulate measurable indicators, determine cognitive levels, and align test items with the blueprint. This indicates that blueprinting problems are not merely technical errors in completing a table, but reflect a broader assessment literacy challenge.

Theoretically, the study contributes to understanding teacher assessment literacy by showing how it operates within a specific classroom assessment task, involving the interaction of conceptual understanding, technical competence, and contextual support. This finding strengthens the view that assessment literacy is situated in teachers' real working conditions and cannot be separated from school assessment culture.

Practically, the findings suggest that English teachers need more than general information about assessment policy; they require hands-on training in writing indicators, classifying cognitive levels, and reviewing teacher-made tests, supported by systematic peer review, simple blueprint templates, and sufficient time for collaborative question preparation.

This study was limited to five English teachers in one junior high school, so the findings should be interpreted as context-specific rather than broadly generalizable. Future research could involve more schools, compare subjects, or examine the quality of teacher-made questions before and after blueprinting training, including the development of a practical blueprinting checklist linking learning objectives, indicators, cognitive levels, item formats, and final test items.

References

Adinda, R., Sosrohadi, S., Syafitri, B. A., & Andini, C. (2025). Cognitive And Cultural Barriers In Synonim Acquisition: A Psycolinguistic Study Of Indonesian Learners Of Korean. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, *32*(4), 881-888.

Andini, C., Youngsun, K., Sunnuraini, S., & Warouw, D. S. (2026). Developing a Chunk-Based Learning Model to Improve Korean Speaking Skills for Beginner Learners. *International Journal of L2CT*, *1*(2), 156-172.

Anggawirya, A. M., Prihandoko, L. A., & Rahman, F. (2021, December). Teacher's role on teaching English during pandemic in a blended classroom. In *International Joined Conference on Social Science (ICSS 2021)* (pp. 458-463). Atlantis Press.

Aria, D., Sukyadi, D., & Kurniawan, E. (2021). Teacher assessment literacy: Indonesian EFL secondary teachers' self-perceived classroom-based assessment practice. *English Review: Journal of English Education*, *10*(1), 15–26.

Association, A. E. R., Association, A. P., & Education, N. C. on M. in. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Aswad, M., Rahman, F., Said, I. M., Hamuddin, B., & Nurchalis, N. F. (2019). A software to increase English learning outcomes: An acceleration model of English as the second language. *The Asian EFL Journal*, *26*(6.2), 157.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brookhart, S. M. (2024). Educational assessment knowledge and skills for teachers revisited. *Education Sciences*, *14*(7), 751.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Coombs, A., & DeLuca, C. (2022). Mapping the constellation of assessment discourses: A scoping review study on assessment competence, literacy, capability, and identity. *Educational Assessment, Evaluation and Accountability*, *34*(3), 279–301.

Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Damayanti, F. Y., Setyaningsih, E., & Sumardi, S. (2024). Teachers' language assessment literacy: Indonesian secondary EFL teachers' knowledge, perspective, and classroom practice. *Tarbiyah: Jurnal Ilmiah Kependidikan*, *13*(2), 171–185.

Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation*, *18*(3), 1–7.

Fu, R. (2025). *Exploring Chinese K–12 Teachers' Formative Assessment Perceptions and Literacy: A Mixed Methods Study*.

Idham, A. Z., Rauf, W., & Chandra, A. (2026). *Assessment and Evaluation in ELT Principles, Methods, and Practices*. Minhaj Pustaka.

Karubaba, S., & Rahman, F. (2025). Code-Switching and Code-Mixing in Indonesian EFL Classrooms: Teacher-Student Interactions in North Biak. *Dialectica Online Publishing Journal*, *1*(1), 107-115.

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, *8*, 1217167.

Popham, W. J. (2018). *Assessment literacy for educators in a hurry*. ASCD.

Rahman, F., Abbas, A., Hasyim, M., Rahman, F., Abbas, A., & Hasyim, M. (2019). Facebook group as media of learning writing in ESP context: A case study at Hasanuddin University. *Asian EFL Journal Research Articles*, *26*(6.1), 153-167.

Richlin, L. (2023). *Blueprint for learning: Constructing college courses to facilitate, assess, and document learning*. Taylor & Francis.

Said, M. M., Rita, F., Weda, S., & Rahman, F. (2021). English language performance development through extracurricular activities at Faculty of Teacher Training and Education Tadulako University Palu. *PalArch's Journal of Archaelogy of Egypt/Egyptology*.

Sumardi, S., & Guci, R. I. (2023). HOTS-based language assessment literacy: Challenges and prospects in English language teaching. *Indonesian Journal of Applied Linguistics*, *12*(3), 818–827.

Weng, F., & Shen, B. (2022). Language assessment literacy of teachers. *Frontiers in Psychology*, *13*, 864582.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162.

Yaumi, M. T. A. H., Rahman, F., & Sahib, H. (2023). Exploring WhatsApp as Teaching and Learning Activities during Covid-19/New Normal era: A Semiotic Technology Analysis. *International Journal of Current Science Research and Review*, *6*(12), 7627-7634.

Yosepha, A., Ali, M., Wahyudin, D., & Rusman, R. (2023). The role of multi-dimensional curriculum design in improving higher-order thinking skills. *International Journal of Learning, Teaching and Educational Research*, *22*(7), 219–239.