Development of Learning Evaluation Tools for Bahasa Indonesia with Rebecca M. Valette Taxonomy and CLIL Approach

Subyantoro[a], Haryadi[b], Mochamad Rizqi Adhi Pratama[c], [a,b]Indonesian Language and Literature Department, Faculty of Language and Art, Semarang State University, Central Java, Indonesia, [c]Department of English Literature, Universitas Ngudi Waluyo, Central Java, Indonesia, Email: [a]bintoro@mail.unnes.ac.id, [b]haryadihar67@mail.unnes.ac.id, [c]rizqi.adhi@unw.ac.id

The purpose of this study is to describe the characteristics of the development of learning evaluation tools for Bahasa Indonesia using the Rebecca M. Valette taxonomy and CLIL approach based on teacher perceptions. The study was quantitative, and the sample consisted of teachers at a selected high school in the Bahasa region of Indonesia. Data was collected through a survey and an observation. The survey was completed to investigate students' proficiency based on a questionnaire, and the observation included an observation on the teacher's perception of competence of students in class XI of the senior high school. The study found that the characteristics of the evaluation tool need to produce short story texts with the taxonomy of Rebecca M. Valette. The CLIL approach based on teacher perceptions is motivated by the principles of developing evaluation tools that can be grouped into language, cultural, literary, and communication. This study wants to want to make the following recommendations. Firstly, researchers could conduct further research on the development of evaluation tools at the elementary, junior high, or high school. Secondly, policymakers should use it as a reference in the preparation of evaluation tools for Indonesian subjects at all levels of elementary, junior high, or high school.

Key words: evaluation tool, CLIL approach, short story text, taxonomy Rebecca M. Valette.

Introduction

Basic competencies about understanding, comparing, analyzing, and evaluating short story texts are still poorly understood by teachers and students. The skills to follow the structure and rules, compare, analyze, and evaluate short story texts are only receptive. The assessment of competencies and abilities has not yet been developed based on the aspects of the 2013 curriculum assessment. It is still difficult for teachers to create short story text material presented in the teacher's handbook. Additionally, there is an absence of a guide or a teacher's manual in line with the 2013 curriculum in evaluating the short story text. Therefore, it is necessary to develop an evaluation tool in the form of a guidebook to help teachers.

Professionals have not sufficiently evaluated teacher guides at ELT. The purpose of this study is to investigate the use of teacher guides to assess their usefulness. This research is a qualitative study, and the data were collected through observations from 15 English classes and interviewed 17 English teachers at Firdausi University College in Mashhad, a city in northeastern Iran. The use of teacher guides is different from the teacher in general. Although teachers use different guidelines, they all want better, more logical ones (Zabihi & Tabataba'ian, 2011).

Based on the evaluation problem, it appears that the form of evaluation using Bloom's taxonomy is not in line with the 2013 curriculum because it places more emphasis on cognitive aspects that are no longer relevant. Bloom's Taxonomy has not been able to identify the elements of skills and issues of attitudes carried out by students specifically (Bloom, 1956). Rebecca M. Valette's taxonomy is a taxonomy that has been specified for language learning. This taxonomy is adapted from the revised Bloom taxonomy. Rebecca M. Valette's taxonomy consists of three domains, namely cognitive, affective, and psychomotor, adapted to the needs of language learning in the 2013 curriculum. In line with the mandate of the 2013 curriculum, the assessment must be done by involving the three aspects, namely cognitive, psychomotor, and affective (religious, social, and cultural attitudes). The three domains are then broken down into fifteen indicators, which are grouped back into four parts or fields. The four sections cover the realm of knowledge and understanding included the following:

- The first domain is the realm of knowledge and understanding. The teacher can do this part in the stages of building the context of students through listening and interpreting activities.
- The second domain is the domain of knowledge and production and includes speaking and writing activities.
- Between the first and second domains, there is a third domain, the manipulation section. In this section, the teacher can ask students to memorize relevant material or provide an evaluation. The three realms combine cognitive and psychomotor aspects.
- The three domains on cognitive and psychomotor aspects ended with evaluation activities, which emphasized more on literary discourse.

The taxonomy of assessment should also have balanced linguistic and literary content. Therefore, language learning does not take place boringly because it deals with language problems. Thus, the taxonomy of language learning is needed to accommodate these needs. The taxonomy of knowledge that can meet several requirements in language learning, as described, is the taxonomy proposed by Rebecca M. Valette.

