

## Developing Teacher Assessment Literacy Questionnaire for University Language Teachers

**Nurdiana**

*Department of English Language and Culture, Universitas Bunda Mulia, Indonesia  
nurdiana@bundamulia.ac.id*

---

**Abstract:** *Language teacher professional development thus far has been focusing on how to teach language learners effectively. Other than improving teaching techniques, they are supposed to enhance their assessment literacy. This is due to the findings of earlier studies on teacher assessment literacy indicating that many teachers are assessment illiterate. Language teachers should be assessment literate as it can help learners improve their language learning by being able to assess their students appropriately and effectively. To investigate language teacher assessment literacy, an instrument called language teacher literacy questionnaire needs to be developed. For this reason, the present study attempts to propose dimensions or indicators and proposition of the questionnaire for university language teachers. Having been developed, the indicators and the propositions were evaluated by 20 university language teachers who have at least five years of teaching experience of teaching a (foreign) language at the tertiary level. They provided feedback or comments on the proposed questionnaire then they were discussed in this research. Results of this study revealed that it was not an easy assignment to determine the indicators of the questionnaire and the construction of the propositions/items was time-consuming. The feedback varied greatly starting from the statement 'the items were good already' to 'complaining about the instruction of completing the questionnaire'. Implicitly, the respondents might not be assessment literate in terms of knowledge of assessment and the questionnaire devised needs a lot of improvement.*

**Keywords:** *Language Teacher Assessment Literacy, Language Teacher Assessment, Literacy Questionnaire*

---

### I. INTRODUCTION

Other than teaching, devising a syllabus, and preparing lessons, assessing students' performance is the teachers' responsibility. Therefore, teachers are obliged to have knowledge of assessment—that is how to assess their students effectively and properly. An early study on assessment literacy reported that without having knowledge of assessment, teachers will not be able to help their students to enhance their performance (Coombe et al., 2012).

Being literate to assessment is paramount due to some reasons. First of all, one of the teachers' routines is assessing students. Teachers obviously need assessment skills and knowledge because they get involved very often in assessment scoring and administration (Plake & Impara, 1997). Secondly, teachers should have the ability to devise valid and reliable assessment methods to examine their students' learning progress and the most important thing is to measure how effective the teaching is (Mellati & Khademi, 2018). They further said that

teachers can improve their teaching, increase their students' motivation to learn and they can help students enhance their achievement by having the ability to employ appropriate assessments. Another underlying principle of why assessment literacy is important for teachers is classroom teaching and learning materials will have no value if teachers have less knowledge of classroom assessment (Susuwele-Banda, 2005; Al-Malki & Weir, 2014).

The importance of assessment literacy, however, has not been fully supported in assessment practices. This is evidenced by earlier studies that have documented that many teachers are illiterate to assessment. They might have the knowledge, but in practice, they cannot assess the students appropriately. For example, a study conducted by Alderson (2005) indicated that teachers make unreliable tests which might have a washback effect on the learning process. Another study by Stiggins (2001) reported that unreliable assessment results resulted from teachers' and administrators' lack of assessment literacy affecting students' opportunities to reach their maximum learning potential.

Assessment literacy, by and large, is defined as "knowing appropriate testing practices, acquiring a wide range of assessment techniques, and utilizing tests that accurately assess higher-order concepts. It consists of two aspects: teacher assessment knowledge and teachers' perspectives on assessment knowledge" (Coombe et al., 2012). In its early development, assessment literacy (AL) aimed to examine teachers' assessment literacy in all subjects or courses integrated in the educational institution curriculum. Later on, it has been developed into language teacher assessment literacy (LTAL). To investigate assessment literacy (AL) and language teacher assessment literacy (LTAL), some standardized competencies, questionnaires, and inventory have been devised by institutions and scholars whose expertise dealing with education.

Most research on AL employs quantitative methods as the instruments used to examine AL are based on statistics and psychometrics (Coombe et al., 2020). Initially, the instrument to measure AL was developed by The American Federation of Teachers (AFT), the National Education Association (NEA), and the National Council on Measurement in Education (NCME). This inventory aimed to measure seven areas of assessment competency embedded in "Standards for Teacher Competence in the Educational Assessment of Students" (Plake & Impara, 1997; DeLuca et al., 2016). Since the 1990s, there have been at least eight inventories devised to investigate AL and LTAL (Coombe et al., 2020). However, a specifically designed inventory for university language teachers has not yet fully developed.

There are two reasons why this inventory needs to be developed. Firstly, not all university language teachers are language assessment literate. This is probably due to their educational background. Many university language teachers do not have adequate knowledge of teaching and testing. They may be able to teach, yet they may not have the ability to write a good test and to assess their students appropriately. Prior research has shown that teachers lacked assessment knowledge (Yamtim & Wongwanich, 2014) and a lot of teachers are assessment illiterate indicated by their classroom practices (Djoub, 2017). Secondly, research on LTAL is still rare, particularly on developing a questionnaire or inventory of language assessment literacy for university language teachers which consists of two elements: teachers' assessment knowledge and teachers' perspectives on assessment knowledge. By knowing how literate the university language teachers are to language assessment, it is hoped that they can maximize their

potential to be professional language teachers who are not only able to teach well but they can also competently assess their students' performance.

As indicated by the title of the present paper, it aims to investigate how the university language teachers perceive the dimensions/indicators and the proposition/items of the devised assessment literacy questionnaire for university language teachers.

## **II. THEORETICAL REVIEW**

### **2.1 Assessment Literacy (AL)**

Stiggins (1995) defines assessment literacy as having knowledge of assessment and have the ability to differentiate reliable and unreliable assessment. According to Paterno (2001, as cited in Mertler, pp. 10-11, 2003), assessment literacy is

the possession of knowledge about the basic principles of assessment and evaluation practice which are the terminology of assessment concepts such as test, measurement, assessment, and evaluation, the development, and use of assessment methodologies and techniques in the classroom, familiarity with different tools and apparatus of language assessment, familiarity with standards of quality in classroom assessment...and familiarity with an alternative to traditional measurements of learning.

