# Development of Evaluation Devices Producing Short Story Text with Rebecca M. Valette's Taxonomy and CLIL Approach to XIth Grade Senior High School Students

**M. Lukman Leksono[1]✉ & Subyantoro[2]**

[1] Public Senior High School 1 Dukuhwaru, Tegal, Jawa Tengah, Indonesia
[2] Universitas Negeri Semarang, Indonesia

## Article Info

## Abstract

The objectives of this research are (1) to describe the development characteristics of short story text evaluation tools of XIth grade Senior High School class participants by using Rebecca M. Valette taxonomy and CLIL approach based on teacher perception, (2) to describe the development of short story text evaluation tool for XIth grade Senior High School students using Rebecca M. Valette's taxonomy and CLIL approach. (3) to describe the validity and reliability  prototypes of short story text evaluation tools for XIth grade Senior High School class participants using Rebecca M. Valette's taxonomy and CLIL approach. The design of this research is R & D (Reseach and Development) approach with research step those are (1) phase I (preliminary study and early development, (2) stage II (development), (3) Stage III (Testing and Validation). This includes the data needs of evaluation tool development according to teacher's perception, data of expert evaluation result, and limited test data in the score form. The sample of this research is evaluation tool produce short story text on XIth grade Senior High School students. Research results are (1) development characteristic of short story text evaluation tool on XIth grade Senior High School students using Rebecca M. Valette's taxonomy and CLIL approach based on teacher's perceptions, (2) development of short story text evaluation tools for XIth grade Senior High School students using taxonomy Rebecca M. Valette's and CLIL approach, (3) validity and the prototype reliability of short story text evaluation tools for XIth grade Senior High School students using Rebecca M. Valette's taxonomy and CLIL approach.

✉ Address correspondence:
Pramuka No.48, Dukuhwaru, Tegal, Jawa Tengah, 52451
E-mail: leksonolukman@yahoo.com

**INTRODUCTION**

Basic competencies about the understanding, comparing, analyzing, and evaluating short story text are still poorly understood by teachers and learners. The skills of understanding structures and rules, comparing, analyzing, and evaluating short story text are receptive only. Assessment of these competencies and skills have not been developed in accordance with aspects of the 2013 curriculum assessment. Teachers still have difficulties in developing short story text material presented in teacher handbooks. In addition, the absence of a manual or teacher manual that is in line with the 2013 curriculum in evaluating short story text (Pujianto, Nuryatin, & Subyantoro, 2015). Therefore, it is necessary to develop evaluation tools in the form of handbooks or guidebooks to help teachers.

Teacher guides have not been adequately evaluated by professionals in ELT to date. The purpose of this study is to investigate the use of teacher guides that are widely used to evaluate their usefulness. The study was a qualitative study and data was collected through observations of 15 English classes and interviewed 17 English teachers at Firdausi University High School in Masyhad, a city in northeastern Iran. It is clear that the use of teacher guidance is different from that other one in general. Although teachers use different guides, all of them want to be better, and more logical (Zabihi & Tabataba'ian 2011).

Based on the evaluation problem, it is seen that the form of evaluation by using Bloom taxonomy which has been used less appropriate with the curriculum 2013 because of more emphasis on cognitive aspects that are no longer relevant. Bloom's taxonomy has not been able to identify aspects of skill and attitude aspects undertaken by learners in particular (Bloom, 1956). In accordance with the mandate of the curriculum 2013, in terms of assessment should be done by involving the three aspects of cognitive, psychomotor, and effective (religious, social, and cultural).

Not only that 3 aspects, the taxonomy of judgments also should have a balanced literary and linguistic content. Therefore, language learning does not go boring because it deals with language problems. Therefore, a taxonomy of language learning is required that can accommodate these needs. The taxonomy of the empire that can meet several needs in language learning as described, is the taxonomy put forward by Rebecca M. Valette.