Rebecca M. Valette's taxonomy also has a fairly complete content or material content. It is indicated by the elaboration of eight material areas, which are grouped into four main objectives. The first purpose is language purpose. For this purpose, there are three materials, namely oral language, written language, and body language (kinesthetic learning). The second purpose is cultural. This educational purpose includes ways to live in a civilized society, knowledge of the love of the motherland, and art. The third is the purpose of literature. The fourth is the purpose of communication. Therefore, in language learning, we should be required to be proficient in communicating, knowing the culture, understanding the relationship of one language to another language, knowing the comparison of one word with another language, and knowing the community or people who have different names. (Valette, 1969, 1997).

Communication skills are an important part of the learning process, including learning communication in languages (Kasiyarno & Apriyanto, 2020). The communication or speaking ability test in a given class must be in line with the learning objectives (Anum & Apriyanto, 2019; Hidayat & Apriyanto, 2019). Some teachers will tend to ignore the actual open competence if the assessment is not carried out based on a communicative approach. They will also ignore class assessments that have an impact on and significantly influence student learning achievement. Likewise, if the measured aspect is not in line with what should be measured, it can lead to meaningless learning (Burrows, 2001; Valette, 1973).

Additionally, the learning approach must also be determined appropriately and based on the 2013 curriculum. Rethinking language assessment must be adjusted to the progress of language learning theory. Therefore, learning Bahasa Indonesia and literature using the CLIL approach is a learning foundation. This is in line with text-based 2013 curriculum learning. Content and Language Integrated Learning (CLIL) is an approach that integrates the teaching of content from a curriculum of subjects with non-native language lessons. In CLIL, students are allowed to learn to think about how to use language, not just to think about language as the main focus of learning. CLIL has four basic components, namely, content, communication, cognition, and culture. Moreover, the selection of this approach is also very compatible with Rebecca M. Valette's taxonomy. Four basic components in the CLIL approach are side by side and complement each other with the learning taxonomy proposed by Rebecca M. Valette, namely social, communication, and culture. (Coyle, 2005; Khomsah & Subyantoro, 2019; Lesca, 2012; McNamara & Lumley, 1997; Mehisto, 2012).

An evaluation tool must also provide assessments that have a positive impact and can overcome differences in cultural backgrounds and provide a genuinely valid assessment system for its users. This positive impact can be seen by testing the validity and reliability of the tool. This course will be useful to improve student learning outcomes. Furthermore, evaluation tools must be practical so that the results of language learning can be maximized (Cismas, 2009; Messick, 1996; Saif, 2006).

The evaluation tools for Bahasa Indonesia compiled in this study ease the teachers to collect data or values of students' abilities (cognitive, attitudes, and psychomotor) directly or indirectly. The concept of evaluation is a process of determining values by considering the results of observations or data collections. Additionally, in evaluating the ability of students, two things must be considered by teachers, namely the learning process and learning outcomes. Both must be balanced so that the learning objectives of language skills can be achieved to the maximum. In evaluating students, the teacher does see not only the final results of student learning but also the student learning process. Moreover, this evaluation tool must be strengthened by developing a learning strategy that combines POL (Project-Oriented Learning) and formative assessment to establish professional competence among students (Boggino, 2009; Cizek, 2005; López-Pastor et al., 2010).

The text to complete this evaluation tool was a short story text and helped to enhances the ability of students to learn four skills, namely listening, speaking, reading, and writing. Short story texts are also more productive and fun because motivation is embedded in the story and is useful for students' pedagogical aspects. Short story text is a valuable and interesting material to improve students' language skills. Furthermore, with short stories, teachers can teach literature and culture. Competence is producing short stories that will be developed by students. Creating the short story text is also expected to be able to have a positive impact on students (Al-Dersi, 2013; Erkaya, 2005; Khatib & Seyyedrezaei, 2017; Pardede, 2011; Saleh Khalaf Ibnian, 2010).

The term competency in producing short story text is a term in the 2013 curriculum which comes from the learning writing skills of a short story. Learning writing skills of a short story based on the genre of the text is an active learning technique as a strategy to improve students' knowledge and ability in writing. Additionally, practical writing skills are highly needed to participate in the community and bring benefits to develop more specific writing skills, for example, mother tongue (De Smedt & Van Keer, 2014; Kusumaningsih, 2013; Pocinho et al., 2014).