Mertler & Campbell (2005, p.6) define assessment literacy as “1) understand which assessment methods to use to gather dependable information and student achievement; 2) communicate assessment results effectively, whether using report card grades, test scores, portfolios, or conferences; 3) can use assessment to maximize student motivation and learning by involving students as full partners in assessment, record keeping, and communication (Center for School Improvement and Policy Studies, Boise State University, n.d.)” According to Fulcher (2012), assessment literacy deals with assessment knowledge, skills, and being able to devise, to improve, and to evaluate standardized or classroom tests.

### **2.2 Language Teacher Assessment Literacy (LTAL)**

In terms of language teacher assessment literacy (LTAL), teachers should be able to design, to conduct, to measure, to evaluate, and to make a decision on students' assessment results (Mellati & Khademi, 2018). Concerning English language assessment, assessment literacy encompasses having knowledge of language test reliability and validity, making open- and closed-ended tasks, being knowledgeable of alternative assessments, and how to assess English language learners (Popham, 2009).

### **2.3 Assessment Literacy Inventory (ALI)**

There have hitherto been some inventories devised to examine teacher assessment literacy such as Assessment Literacy Inventory, Assessment Practices Inventory, Assessment Self-Confidence Survey, Assessment in Vocational Classroom Questionnaire, Coombe et al.

Language Testing in Asia, Measurement Literacy, the revised Assessment Literacy Inventory, and the Teacher Assessment Literacy Questionnaire (Coombe et al., 2020).

In the beginning, assessment literacy inventory or questionnaire was developed in the 1990s by The American Federation of Teachers (AFT), the National Education Association (NEA), and the National Council on Measurement in Education (NCME). There are seven competencies that become the standards of assessment literacy namely “1) Choosing assessment methods appropriate for instructional decisions; 2) Developing assessment methods appropriate for instructional decisions; 3) Administering, scoring, interpreting the results of both externally produced and teacher-produced assessment methods; 4) Using assessment results when making decisions about individual students, planning instruction, developing curriculum, and improving schools; 5) Developing valid pupil grading procedures; 6) Communicating assessment results to students, parents, other lay audiences, and other educators; 7) Recognizing unethical, illegal, and other appropriate methods and uses of assessment information.” (Plake & Impara, 1997; DeLuca et al., 2016 ).

These standards were then argued by Stiggins (1999b, as cited in Mertler & Campbell, 2005) because they did not measure real classroom test practices. He further proposed seven areas of assessment literacy represented through the following competencies:

1. Connecting assessments to clear purposes
2. Clarifying achievement expectations
3. Applying proper assessment methods
4. Developing quality assessment exercises and scoring criteria and sampling appropriately
5. Avoiding bias in assessment
6. Communicating effectively about student achievement
7. Using an assessment as an instructional intervention

The questionnaire devised in this study was based on the one developed by Giraldo (2018). The total number of the items is 66 and it consists of three descriptors, i.e. ‘Knowledge’, ‘Skills’, and ‘Principles’. ‘Knowledge’ measures a) awareness of applied linguistics, b) awareness of theory and concepts, and c) awareness of own language assessment context. ‘Skills’ measures a) instructional skills, b) design skills for language assessment, c) skills in educational measurement, and d) technological skills. The last descriptor, ‘Principles’, consists of awareness of and actions towards critical issues in language assessment.

### **III. RESEARCH METHODOLOGY**

The present study employed a descriptive-qualitative method. The objective of a descriptive study is to describe any researchable phenomena and their characteristics. It is also more concerned with ‘what’ instead of ‘how’ or ‘why’ (Gall, Gall, & Borg, 2007). Hence, qualitative research is a research for exploring and understand the meaning individuals or groups ascribe to social or human problem. It means qualitative is a research design where the researcher presenting the data with using description (Pasaribu et al, 2020:15; Sinaga et al,

2020:33). This research deployed a qualitative method as the data analysed were not in figures or numbers and it did not involve statistical programs. (Dornyei, 2007; Hutabarat et al, 2020)

### **3.1 Research Procedures**

The research procedures discuss instrument development, data collection, and data analysis procedures. Data collection describes how the researcher collects the data and data analysis explains how the researcher analyses the data.

### **3.2 Instrument Development**

The present study aims to develop an instrument of Teacher Assessment Literacy questionnaire for university language teachers. The questionnaire was in the rating scale format. Hamzah & Susanti (2020) said that a rating scale is a data collection instrument used to explain, classify, and assess individuals or particular situations. In addition, it is an instrument that can be used to measure people's perceptions of certain phenomena, such as one's social-economic status or one's knowledge and abilities of particular subjects. According to Dörnyei & Taguchi (2010, pp.11-12), to construct a good questionnaire, there are some steps and procedures involved:

1. Deciding on the general features of the questionnaire, such as the length, the format, and the main parts.
2. Writing effective items/questions and drawing up an item pool.
3. Selecting and sequencing the items.
4. Writing appropriate instructions and examples.
5. Translating the questionnaire into a target language if it was not originally written in that language.
6. Piloting the questionnaire and conducting item analysis.

Hamzah and Susanti (2020) suggest some steps to construct a questionnaire:

1. Referring to the variables of the research.
2. Determining the indicators which will be measured from the variables.
3. Writing the items focusing on the indicators which have been determined.

Hamzah and Susanti (2020) further said that determining indicators or dimensions of the research variables should be based on relevant literature, theoretical framework, or previous studies. In this research, piloting the questionnaire and conducting item analysis were not carried out. Therefore, the items of the questionnaire were only evaluated by the expert judgment—university language teachers who have at least 5 years of teaching experience at the tertiary level—and the results were reported through the present study. The method or the step-by-step of developing this questionnaire was based on Mukundan et al.'s (2011) studies on developing an English language textbook evaluation checklist. The first step of the studies was determining the evaluative criteria (Mukundan et al., 2011) and the criteria were evaluated by some experts through an unstructured interview in which the participants brainstormed on the evaluative criteria that should be considered in evaluating ELT textbooks. This was followed by a

structured interview in which the participants were provided with a copy of the checklist. They were free to reword, delete or add items that they considered necessary.