Rebecca M. Valette Taxonomy is a taxonomy that has been specified for language learning purposes. This taxonomy adapts from Bloom's revised taxonomy. Rebecca M. Valette's taxonomy consists of three domains: cognitive, affective, and psychomotor that have been adapted to the needs of language learning in the 2013 curriculum. The three domains are then spelled out into fifteen indicators that are reassembled into four parts/domains. The four sections cover the realm of knowledge and understanding. The first domain is the realm of knowledge and understanding. This section can be done by teachers in the phases of building the context of learners through listening and interpreting activities. The second domain is the domain of knowledge and production. This domain is usually done with talking and writing activities. Among the first and second spheres, there is a third realm of manipulation. In this section the teacher can ask learners to memorize important material or provide evaluation. These three domains combine aspects of cognitive and psychomotoric. These three domains of the cognitive and psychomotor aspects end with more evaluation activities emphasized on the literary discourse. In addition to the three domains that have been explored and adapted to the needs of language learning, Rebecca M. Valette taxonomy also has a content or material content that is quite complete. This is indicated by the description of eight areas of material grouped into four main objectives. The first goal is the purpose of language, in the purpose of this language there are three materials namely spoken language, written language, and body language (kinesik). The second goal is a cultural goal. In this cultural purpose includes a way to live in a

cultured society, knowledge of the love of the homeland, and art. The third goal is the literary goal. Then the fourth goal is the purpose of communication. So, in language learning, we should be adept at communicating, knowing the culture, knowing one language's relationship with other languages, knowing language comparisons with other languages, and knowing communities or communities that have language. (Valette, 1969; Valette ,1997).

Communication skills are an important part of the learning process. Includes learning communication in the language. The communication or speaking test given at the classroom must be accordance with the learning objectives. The phenomenon that occurs in the field, i.e. some teachers will tend to ignore the actual communicative competence if the assessment is not done based on communicative approach, they will also ignore the class assessment that impacts on the contribution and significant on student achievement. Similarly, if the measured aspect is not in accordance with what should be measured it can lead to meaningless learning (Valette, 1973; Burrows, 2004).

In addition to appropriate and appropriate taxonomies used in Indonesian language and literature learning, approaches in learning should also be determined appropriately and in accordance with the 2013 curriculum. In addition, rethinking in language assessment should be tailored to the progress of language learning theories. Therefore, Indonesian language and literature learning use CLIL approach as the foundation of learning. This is in line with text-based curriculum learning 2013. Content and Language Integrated Learning (CLIL) is an approach that integrates content teaching/content of a subject curriculum with non-indigenous language learning (Dewi & Zulaeha, 2017). In CLIL, learners are given the opportunity to learn to think how to use the language, not just thinking about the language as the main focus of learning. CLIL has four basic components of content, communication, cognition, and culture. In addition, the choice of this approach also fits well with Rebecca M.

Valette's taxonomy because the four basic components of the CLIL approach coexist and complement each other with Rebecca M. Valette's economic taxonomy, which is social, communication, and culture. (McNamara, 2001; Coyle, 2005; Lesca, 2012; Mehisto. 2012; Subyantoro, 2014).

Meanwhile, an evaluation tool also should provide an assessment that has a positive impact and can overcome cultural background differences as well as provide a truly valid assessment system for its users. This positive impact can be seen by testing the validity and reliability test of the tool. This will certainly be useful for improving learners' learning outcomes. In addition, evaluation tools must be effective in order to maximize language learning outcomes (Messick, 1996; Saif 2006; Cismas, 2009).

The Indonesian evaluation tool developed in this research is a tool that facilitates teachers in collecting data or value of learners' ability (cognitive, attitude, and psychomotor) directly or indirectly. The evaluation concept is a process of determining the value by considering the results of observation or collection of data obtained. In addition, in evaluating the ability of learners there are two things that must be considered by teachers, learning process and learning outcomes. Both of these must be balanced so that the goal of learning language skills can be achieved with the maximum. In evaluating students, teachers not only look at the students' learning outcomes but teachers also see the learning process of students. In addition, this evaluation tool should be strengthened by developing a learning strategy that combines POL (Project-Oriented Learning) and formative assessment to develop professional competence among learners. (Cizek, 2000; Norberto Boggino, 2009, Pastor et al., 2010).