Based on these statements, it can be revealed that the use of an appropriate taxonomy with short story text learning and the CLIL approach that integrates content teaching from the curriculum of subjects with non-native language teaching becomes very compatible with the implementation of the 2013 curriculum. The existence of the four basic components in the

CLIL approach is very appropriate if it is applied side by side with the learning taxonomy put forward by Rebecca M. Valette. This is because some of the elements have the same and complementary goals.

Therefore, this study seeks to answer the following main- and sub-research questions:

How can learning evaluation tools be developed through Rebececca M Vallette's taxonomy and CLIL approach?

The main research question was the foundation of the study, and the main aim was to develop an evaluation tool for Bahasa Indonesia, especially short story texts with the taxonomy of Rebecca M. Valette and the CLIL approach for the 2013 curriculum in senior high school.

Methods

This study employed an R&D approach (Research and Development) with research steps, namely (1) phase I (preliminary research and initial development), (2) phase II (development), (3) Phase III (Testing and Validation). The research data include the data on the needs of developing evaluation tools according to teacher perceptions, data on expert assessment results, and limited trial data in the form of scores. The sample of this study were teachers at selected high schools in the Bahasa region of Indonesia. Data collection is done through surveys and observations. The survey was carried out to investigate students' proficiency in languages and the observation on the teacher's perception of competence of students in class XI of the senior high school. Data were grouped into three groups and analysed in the three groups.

Results and Discussion

The results of this study are the characteristics of short story evaluation tools using the Rebecca M. Valette taxonomy and CLIL approach based on teacher perceptions which include 1) dimensions/objectives of language assessment, 2) dimensions/objectives of cultural assessment, 3) dimensions/objectives of the literary evaluation, and 4) communication assessment dimensions. The four dimensions are developed into the principles of developing evaluation tools producing short story texts with the taxonomy of Rebecca M. Valette and the CLIL approach for eleventh graders of senior high school based on the results of analyzing the characteristics of teacher needs. Table 1 explains the analysis of the needs of evaluation tools producing short story texts with Rebecca M. Valette's Taxonomy and CLIL Approach in more detail below:

Table 1. Analysis of the Needs of Evaluation Tools Producing Short Story Texts with Rebecca M. Valette's Taxonomy and CLIL Approach

Aspects	Principles of Developing Evaluation Tool
Language Purpose	
1. Assessment of spoken language knowledge to foster a positive attitude towards a phenomenon	The form of assessment is in the form of multiple-choice; the number of paragraphs is 3; the limitation of the evaluation is the answer containing criteria, the duration of time is 40 minutes / 10 questions. The level of difficulty of the problem is 30% easy questions - 30 difficult questions.
2. Assessment of the ability to distinguish and classify language participants and motivate students in responding to a phenomenon	The evaluation form is performance appraisal, the number of paragraphs is 3, the number of aspects of the assessment indicators is three aspects or signs, the duration of time is 20x40 minutes/activity, and the scale of the scoring range is six scaled score.
3. Assessment of knowledge about the rules and patterns of written language to be able to appreciate and live the teachings of the religion they hold	The assessment form is multiple choice, the number of paragraphs is 3, the limitation of the assessment is the answers containing criteria, the duration of time is 40 minutes / 10 questions, and the difficulty level of the items is 30% easy questions, 40% medium questions, and 30% difficult questions
4. Assessment of the ability to distinguish and classify rules and patterns of written language so that students have a positive attitude towards a phenomenon	The form of evaluation is performed, there are three paragraphs, the number of aspects/indicators of assessment are three aspects or signs, the duration of time is 2x40 minutes, and the range/scale of scoring is four ranges.
5. Assessment of the ability to produce elements and patterns of spoken language so that students have a positive attitude towards a phenomenon	The form of evaluation is performed, the number of paragraphs is 4, the number of aspects/indicators of assessment is four aspects or signs, the duration of time is 2x40 minutes, and the range or scale of scoring is four ranges.
6. Assessment of the ability to produce speech or patterns of spoken language so that students are motivated to respond to a phenomenon	The form of assessment is in the form of essays/descriptions, the number of paragraphs is 3, the limitation of the evaluation is the answer based on the guidelines, the duration of time 40 minutes / 10 questions, and the difficulty level of the problem is 30% easy questions, 40% medium questions, and 30% questions difficult.