### **3.3 Data Collection Procedures**

Following Mukundan et al.'s studies (2011), to address the research question, the data were supposed to be collected through an unstructured interview. The purpose of the interview is to give feedback on the dimensions/the indicators and the proposition/the items of the questionnaire. However, due to time constraints, the data collection was conducted through the google form application. The application recorded 20 responses having been distributed.

### **3.4 Data Analysis Procedures**

The responses to the questionnaire were analyzed through the following steps.

1. The feedback or comments for each item from the respondents were classified in a table.
2. The feedback was then summarized and discussed in the discussion session
3. The discussion referred to the theoretical framework of language teacher assessment literacy.

## **IV. FINDINGS AND DISCUSSION**

### **4.1 Findings**

The main objective of this study is to develop a questionnaire investigating university language teachers' assessment literacy. It took three months to develop the indicators and the items of the questionnaire. Although there have been language teacher assessment literacy questionnaires developed in previous studies, the one made in this research was simplified and each item of the questionnaire incorporated language assessment terms, such as the differences between assessments and tests, scoring rubric, validity, practicality, and reliability as well as test specifications. Moreover, the items constructed were evaluated descriptively by the respondents of the questionnaire which might be the research gap of this study. Most of the research discussing similar topics did not indicate that the questionnaire developed prior to the present research did not base on any feedback or suggestions from any language teachers or people who have a job relevant to the items of the questionnaire. Table 1 below shows the indicators of the questionnaire from which the items were constructed.

**Table 1. Indicators and Sub-indicators of the Questionnaire**

<b>No.</b>	<b>Indicators</b>	<b>Sub-indicators</b>
------------	-------------------	-----------------------

No.	Indicators	Sub-indicators
1. Knowledge (Giraldo, 2018)		<p>The theories/concept of language assessment and language testing:</p> <ul style="list-style-type: none"> <li>• Definitions of Assessment</li> <li>• Definitions of Tests</li> <li>• Types of assessment</li> <li>• Types of tests</li> <li>• Types of test items</li> <li>• Reliability</li> <li>• Validity</li> <li>• Practicality</li> <li>• Assessment and test score</li> <li>• Scoring rubric</li> <li>• Test specifications</li> </ul>
2. Skills (Giraldo, 2018)		<ul style="list-style-type: none"> <li>• Constructing test /assessment items</li> <li>• Making test instructions</li> <li>• Designing scoring rubric</li> <li>• Weighing assessment/test score</li> <li>• Making test specifications</li> <li>• Measuring test reliability and item validity</li> <li>• The implementation of assessment and test in the classroom</li> </ul>

The indicators above were then developed into 30 items. To avoid misunderstanding when completing the questionnaire (to give feedback and evaluate the items of the questionnaire), it was constructed in the Indonesian language (See table 2)

**Table 2. Sub-indicators and Questionnaire items**

Sub-indicators	Questionnaire items
<p>The theories/concept of language assessment and language testing:</p> <ul style="list-style-type: none"> <li>• Definitions of Assessment</li> <li>• Definitions of Tests</li> <li>• Types of assessment</li> <li>• Types of tests</li> </ul>	1. Saya memahami konsep atau teori asesmen bahasa ('language assessment') secara umum.
	2. Asesmen ('assessment') dan tes ('test') adalah dua konsep yang berbeda.
	3. Asesmen ('assessment') dan evaluasi ('evaluation') adalah dua konsep yang berbeda.
	4. Yang saya kerjakan sehari-hari di dalam kelas adalah melakukan asesmen, bukan memberikan tes.
	5. Asesmen dan tes terbagi atas beberapa jenis.
	6. Saya mengetahui jenis asesmen dan tes.

Sub-indicators	Questionnaire items
	7. Saya memahami jenis asesmen dan tes dan tujuan dari tiap jenis asesmen dan tes tersebut.
	8. Saya memahami tiap jenis asesmen dan tes yang saya berikan didalam kelas.
	9. Pekerjaan Rumah (PR), tugas, dan kuis adalah contoh asesmen.
	10. Rapor merupakan salah satu bentuk asesmen.
	11. Pertanyaan spontan yang ditanyakan saat mengajar dalam kelas termasuk bentuk tes.
	12. Untuk mengetahui kelemahan dan kekuatan mahasiswa dalam kemahiran ('skills') tertentu, kita dapat melakukan 'placement test'.
	13. "Proficiency test" digunakan untuk mengukur kemampuan bahasa mahasiswa secara keseluruhan ('global competence').
	14. 'Summative Assessment' lebih bersifat evaluatif karena tidak bertujuan untuk kekuatan dan kelemahan mahasiswa.
	15. Istilah lain dari 'Summative Assessment' adalah 'Formative Assessment'.
<ul style="list-style-type: none"> <li>Reliability</li> <li>Validity</li> <li>Practicality</li> <li>Assessment and test score</li> <li>Scoring rubric</li> <li>Test specifications</li> </ul>	16. Saya memahami dan mengetahui jenis-jenis 'reliability' dan 'validity' terkait dengan asesmen dan tes.
	17. Penyelenggaraan atau administrasi asesmen terkait dengan konsep 'practicality'.
	18. Pembuatan instruksi asesmen atau tes harus memperhatikan faktor 'reliability'.
	19. Kondisi mahasiswa saat kinerjanya dinilai dapat mempengaruhi 'reliability' dari hasil asesmen.
	20. Penyesuaian soal asesmen dengan silabus atau RPS merupakan implementasi dari 'validity'.
	21. 'Validity' item soal/asesmen dapat dilihat dari mutu item soal.
	22. Bobot nilai/skor soal harus memperhatikan tingkat kesulitan soal.
	23. Semakin sulit item/soal tes, nilai/skor harus makin tinggi.
	24. Saya mengetahui dan memahami jenis-jenis 'scoring rubric'.
	25. Salah satu fungsi 'scoring rubric' adalah untuk mengurangi bias pengajar saat menilai kinerja mahasiswa.