Then the text is taken to complete this evaluation tool, is a short story text. Short story text selected because this text is one of the literary texts that help and improve the ability of learners to learn four skills of listening, speaking, reading and writing. Short story text is also more effective and fun because the motivation is embedded in the story and useful for pedagogical learners. Short story text is a valuable and interesting

material to improve the language skills of learners. In addition, with short stories, teachers can teach literature, culture. The Competence is to produce short story texts that will be developed on learners. Producing text of this short story is also expected to give the positive impact on the students (Erkaya, 2005; Ibnian, 2010; Pardede, 2011; Al Dersi, 2013; Khatib & Sayyedrezaei, 2013).

The term competence of producing short story text is a term in the 2013 curriculum which basically comes from learning the skills of writing short story text. The learning of short story writing skills based on text genres is an effective learning technique as a strategy to increase the knowledge and ability of learners in writing. In addition, effective writing skills are indispensable for participation in the community and bring benefits to enhance more specialized writing skills, such as mother tongue (Kusumaningsih, 2013; Smedt, 2013; Pocinho et.al., 2014).

Based on these statements it can be argued that the appropriate use of taxonomy with short story text learning and CLIL approach that integrates content teaching/content from the curriculum of subjects with non-indigenous language teaching is very suitable to the implementation of the 2013 curriculum. The four basic components in CLIL approach is suitable if applied side by side with the learning taxonomy put forward by Rebecca M. Valette. This is because some elements have the same purpose and complement each other. Therefore, it is necessary to develop evaluation tools in the Indonesian language especially short story text with the taxonomy of Rebecca M. Valette and CLIL approach for Senior High School curriculum 2013.

**METHODS**

This research used R & D (Reseach and Development) approach with research step that is (1) phase I (preliminary study and early development, (2) stage II (development), (3) Stage III (Testing and Validation). including the data of evaluation tool development needs according to teacher's perception, data of expert evaluation result, and limited test data in the form of the score. The sample of this research is evaluation tool producing short story text on XIth grade Senior High School students.

Data retrieval is done by test and non-test technique. The test technique is done to know the ability of the students to produce short story text, and the non-test technique using a questionnaire to develop the teacher device based on the expert's knowledge of the device in XIth grade Senior High School. Prior to the research, conducted questionnaire collection needs of device development based on teacher perceptions and questionnaires by experts. Data analysis techniques are currently grouped into three data analyzes. Three groups of data analyzed are (1) data of requirement analysis result obtained from teacher questionnaire, (2) data of expert validation result, (3) data Product effectiveness test result (validity and reliability).

**RESULTS AND DISCUSSION**

The results of this study are the characteristics of the tools of text evaluation of the short stories using the taxonomy of Rebecca M. Valette and CLIL approach based on teachers' perceptions that include (1) the dimension/purpose of language assessment, (2) the dimensions/objectives of cultural assessment, (3) the dimensions/purpose of the literary assessment, and (4) dimension/ communications assessment.

These four dimensions are developed into principles of evaluation tool development producing short story text with Rebecca M. Valette taxonomy and CLIL approach for high school class XI students prepared based on the analysis of teacher requirement characteristics. This can be seen in the table below.