7.	Assessment of the ability to produce elements and written language patterns so that students can be motivated to respond to a phenomenon	The form of evaluation is performed, the number of paragraphs is 4, the number of aspects/indicators of assessment is four aspects or signs, duration of time is 1x40 minutes, scoring range or scale is three ranges.
8.	Assessment of the ability to produce utterances or patterns to convey the implicit intent of written language so that students have a continuous desire to improve their competence and understanding	The form of assessment is in the form of essays/descriptions. The number of paragraphs is 3, the limitation of the evaluation is the answers based on the guidelines, the duration of time is 40 minutes/10 questions, and the difficulty level of the problem is 40% easy questions, 30% medium questions, 30% difficult questions.

Cultural Purposes

1.	Assessment of the ability to understand the explicit meaning of socio-cultural patterns so that students can appreciate and live the teachings of their religion	The form of assessment in the form of essays or descriptions, the number of paragraphs is 3, the limitation of the evaluation is that there are answers with criteria, duration of time is 40 minutes/10 questions, the level of difficulty is 40% easy questions, 30% medium questions, 30% difficult questions
2.	Assessment of the ability to distinguish and classify several elements of artistry, so students are motivated to respond to a phenomenon	The form of evaluation is performed, the number of paragraphs is 3, the number of aspects/indicators is three aspects or signs, the duration of time is 2x40 minutes, and the range/scale is 4.
3.	Assessment of the ability to produce elements and socio-cultural patterns so that students have a continuing desire to improve competence and understanding	The form of evaluation is performed, the number of paragraphs is 3, the limitation of assessment is the answers based on the guidelines, the duration of time is 1x40 minutes, the range/scale of scoring is 6
4.	Assessment of the ability to evaluate a work or artistic phenomenon to appreciate and live the teachings of the religion they hold	The form of assessment is multiple choice, the number of paragraphs is 3, the assessment limit is the answers based on the guidelines, the duration of time is 40 minutes / 10 questions, the level of problem

		difficulty is 25% easy questions, 50% medium questions, 25% difficult questions

Literature Purpose

1.	Evaluation of knowledge about literature so that students are motivated to respond to a phenomenon	The form of assessment is in the form of multiple-choice, the number of paragraphs is 3, the limitation of the evaluation is the answers with criteria, the duration of time is 30 minutes / 10 questions, the level of difficulty of the items is 25% easy questions, 50% medium questions, 25% difficult questions.
2.	Assessment of the ability to evaluate a work or literary phenomenon so that students have a positive attitude towards a phenomenon	The form of assessment in the form of multiple-choice, the number of paragraphs is 3, the limitation of the assessment is the answers with criteria, the duration of time is 30 minutes / 10 questions, the level of difficulty of the items is 25% easy questions, 50% medium questions, 25% difficult questions.

Communication Purpose

1.	Assessment of the ability to analyze utterances or implicit patterns of communication so that students can appreciate and live the teachings of the religion they hold	The form of assessment in the form of essays/descriptions, the number of paragraphs is 3, the limitation of the assessment is the answers based on the guidelines, the duration of time is 20 minutes / 10 questions, the level of difficulty is 25% easy questions, 50% medium questions, 25% difficult questions
2.	Assessment of the ability to produce utterances or patterns to convey the explicit meaning of communication	The form of assessment is multiple choice, the number of paragraphs is 3, the assessment limit is the answers with guidelines, the duration of time is 40 minutes / 10 questions, the level of difficulty of the problem is 25% easy questions, 50% medium questions, 25% difficult questions

The evaluation tool developed in this study is a guidebook containing guidelines for evaluating the competency of producing short stories based on Basic Competency 3.8, 3.9, 4.8, and 4.9 in class XI of senior high school. These basic competencies combined with Rebecca M. Valette's taxonomy and CLIL's approach, which created a guidebook for the development of evaluation tools producing short stories. There are two stages in developing an evaluation tool handbook providing short story texts, namely the initial product and the final product. The initial product consists of two sets of questions, namely a package of items of knowledge and skills in which there is an integrated attitude assessment. The test packages

are tested for validity and reliability. The final product, the evaluation tool handbook, consists of five chapters compiled after the validity and reliability tests were held. The initial product consists of a package of knowledge questions in the form of multiple-choice, checklist questions, an essay, and a skill question in the way of an independent task (performance or project). Temporarily, an introduction of the final draft consists of the foundation of development, goals, objectives, and scope, concepts, and theories. It includes the understanding of learning evaluation, short story texts, Rebecca M. Valette's taxonomy, and CLIL approach, principles of developing evaluation tools. It consists of principles of developing an evaluation tool with the taxonomy of Rebecca M. Valette and CLIL approach, a knowledge evaluation tool that includes a learning syllabus, a lesson plan or RPP, a grid of knowledge and attitude competencies, and key answers to questions of knowledge and attitude competencies, and skill evaluation tools that include learning syllabus, lesson plans (RPP), skills and attitude competency grids, skills and attitude competency test questions, and key answers to skills and attitude competencies.:

The validity and reliability test of the prototype in the limited trial, among 32 multiple choice questions revealed the following:

- there are 25 valid items with the reliability of 0.702 categorized as good and sufficient, among eight checklist questions;
- five items were valid with the reliability of 0.393 categorized as good and very good, among five items of essay questions;
- five items were valid with the reliability of 0.431 in the good category, then 1 question in the skill test with eight aspects of the assessment, it was valid, and the reliability was 0.517 in the good and sufficient category. Table 2 describes the results of a limited trial of validity and reliability in more detail.

Table 2. Results of Analysis of Validity, Reliability, Discriminating Power, and Difficulty Level of Multiple Choice Questions

No.	validity		Reliability		Discriminating Power		Difficulty Level	
	r	Criteria	r 11	Criteria	DP	Criteria	P	Criteria
1	0,524	Valid	0,702	Reliable	0,438	Good	0,719	Easy
2	0,754	Valid			0,668	Good	0,656	Moderate
3	0,356	Valid			0,313	Moderate	0,656	Moderate
4	0,546	Valid			0,375	Moderate	0,813	Easy
5	0,443	Valid			0,313	Moderate	0,594	Moderate
6	0,443	Valid			0,313	Moderate	0,594	Moderate
7	0,393	Valid			0,375	Moderate	0,688	Moderate
8	0,398	Valid			0,313	Moderate	0,531	Moderate
9	-0,118	Invalid			-0,188	Poor	0,281	Difficult
10	0,539	Valid			0,438	Good	0,719	Easy
11	0,386	Valid			0,250	Moderate	0,750	Easy
12	0,538	Valid			0,500	Good	0,688	Moderate
13	0,558	Valid			0,500	Good	0,750	Easy
14	-0,388	Invalid			-0,313	Poor	0,281	Difficult
15	0,495	Valid			0,375	Moderate	0,750	Easy
16	0,440	Valid			0,438	Good	0,344	Moderate
17	-0,343	Invalid			-0,313	Poor	0,281	Difficult
18	0,433	Valid			0,438	Moderate	0,719	Easy
19	0,406	Valid			0,375	Moderate	0,625	Moderate
20	0,425	Valid			0,313	Moderate	0531	Moderate
21	0,466	Valid			0,438	Good	0,531	Moderate
22	0,511	Valid			0,250	Moderate	0,750	Easy
23	0,498	Valid			0,438	Good	0,594	Moderate
24	-0,343	Invalid			-0,313	Moderate	0,281	Difficult
25	-0,283	Invalid			-0,313	Moderate	0,281	Difficult
26	-0,373	Invalid			-0,313	Moderate	0,281	Difficult
27	-0,103	Invalid			0,063	Poor	0,281	Difficult
28	0,457	Valid			0,438	Good	0,459	Moderate
29	0,385	Valid			0,500	Good	0,500	Moderate
30	0,495	Valid			0,375	Moderate	0,750	Easy
31	0,466	Valid			0,438	Good	0,531	Moderate
32	0,511	Valid			0,250	Moderate	0,750	Easy

Based on the results of the first stage, there are still items that are not yet valid and reliable. It is necessary to improve the validity and reliability analysis in the second stage. The steps are to reduce the level of difficulty of the questions to be moderate or easy and change the

difference in power from bad to good so that the second stage can obtain validity and reliability on all items.

Based on these results, it turns out that the needs of evaluation tools produce short story texts with the taxonomy of Rebecca M. Valette and CLIL approach in class XI high school and include the following three areas:

- The principles of developing evaluation tools consist of language purpose, cultural purpose, literary purpose, and communication purpose,
- The development of the contents of an evaluation tool consists of four parts, namely, the conceptual development of basic competence material producing short stories, developing lesson plans competence in producing short story texts, developing guidelines, and developing instruments (questions).
- Validity and reliability of evaluation tool prototype producing short story text.