Sub-indicators	Questionnaire items
	26. Sebelum membuat soal atau melakukan asesmen, pengajar dianjurkan membuat 'test specifications' terlebih dahulu.
	27. 'Test specifications' diperlukan agar item asesmen atau tes mengukur apa yang seharusnya diukur dan tepat sasaran.
	28. Item soal sebaiknya disusun dari yang mudah sampai yang paling sulit.
	29. Saya mampu membuat scoring rubric, baik yang berbentuk 'analytic' maupun 'holystic'.
	30. 'Item formats' dan 'task formats' adalah dua konsep yang berbeda dalam asesmen.

The instruction of a questionnaire is a crucial part as it may affect the results of a study. To ensure that the respondents of the present research understood the instruction, it was written in the Indonesian language. The instruction of the questionnaire is as follows.

### Language Teacher Assessment Literacy Questionnaire

Mohon instruksi di bawah ini dibaca dengan cermat.

Izinkan saya memperkenalkan diri terlebih dahulu. Nama saya Nurdiana, dosen tetap prodi Bahasa dan Budaya Inggris, Universitas Bunda Mulia, Jakarta. Saat ini saya sedang melakukan penelitian dengan judul 'Developing Teacher Assessment Literacy Questionnaire for University Language Teachers'.

Tujuan penelitian ini adalah untuk mengembangkan kuesioner 'Teacher Assessment Literacy for University Language Teachers' dengan cara meminta responden penelitian memberikan umpan balik ('feedback') terhadap item kuesioner. Jadi, responden **diminta untuk menjadi EVALUATOR item kuesioner yang sedang dikembangkan agar kuesioner tersebut menjadi 'reliable' dan 'valid'**.

Untuk itu, saya membutuhkan bantuan bapak/ibu untuk menjadi evaluator item kuesioner dengan cara sebagai berikut:

**1. Bacalah dengan cermat setiap item kuesioner.**

**2. Berikan umpan balik ('feedback') terhadap setiap item dengan cara:**

- a. merevisi item SESUAI atau BERDASARKAN PEKERJAAN/KEGIATAN bapak/ibu sebagai PENGAJAR BAHASA di UNIVERSITAS**
- b. jika menurut bapak/ibu item tersebut tidak perlu direvisi, bapak/ibu tidak perlu mengisi apapun di kolom yang sudah disediakan di bawah setiap item kuesioner.**
- c. bapak/ibu boleh memberikan saran dalam bentuk membuat item yang baru di kolom saran.**

The following table (table 3) is the responses to the questionnaire recorded in the google form application. Unfortunately, some of the respondents did not give expected responses. Therefore, only the relevant ones were reported in the table below.

**Table 3. Feedback on Questionnaire Items**

No.	Items	Feedback/comment/evaluation
1.	Saya memahami konsep atau teori asesmen bahasa ('language assessment') secara umum.	R1: Menurut saya tidak perlu dibuat dalam dua bahasa seperti assessmen bahasa (language assessment) karena istilah itu sudah umum. Apalagi target responden dosen
2.	Asesmen ('assessment') dan tes ('test') adalah dua konsep yang berbeda.	R1:Asesmen dan tes adalah dua sistem penilaian yang berbeda. R2: saya memahami Asesmen ('assessment') dan tes ('test') adalah dua konsep yang berbeda
3.	Asesment ('assessment') dan evaluasi ('evaluation') adalah dua konsep yang berbeda.	R1: Asesmen dan evaluasi adalah dua sistem penilaian yang berbeda. R2: saya memahami Asesment ('assessment') dan evaluasi ('evaluation') adalah dua konsep yang berbeda.
4.	Yang saya kerjakan sehari-hari di dalam kelas adalah melakukan asesmen, bukan memberikan tes.	R1: Sepertinya perlu disesuaikan dengan masa sekarang bu. Karena daring, baik asesmen atau tes, mirip". Mungkin redaksinya bisa lebih dikerucutkan yang dikerjakan itu apa. R2: Terkadang saya melakukan asesmen dengan memberikan tes R3: Mengapa setelah asesmen koma?
5.	Asesmen dan tes terbagi atas beberapa jenis.	R1: Menurut saya tidak perlu bu. R2: Item ini kurang sesuai untuk skala Likert. Terlebih lagi ada dua variabel pada item ini (asesmen dan tes). R3: Mohon dispesifikan jenisnya R4: Sebutkan jenisnya secara lbh rinci R5: ada beberapa jenis assessmen dan tes R5: Item ini kurang sesuai untuk skala Likert. Terlebih lagi ada dua variabel pada item ini (asesmen dan tes)

No.	Items	Feedback/comment/evaluation
6.	Saya mengetahui jenis asesmen dan tes.	R1: Menurut saya bisa langsung ditanyakan misalnya "Berikut yang merupakan jenis asesmen adalah" untuk yang tes, sebaiknya dipisah. Namun, pilihan untuk dua butir tsb disamakan. R2: Saya tahu dan paham ada beberapa jenis asesmen dan tes R3: Ada dua variabel pada item ini. Sebaiknya variabelnya dipisah menjadi dua pertanyaan yang berbeda. R4: Saran untuk membaca ulang konsep tersebut sehingga dapat membuat item ini lebih jelas R5: da dua variabel pada item ini. Sebaiknya variabelnya dipisah menjadi dua pertanyaan yang berbeda.
7.	Saya memahami jenis asesmen dan tes dan tujuan dari tiap jenis asesmen dan tes tersebut.	R1: Mirip" dengan nomor 8 bu. Mungkin nomor 8 bisa diganti dengan "Saya memahami tujuan dari....." Bu R2: Saya memahami asesmen, tes dan tujuan masing-masing asesmen dan tes. R3: Membingungkan R4: Langsung tujuan saja, jenis sdh di no. 6. R5:
8.	Saya memahami tiap jenis asesmen dan tes yang saya berikan didalam kelas.	R1: Mungkin nomor 8 bisa diganti dengan "Saya memahami tujuan dari....." Bu karena yang ini sudah di-cover di nomor 7 R2: Apakah tidak mirip dengan no. 7?
9.	Pekerjaan Rumah (PR), tugas, dan kuis adalah contoh asesmen.	
10.	Rapor merupakan salah satu bentuk asesmen.	
11.	Pertanyaan spontan yang ditanyakan saat mengajar dalam kelas termasuk bentuk tes.	
12.	Untuk mengetahui kelemahan dan kekuatan mahasiswa dalam kemahiran ('skills') tertentu, kita dapat melakukan 'placement test'.	R1: Konteksnya di awal pembelajaran kah?
13.	"Proficiency test" digunakan untuk mengukur	R1: Mengapa diberikan tanda petik?

No.	Items	Feedback/comment/evaluation
	kemampuan bahasa mahasiswa secara keseluruhan ('global competence').	
14.	'Summative Assessment' lebih bersifat evaluatif karena tidak bertujuan untuk kekuatan dan kelemahan mahasiswa.	R1: Mengapa diberikan tanda petik
15.	Istilah lain dari 'Summative Assessment' adalah 'Formative Assessment'.	R1: Item ini kurang cocok untuk skala Likert.
16.	Saya memahami dan mengetahui jenis-jenis 'reliability' dan 'validity' terkait dengan asesmen dan tes.	R1: Sebaiknya variabel-variabelnya dipisah menjadi pertanyaan individu. R2: Untuk reliability 1 nomer, validity 1 nomer.
17.	Penyelenggaraan atau administrasi asesmen terkait dengan konsep 'practicality'.	R1:Mungkin bisa lebih dijelaskan apa yang dimaksud dengan 'practicality' R2: Belum jelas arah pertanyaan
18.	Pembuatan instruksi asesmen atau tes harus memperhatikan faktor 'reliability'.	R1:Mungkin bisa lebih dijelaskan apa yang dimaksud dengan 'reliability' R2: Idem mengapa tanda petik
19.	Kondisi mahasiswa saat kinerjanya dinilai dapat mempengaruhi 'reliability' dari hasil asesmen.	
20.	Penyelarasan soal asesmen dengan silabus atau RPS merupakan implementasi dari 'validity'.	R1:Kenapa tanda petik
21.	'Validity' item soal/asesmen dapat dilihat dari mutu item soal.	R1:'Validity' item dapat diketahui dari mutu item. R2:Mungkin bisa lebih dijelaskan apa yang dimaksud dengan 'validity'
22.	Bobot nilai/skor soal harus memperhatikan tingkat kesulitan soal.	R1:Bobot nilai harus memperhatikan tingkat kesulitan item. R2:penentuan Bobot nilai/skor soal harus memperhatikan tingkat kesulitan soal.
23.	Semakin sulit item/soal tes, nilai/skor harus makin tinggi.	R1:Tidak perlu seperti itu.
24.	Saya mengetahui dan memahami jenis-jenis 'scoring rubric'.	R1:Bisa dibahas Indonesia kan menjadi rubrik penilaian
25.	Salah satu fungsi 'scoring rubric' adalah untuk mengurangi bias pengajar saat menilai kinerja mahasiswa.	
26.	Sebelum membuat soal atau melakukan asesmen, pengajar dianjurkan membuat 'test specifications' terlebih dahulu.	R1: Mengapa tanda petik?

No.	Items	Feedback/comment/evaluation
27.	'Test specifications' diperlukan agar item asesmen atau tes mengukur apa yang seharusnya diukur dan tepat sasaran.	
28.	Item soal sebaiknya disusun dari yang mudah sampai yang paling sulit.	
29.	Saya mampu membuat scoring rubric, baik yang berbentuk 'analytic' maupun 'holystic'.	R1:Cek ejaan holystic atau holistic? R2: Tidak semua dosen punya ability membuat scoring rubric R3: Pisahkan nomernya antara analytic dan holystic.
30.	'Item formats' dan 'task formats' adalah dua konsep yang berbeda dalam asesmen.	

**R= Respondents**

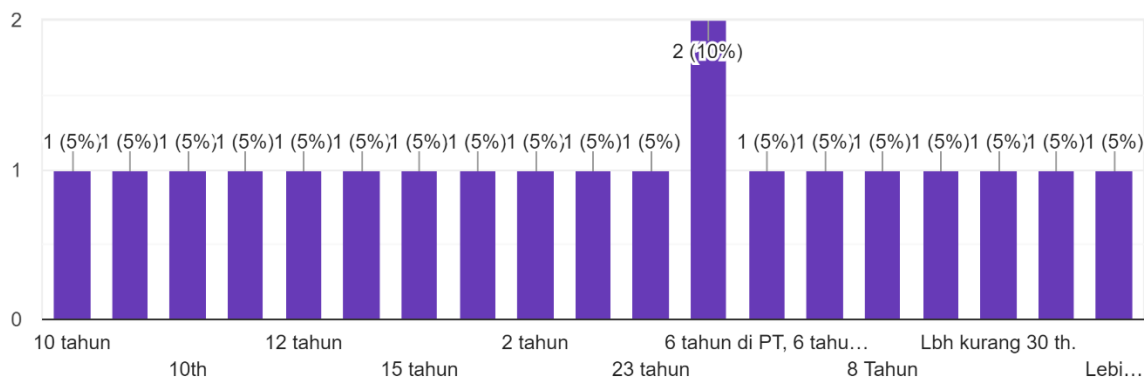
Besides constructing 30 items, 2 closed-ended questions and 2 open-ended questions were embedded in the questionnaire. The two 2 closed-ended questions are:

1. Sudah berapa lama anda mengajar?
2. Kemahiran ('skills') dan aspek bahasa ('language aspects') apa yang anda ajarkan?

The graph below (Graph 1) indicates that most of the respondents have been teaching at the university for more than 5 years. There was only one (1) respondent who has been teaching a language for two years.