These areas are in line with previous research on evaluations and assessments that have been conducted (Boggino, 2009; Burrows, 2001; Khomsah & Subyantoro, 2019; McNamara & Lumley, 1997; Messick, 1996; Nurgiyantoro, 2013).

The principles in the development of this evaluation tool are based on four objectives proposed by Rebecca M. Valette's taxonomy, namely language purpose, cultural purpose, literary purpose, and communication purpose. These four purposes are in line with the learning objectives of the 2013 curriculum employing the CLIL approach. This can be seen in the syllabus of Bahasa Indonesia subjects (mandatory) SMA/MA/SMK/MAK in 2016 on page 7. Furthermore, in the syllabus of Bahasa Indonesia subjects (compulsory) SMA / MA / SMK / MAK / 2016, the four objectives in this evaluation tool play an important role in learning Bahasa Indonesia with curriculum 2013 which demands an integrated, interconnected, and mutually supportive material (language, literature, and literacy) scope.

A similar notion was expressed by (Valette, 1997), in her article entitled National Standards and the Role of the Imagination in Foreign Language Learning. This article discussed the purpose of the new standard in language learning, which is adept at communicating, understanding culture, the relationship of foreign languages with other languages, the comparison of foreign languages with other words, and the community or people who have different styles. The process of preparing this evaluation tool is different from the previous evaluation tool. The present evaluation tool used the Rebecca M. Valette taxonomy and CLIL approach, which is in line with the demands of the 2013 curriculum.

Rebecca M. Valette's Taxonomy is a learning taxonomy proposed by Rebecca M. Valette adapted from the revised Bloom Taxonomy. Rebecca M. Valette's taxonomy contains three

domains, namely the cognitive domain, the affective domain, and the psychomotor domain. If Rebecca M Valette's taxonomic features are compared to Bloom's taxonomy, it has presented four specific objectives, namely, language purpose, cultural purpose, literary purpose, and communication purpose. Additionally, Rebecca M. Valette's taxonomy has focused on four language skills (listening, reading, speaking, and writing). Therefore, this learning taxonomy is very suitable when applied to language learning (Valette, 1969, 1997).

Meanwhile, the CLIL approach is an approach that is integrated into learning the language content that proposes 4C as the application of CLIL, namely content, communication, cognition, culture (community or citizenship). The content is related to what topics (in this case, science topics such as ecosystems). Communication relates to what type of language is used (e.g., comparing, reporting). The genre concept is applied. It means that a text type is arranged and used. Cognition deals with what thinking skills are required regarding the topic (identifying, classifying). Culture is related to the local content of the surrounding environment which is associated with the problem, for example, the peculiarities of plants that exist in the area where students learn, including also the issue of character and language attitudes (Coyle, 2005; Khomsah & Subyantoro, 2019; Lesca, 2012; McNamara & Lumley, 1997; Mehisto, 2012).

Hence, the CLIL learning approach is suitable for developing students' competence in language and literature in the 2013 curriculum (DITJEN & Kependidikan, 2016). The purpose of this study is to obtain a description of general criteria that can be applied in the development of all types of learning materials, especially CLIL. One of the objectives of this tool is the purpose of communication. The use of communication is essential to the development of the skill of how to produce short story text because it can influence the learning outcomes of students to obtain maximum and complete results. This is reinforced by articles written by (Valette 1973) and Burrows, 2001).

Meanwhile, the basic competencies chosen to developed into evaluation tools are the basic competencies in producing short story texts. The short story was chosen in this study because this text is unique and interesting to improve students' language skills. This was confirmed by studies conducted by (Al-Dersi, 2013; Erkaya, 2005; Khatib & Seyyedrezaei, 2017; Pardede, 2011; Saleh Khalaf Ibnian, 2010).

The results of the validity and reliability test of the evaluation tool prototype with the taxonomy of Rebecca M. Valette and CLIL approach showed validity and reliability of 100% of all questions (multiple-choice, checklist questions, descriptions, and skills). Validity and reliability test were carried out twice, namely the first and second stages. The first stage was obtained from 32 multiple-choice items tested to students of class XI. Seven items were invalid, namely numbers 9, 14, 17, 24, 25, 26, 27. The invalidity was due to items number 9, 14, 17, 24, 25, 26, 27 are considered difficult and the value of r <value of r table (N = 32;

5%) = 0.349. Meanwhile, 25 items were valid. The reliability of 32 items tested to students was 0.702.