Sudah berapa lama anda mengajar?

20 responses

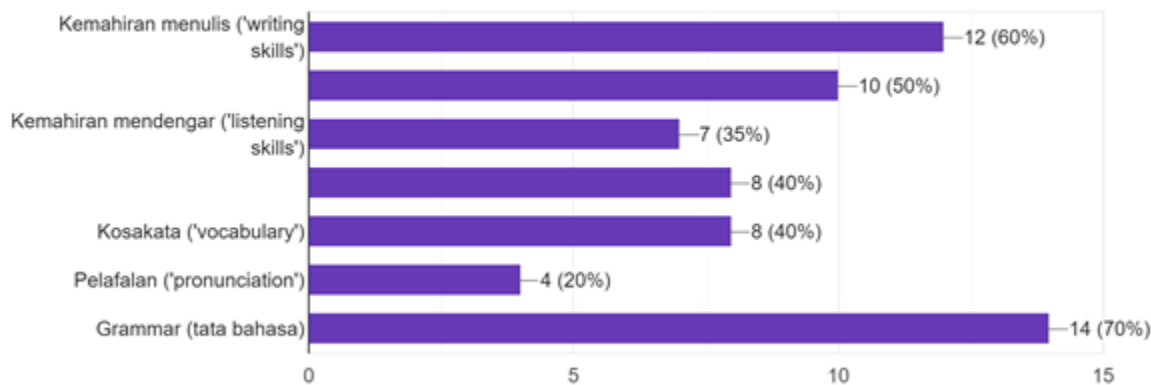


**Graph 1. Teaching Experience**

Meanwhile, the following graph (Graph 2) shows the skills and the language aspects the respondents teach.

Kemahiran ('skills') dan aspek bahasa ('language aspects') apa yang anda ajarkan?

20 responses



**Graph 2. Skills and Language Aspects Taught**

It can be seen in the chart that 70% of the respondents teach grammar. The least percentage, 20%, indicates that there were just a few respondents who teach pronunciation.

The two open-ended questions/items are as follows.

1. Apa persepsi anda terhadap kuesioner yang sedang dikembangkan ini?
2. Saran anda untuk item kuesioner ini secara keseluruhan.

Below (Table 4) is the answers to question 1.

**Table 4. Responses to Question 1**

<b>Apa persepsi anda terhadap kuesioner yang sedang dikembangkan ini?</b>
1. Itemnya cukup baik.
2. Bagus namun sebaiknya multiple choices saja.
3. Bagus
4. Refleksi utk pengajar dalam melakukan pembelajaran dan penilaian serta evaluasi.
5. Item pada kuesioner sdh cukup detil utk mencari topik yg diteliti.
6. Saya kurang paham yang dimaksud dengan persepsi disini
7. Beberapa item terasa janggal karena seolah-olah berupa pertanyaan benar atau salah (true or false), bukan pertanyaan opini yang bisa dicari tahu melalui skala Likert. Bahasa yang digunakan agak teknis sehingga akan membingungkan responden yang tidak mendalami ilmu asesmen.
8. Untuk bahan evaluasi diri
9. Cukup bagus
10. Terkadang bingung, ekspektasi jawabannya apa
11. Pertanyaan nya seperti pertanyaan yang harus mempelajari definisi-definisi istilah (term) tentang assesment di pengajaran.
12. Perlu dikaji ulang untuk setiap item. Untuk menguji validitas sebaiknya diberikan kepada ahli assessment.
13. Sudah cukup bagus
14. Berhubung background saya bukan pengajaran, banyak item2 yang terlalu spesifik dan kurang saya pahami
15. Sudah bagus.
16. Cukup baik, karena semua aspek mengenai assessment di tanyakan

The following table (Table 5) is the respondents' suggestions for the questionnaire devised.

**Table 5. Respondents' Suggestions for the Items of the Questionnaire**

<b>Saran anda untuk item kuesioner ini secara keseluruhan.</b>
1. Sudah ada di butir kuesionernya
-
2. Lebih pada contoh kasus yg dijumpai dan jawaban dalam bentuk pilhan ganda.
3. Utk terminologi yg spesifik sebaiknya diberikan keterangan utk menyamakan persepsi responden.
4. Seharusnya ada pilihan untuk menjawabnya, mis: Ya, Tidak, dan Tidak Tahu
5. Item-item yang ada perlu dipertimbangkan kembali.

6. Tolong petunjuknya agar lebih jelas ya....
7. pertanyaan lebih singkat.
8.1. Sebaiknya itemnya disesuaikan dg response yg diharapkan. Kadang ada yg butuh MC (T/F or bbrp opsi), short response, long essay, ato model Likert. Tdk semua, your answer. Kalo bukan pertanyaan, hanya statements, trs dijawab apa? 2. No. 1 itu jg dg pertimbangkan efisiensi & kesesuaian response. 3. Ada item yg kurang jelas konsepnya, e.g. No 15. Summative & Formative itu tdk sama. No. 12, I think diagnostic test. K1 placement itu utk mengelompokkan kemampuan testees. 4. Angket ini perlu lbh jls tujuannya. Apakah utk mengukur pengetahuan respondents ttg Assessment, ato mengukur kejelasan pengkalimatan items, ato apa? Item2nya campur, tdk menggambarkan tujuan yg jelas. 5. Maaf & terima kasih.
9. Perlu dibuat lebih runtut dan lengkap lagi
10. Sepertinya pertanyaan nya bukan open-ended. Mungkin kalau bentuknya option atau multiple choice lebih enak jawabnya.
11. Perlu diperbaiki kembali
12. Sebaiknya tambahkan dengan item tentang alternatives in assessment, standardized testing, standardized based assessment.
13. Apabila ini dikhususkan utk org2 dgn background pengajaran mgk tidak terlalu masalah, tp apabila ditujukan utk org2 lain dgn background bukan pengajaran akan sedikit sulit utk memahaminya dengan baik
14. Disusun secara sistematis mulai dr yg umum ke yg lbh khusus.
15. Sudah bagus.
16. Disusun secara sistematis mulai dari yang umum ke yang lebih khusus.