In the eight items of the checklist tested to the eleventh graders, there were three invalid items, namely numbers 33, 37, and 40. The invalidity was due to item number 33, 37, and 40 have the value of r <value of r table (N = 32 ; 5%) = 0.349. Meanwhile, for valid items were five items. The reliability of the eight items tested to students was 0.393. Furthermore, on the 5 item description questions tested to eleventh graders. From the five items obtained, two items were invalid. The invalidity is due to items number 2 and 4 which have value r <value of r table (N = 32; 5%) = 0.349. Then, the reliability of the five items tested to students is 0.431. Finally, in the matter of skills, there are eight aspects of assessment tested to eleventh graders. From these eight aspects, two aspects are invalid. The invalidity is due to aspects 2 and 4 which have value of r <value of r table (N = 32; 5%) = 0.349. The reliability of the eight aspects tested to students was 0.517. Based on the results of the first stage, there are still items that are not yet valid and reliable. It is necessary to improve the validity and reliability analysis in the second stage. The step taken by the researcher is to reduce the level of difficulty of the questions to be moderate or easy and change the difference of power from bad to good so that in the second stage, validity and reliability on all items can be obtained.

The evaluation tool is effective and feasible to assess aspects of knowledge, attitudes, and skills that exist in learning Bahasa Indonesia for eleventh graders of a senior high school implementing the 2013 curriculum based on the findings above. It especially the basic competence of producing short stories. The evaluative tool could support the assessment of learning Bahasa Indonesia by implementing the text-based 2013 curriculum. This is because this tool has been equipped with the aim of learning Bahasa Indonesia, which includes aspects of language, culture, literature, communication in line with the demands of the 2013 curriculum (Ditjen & Kependidikan, 2016). Additionally, this tool has been detailed with an assessment that includes aspects (language, culture, literature, communication) in the form of multiple-choice questions, checklist questions, descriptions, and skills so that it is believed to facilitate teachers in conducting assessments for their students.

Conclusion

Based on the results of the research and discussion presented in this paper, conclusions can be drawn as follows. First, the characteristics of the evaluation tool need to produce short story texts with the taxonomy of Rebecca M. Valette and the CLIL approach based on teacher perceptions are motivated by the principles of developing evaluation tools that can be grouped into four objectives including 1) language purpose, 2) cultural purpose, 3) literary purpose, and 4) communication purpose.
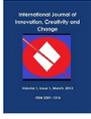
Researchers also propose the following suggestions. For teachers, Indonesian language teachers should need to apply the taxonomy of Rebecca M. Valette and CLIL's approach to improve the quality of learning outcomes in producing short stories in high school class XI. Furthermore, Rebecca M. Valette's taxonomic model and CLIL approach can be used as a choice in evaluating student learning outcomes in different classes. So that teachers will be accustomed to developing evaluation tools, increasing teacher creativity, and creating creative learning evaluation activities with students. For researchers, it should be able to conduct further research on the development of evaluation tools with the taxonomy of Rebecca M. Valette and the CLIL approach at all levels, whether elementary, junior high, or high school. For policymakers, the findings of this study should be used as a reference in the preparation of evaluation tools for Indonesian subjects at all levels of elementary, junior high, or high school. Additionally, it should be used as an innovation development in the development of evaluation tools for Bahasa Indonesia, especially the competence of producing short story texts.

REFERENCES

Al-Dersi, Z. E. M. (2013). The use of short-stories for developing vocabulary of EFL learners. *International Journal of English Language & Translation Studies*, *1*(1), 72–86.

Anum, A., & Apriyanto, S. (2019). DETECTING GENDER'S STRATEGIES IN LEARNING SPEAKING. *Premise: Journal of English Education*, *8*(1), 57. https://doi.org/10.24127/pj.v8i1.1932

Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, *20*, 24.

Boggino, N. (2009). Evaluation as a strategy for teaching. Evaluating processes and results. *Sísifo. Educational Sciences Journal*, *9*, 79–86.

Burrows, C. (2001). Searching for washback: The impact of assessment in the certificate in spoken and written English. *Studies in Immigrant English Language Assessment*, *2*, 95–187.

Cismas, S. C. (2009). Culture-fair assessment of foreign language skills in engineering education. *Proceedings of the 8th WSEAS International Conference on E-Activities and Information Security and Privacy*, 263–267.

Cizek, G. J. (2005). Pockets of Resistance in the Assessment Revolution. *Educational Measurement: Issues and Practice*, *19*(2), 16–23. https://doi.org/10.1111/j.1745-3992.2000.tb00026.x

Coyle, D. (2005). CLIL: Planning tools for teachers. The *University of Nottingham*.

De Smedt, F., & Van Keer, H. (2014). A Research Synthesis on Effective Writing Instruction in Primary Education. *Procedia - Social and Behavioral Sciences*, *112*, 693–701. https://doi.org/10.1016/j.sbspro.2014.01.1219

Ditjen, G., & Kependidikan, T. (2016). Panduan Operasional Penyelenggaraan Bimbingan dan Konseling Sekolah Dasar. *Jakarta: Kemendikbud*.

Erkaya, O. R. (2005). Benefits of Using Short Stories in the EFL Context. *Online Submission*, *8*.

Hidayat, O., & Apriyanto, S. (2019). Drama Excerpt: Tool in Enhancing Speaking Ability for Junior High School. *IJECA (International Journal of Education and Curriculum Application)*, *2*(3), 1. https://doi.org/10.31764/ijeca.v2i3.2029

Kasiyarno, D. S., & Apriyanto, S. (2020). Intercultural Communication and Business Chain Formation of America and Japan. *Journal of Talent Development and Excellence*, *12*(2s), 2042–2052.

Khatib, M., & Seyyedrezaei, S. H. (2017). A short story based language teaching (SSBLT): A literature-based language teaching method. *Journal of Humanities Insights*, *1*(04), 190–195.

Khomsah, A. N., & Subyantoro, S. (2019). Developing Evaluative Descriptive Text with Rebecca M. Valette's Taxonomy and CLIL Approach. *Seloka: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, *8*(2), 79–85.

Kusumaningsih, D. (2013). Indonesian text role as draft science in curriculum 2013: assessment introduction text structure strategies in an Indonesian book. *Asian Journal of Sciences and Humanities*, *2*(4), 1–4.

Lesca, U. (2012). An introduction to CLIL Notes based on a CLIL course at British Study Center–Oxford. *Istituto Di Istruzione Superiore-"Quintino Sella"–Biella*.

López-Pastor, V. M., Arribas, J. C. M., Aguado, R. M., & Fernandez, J. M. G. (2010). Formative assessment in project-oriented learning to improve academic performance. *Assessment, Teaching & Learning Journal*, *9*.

McNamara, T. F., & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational I settings. *Language Testing*, *14*(2), 140–156.

Mehisto, P. (2012). Criteria for Producing CLIL Learning Material. *Online Submission*.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241–256. https://doi.org/10.1177/026553229601300302

Nurgiyantoro, B. (2013). Model Penilaian OTentik Dalam Pembelajaran Bahasa. *Litera*, *10*(2). https://doi.org/10.21831/ltr.v10i2.1157

Pardede, P. (2011). Using Short Stories to Teach Language Skills. *JET (Journal of English Teaching)*, *1*(1), 14. https://doi.org/10.33541/jet.v1i1.49

Pocinho, M., Sousa, A., & Carvalho, R. G. (2014). Learning to write and get school success: A quasi-experimental study among 9th formers. *Journal of Linguistic and Intercultural Education-Jolie*, *7*, 119–142.
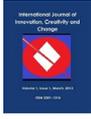
Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language Testing*, *23*(1), 1–34. https://doi.org/10.1191/0265532206lt322oa

Saleh Khalaf Ibnian, S. (2010). The Effect of Using the Story- Mapping Technique on Developing Tenth Grade Students' Short Story Writing Skills in EFL. *English Language Teaching*, *3*(4). https://doi.org/10.5539/elt.v3n4p181

Valette, R. M. (1969). *Directions in Foreign Language Testing.*

Valette, R. M. (1973). Developing and Evaluating Communication Skills in the Classroom. *TESOL Quarterly*, *7*(4), 407. https://doi.org/10.2307/3585871

Valette, R. M. (1997). National Standards and the role of imagination in foreign language learning. *Journal of the Imagination in Language Learning*, *4*, 18–25.

Zabihi, R., & Tabataba'ian, M. (2011). Teachers' evaluation and use of teacher's guides in foreign language classes. *Continental Journal of Arts and Humanities*, *3*(1), 1–10.