## 4.2 Discussions

Based on the results and the findings of this study, it was not an easy assignment to determine the indicators and to construct the items of the questionnaire. It was time-consuming and the dictions should be carefully taken into account. The respondents' feedback to the items, their perception towards the items and the questionnaire as well as their suggestions for the questionnaire indicate that this questionnaire obviously needs a lot of improvement, particularly on the items and the questionnaire. There are several categories of what should be enhanced to make the questionnaire valid and reliable.

### 1. The instruction on how to complete the questionnaire.

Many responses from the respondents gave unexpected responses such as 'Ya', 'Tidak', 'Belum' whereas they were not required to 'answer' the item. What they had to was they have to EVALUATE or give FEEDBACK on each item of the questionnaire. This implies that in the future, the instruction on how to complete the questionnaire should be revised.

### 2. The respondents should be informed of the objective of the questionnaire devised.

It turned out that the aim of the questionnaire should have been informed to avoid unexpected responses. This seems to be significant input as there was a respondent who said that

the objective of why this questionnaire was developed was unclear. Thus, it might have led the respondents to confusion when completing the questionnaire.

### 3. The type/the format of the questionnaire

One of the findings suggests the format of the questionnaire. It was supposed to be in Likert Scale format; however, based on the respondents' suggestions, other types/format of answer choices may be in multiple-choice, true or false (T/F), short/long response, or essays.

The questionnaire referred to in this study was the one developed by Giraldo (2018). Below are the samples of the indicators/dimensions and the items of the questionnaire.

Knowledge	
<b>Awareness of applied linguistics</b>	
1	<i>Compares</i> approaches for language teaching and assessment; e.g., communicative language testing; task-based assessment.
2	<i>Explains</i> major issues in applied linguistics; e.g., bilingualism, language policy and planning, pragmatics, sociolinguistics, etc.
3	<i>Analyzes</i> trends in second language acquisition and their impact on language assessment; e.g., motivation, cross-linguistic influence, learner strategies.
4	<i>Integrates</i> theories related to language and language use; e.g., models of language ability, discourse analysis, and grammar teaching.
<b>Awareness of theory and concepts</b>	
5	<i>Illustrates</i> history of language testing and assessment, and its impact on current practices and society.
6	<i>Interprets</i> reliability in language assessment and its implications: dependability, classical test theory, item analysis, threats, calculating reliability of tests and items, inter- and intra-rater reliability, etc.
7	<i>Interprets</i> validity in assessment and its implications: construct, content, and criterion validities, construct validity as unitary, Messick's (1989) consequential validity; validity as argument.
8	<i>Calculates</i> statistics procedures for investigating validity such as Pearson Product Moment Correlation (PPMC).
9	<i>Interprets</i> major qualities for language assessment practices (apart from reliability and validity), and their implications for language assessment: authenticity, practicality, interactiveness, fairness, ethics, and impact (including washback).
10	<i>Computes basic statistical analyses:</i> mean, mode, median, range, standard deviation, score distribution, etc.
11	<i>Differentiates</i> concepts related to assessment paradigms: traditional versus alternative; norm-referenced and criterion-referenced testing.
12	<i>Differentiates</i> major purposes and related decision-making for language testing: placement, achievement, proficiency, etc.
13	<i>Explains</i> major steps in developing tests: test purpose, construct definition, content specifications, test specifications, etc.
<b>Awareness of own language assessment context</b>	
20	<i>Explains</i> own beliefs, attitudes, context, and needs for assessment.
21	<i>Evaluates</i> the test and assessment policies that influence his/her teaching.
22	<i>Assesses</i> the existing tensions that influence language assessment in his/her school.
23	<i>Illustrates</i> the general guidelines and policies that drive language learning and assessment in his/her context; for example, type of language curriculum.
24	<i>Criticizes</i> the kind of washback assessments usually have on his/her teaching context.

## Skills

### **Instructional skills** has the ability to:

25	align curriculum objectives, instruction, and assessment.
26	plan, implement, monitor, record, and report student language development.
27	provide feedback on students' assessment performance (norm- and criterion-referenced).
28	collect formal data (e.g., through tests) and informal data (while observing in class) of students' language development.
29	improve instruction based on assessment results and feedback.
30	utilize alternative means for assessment; for example, portfolios.
31	use language assessment methods appropriately: to monitor language learning and nothing else.
32	provide motivating assessment experiences, giving encouraging feedback, or setting up self-assessment scenarios.
33	communicate norm- and criterion-referenced test results to a variety of audiences: students, parents, school directors, etc.
34	use multiple methods of assessment to make decisions based on substantive information.
35	incorporate technologies in assessing students.

### **Skills in educational measurement (advanced skills not always needed)** has the ability to:

48	interpret data from large-scale tests, namely descriptive statistics such as means, modes, medians, bell curves, etc.; has the ability to calculate descriptive statistics.
49	infer students' strengths and weaknesses based on data.
50	criticize external tests and their qualities based on their psychometric characteristics.
51	interpret data related to test design, such as item difficulty and item discrimination.
52	calculate reliability and validity indices by using appropriate methods such as Kappa, PPMC, and others.
53	investigate facility and discrimination indices statistically.

### **Technological skills** has the ability to:

54	use software such as Statistical Package for the Social Sciences.
55	run operations on Excel; for example, descriptive statistics and reliability correlations.
56	use internet resources such as online tutorials and adapt contents for his/her particular language assessment needs.

## Principles

### **Awareness of and actions towards critical issues in language assessment**

57	Clearly informs the inferences and decisions that derive from scores in assessments.
58	Uses assessment results for feedback to influence language learning, not other construct-irrelevant sources (e.g., personal bias towards a student).
59	Treats all students, or users of language assessment, with respect.
60	Uses tests, test processes, and test scores ethically.
61	Provides assessment practices that are fair and non-discriminatory.
62	Critiques the impact and power standardized tests can have and has a stance towards them.
63	Observes guidelines for ethics used at the institution in regard to language assessment.
64	Criticizes external tests based on their quality and impact.
65	Implements transparent language assessment practices; informs students of the what, how, and why of assessment.
66	Implements democratic language assessment practices, by giving students opportunities to share their voices about assessment.

According to Giraldo (2018), the list above can be used in five methods.

1. Using the list/the questionnaire in Yes/No checklist.
2. As an observation sheet when observing language teachers.
3. As a need analysis checklist to examine what language teachers need to improve as part of their professional development.
4. Using the list to see what teachers need to think about new ideas or activities.
5. Using the list as an overview of language assessment literacy.

In conclusion, language teachers may use the list for their purposes. Regarding the list of the items devised in the present study, therefore, the items can be utilized similarly in a way Giraldo (2018) proposed for his list.

## V. CONCLUSION

Designing language teacher assessment literacy is a challenging assignment. It obviously requires knowledge of language assessment and the objectives of the questionnaire should be crystal clear and the diction of each item should be taken into account to avoid misunderstanding when completing the questionnaire.

Not only should the items and indicators be paid attention to, but the format of how to complete the questionnaire also needs to weigh up whether it is used as Yes/No list, True or False, Multiple-choice, or it is used in the Likert Scale format. It all depends on teachers' purposes.

Suggestions for future research mostly come from the responses to the questionnaire. There are significant recommendations for the present study. Firstly, the propositions or the items must be revised. Secondly, the instruction of the questionnaire completion (how to complete the questionnaire) should be fixed or should be allowed in a suitable and various format. Thirdly, the objectives of the questionnaire must be crystal clear to prevent misunderstanding when working on the questionnaire. Therefore, this study should be carried on in order to enhance the list of items and to make the questionnaire more reliable and valid so that language teachers can take advantage of the questionnaire.

## REFERENCES

- [1] Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London, UK: Continuum.
- [2] Al-Malki, M. A. & Weir, K. (2014). A comparative analysis between the assessment criteria used to assess graduating teachers at Rustaq College (Oman) and Griffith University (Australia) during the teaching practicum. *Australian Journal of Teacher Education*, 39(12), 28-42.
- [3] Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: issues, challenges, and recommendations. In O'Sullivan, B., Stoyanoff,

- S. Davidson, P. & Coombe, C. (Eds.), *The Cambridge Guide to Second Language Assessment*. Cambridge: Cambridge University Press
- [4] Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*. 10:3
- [5] Deluca, C., Lapointe-McEwan, D. & Luhanga, U. (2016). Approaches to classroom inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21:4, 248-266.
- [6] Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari, & V. Thakur (Eds.), *Revisiting EFL Assessment: Critical perspective* (pp. 9–27)
- [7] Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- [8] Dörnyei, Z. & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge
- [9] Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132
- [10] Gall, M.D., Gall, J.P., and Borg, W.R. (2007), *Educational research: An introduction* (8th ed.). Boston: Pearson.
- [11] Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues Teach. Prof. Dev.*, Vol. 20 No. 1, pp.179-195
- [12] Hamzah, A. & Susanti, L. (2020). *Metode penelitian kualitatif*. Malang: Literasi Nusantara
- [13] Hutabarat, E., Herman, Silalahi, D.E., and Sihombing, P. S. R. (2020). An Analysis of Ideational Metafunction on News Jakarta Post about Some Good Covid-19 Related News. *VELES Voices of English Language Education Society*, Vol 4, No 2 (2020), e-ISSN 2579-7484, PP. 142-151.
- [14] Mackey, A. & Gass, S.M. (2005). *Second language research*. London: Lawrence Erlbaum Associates, Inc.
- [15] Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education*, 43(6).
- [16] Mertler, C.A. (2003). Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference? *Paper presented at the Annual Meeting of the Mid-Western Educational Research Association* (Columbus, OH, October 15-18)
- [17] Mertler, C.A. & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory. *Paper presented at the annual meeting of the American Educational Research Association*, Montréal, Quebec, Canada, April 11–15
- [18] Mukundan, J., Hajimohammadi, R. & Nimehchisalem, V. (2011). Developing an English language textbook evaluation checklist. *Contemporary Issues in Education Research*, Volume 4, Number 6
- [19] Mukundan, J., Hajimohammadi, R. & Nimehchisalem, V. (2011). Developing an English language textbook evaluation checklist: A Focus Group Study. *International Journal of Humanities and Social Science* Vol. 1 No. 12

- [20] Pasaribu, B., Herman, and Hutahaeen, D. T. (2020). Students' Difficulties In Translating Narrative Text From English Into Indonesia At Grade VIII Of SMP Negeri 9 Pematangsiantar. *Acitya: Journal of Teaching & Education*, Vol. 2 No. 1 2020, PP. 12-18.
- [21] Plake, B.S. & Impara, J.C. (1997). Teacher assessment literacy: What do teachers know about assessment? *Handbook of Classroom Assessment*. Academic Press, Inc.
- [22] Popham, W. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4-11.
- [23] Sinaga, H., Herman., and Hutaurok, B. S. (2020). Students' Difficulties in Using Personal Pronouns in Writing Recount Text. *Scientia: Jurnal Hasil Penelitian*, 5(1), 29-36. <https://doi.org/10.32923/sci.v5i1.1341>
- [24] Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- [25] Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.
- [26] Susuwele-Banda, W. J. (2005). Classroom assessment in Malawi: Teachers' perceptions and practices in Mathematics. *Unpublished Doctoral Dissertation*, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- [27] Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, 116, 2998–3004.