**Table 1**. Aspect and Principles of Development Evaluation Tool

| Aspect | Principles of development evaluation tool |
|---|---|
| **Language Objectives** | |
| 1  Assessment of oral language knowledge to cultivate a positive attitude towards a phenomenon | The form of the assessment is a multiple choice, the number of paragraphs is 3, the definition of the criteria is the answer, the duration is 40 minutes / 10 questions, and the difficulty level is 30% |
| 2  Assessment of the ability to distinguish and group elements of spoken language to motivate learners in response to a phenomenon | Form of assessment in the form of performance, the number of paragraphs there are 3, the number of aspects/indicators of assessment are 3 aspects/indicators, the duration of time is 2x40 minutes/activities, and scaling range/scale is 6 scale scoring |
| 3  Assessment of knowledge of rules and patterns of written language to be able to appreciate and appreciate the religious teachings it embraces | The form of assessment is a multiple choice, the number of paragraphs is 3, the definition of the criteria is the answer, the duration of time 40 minutes/10 questions, and the difficulty level of the specified problem is 30% easy problem -40% about moderate-30% problem difficult |
| 4  Assessment ability to differentiate and group rules and writing patterns so that learners have a positive attitude towards a phenomenon | Form of assessment in the form of performance, the number of paragraphs there are 3, the number of aspects/indicators of assessment are 3 aspects/indicators, duration of 2x40 minutes, and range/scale scoring ie four ranges |
| 5  Assessment of the ability to produce elements and patterns of spoken language so that learners have a positive attitude toward a phenomenon | Form of assessment in the form of performance, the number of paragraphs there are 4, the number of aspects/indicators of assessment are 4 aspects/indicators, duration of 2x40 minutes, and range/scale scans four ranges |
| 6  Assessing the ability to produce speech or patterns to convey the implicit intentions of spoken language so that learners are able to be motivated to respond to a phenomenon | The form of the assessment is essay/description, the number of paragraphs is 3, the assessment limit is the answer based on the lattice, the duration of time 40 minutes / 10 questions, and the difficulty level of the specified problem is 30% easy problem -40% about moderate-30% difficult problem |
| 7  Assessment of the ability to produce elements and patterns of written language so that learners can be motivated to respond to a phenomenon | Form of assessment in the form of performance, the number of paragraphs there are 4, the number of aspects/indicators of assessment that is 4 aspects/indicators, duration 1x40 minutes, range/scale scoring ie three ranges |
| 8  Assessment of the ability to produce speech or pattern to convey the implicit intention of written language so that learners have an ongoing desire to improve competence and understanding | The form of the assessment is essay/description, the number of paragraphs is 3, the assessment limit is the answer based on the grid, the duration is 40 minutes / 10 questions, and the difficulty level of the specified problem is 40% easy problem-30% about moderate-30% difficult problem |
| **Cultural Objectives** | |
| 1  Assessment of the ability to understand the explicit meaning of socio-cultural patterns so that learners can appreciate and appreciate the religious teachings it embraces | Form of assessment in the form of essays/descriptions, the number of paragraphs there are 3, the limits of the assessment that is the answer there are criteria, the duration of time 40 minutes / 10 questions, the difficulty level of the specified problem is 40% easy problem-30% about medium-30% problem difficult |
| 2  Assessment ability to differentiate and classify some elements of art so that learners are motivated to respond to a phenomenon | Form of assessment in the form of performance, the number of paragraphs there are 3, the number of aspects / indicators are 3 aspects/indicators, duration of 2x40 minutes, and range / scale scoring ie four |
| 3  Assessment of the ability to produce elements and socio-cultural patterns so that learners have an ongoing desire to improve competence and understanding | The form of assessment in the form of performance, the number of paragraphs there are 3, the assessment limit is the answer based on the grille, the duration of 1x40 minutes, the scaling range of six |
| 4  Assessment of the ability to evaluate a job or artistic phenomenon so as to appreciate and appreciate the religious teachings it embraces | The form of assessment is a multiple choice, the number of paragraphs is 3, the assessment limits are the answers based on the grid, the duration is 40 minutes / 10 questions, the difficulty level is 25% |
| **The Purpose of Literature** | |
| 1  Assess the knowledge of literature so that learners are motivated to respond to a phenomenon | Form of assessment in the form of multiple choice, the number of paragraphs there are 3, the limits of the assessment that is the answer there are criteria, duration of time 30 minutes / 10 questions, the difficulty level of the specified problem is 25% easy problem-50% about moderate-25% problem difficult |
| 2  Assessment of the ability to evaluate a job or literary phenomenon so that learners have a positive attitude toward a phenomenon | The form of assessment is a multiple choice, the number of paragraphs is 3, the assessment limit is the answer based on the grid, the duration is 40 minutes / 10 questions, the difficulty level is 25% |
| **Communication Objectives** | |
| 1  Assessment of the ability to analyze speech or implicit patterns of communication so that learners are able to appreciate and appreciate the religious teachings it embraces | Form of assessment in the form of essays/descriptions, the number of paragraphs is 3, the assessment limit is the answer based on the grid, the duration of 20 minutes / 10 questions, the difficulty level is 25% |
| 2  Assessment of the ability to produce speech or pattern to convey the explicit meaning of communication | The form of assessment is a multiple choice, the number of paragraphs is 3, the assessment limits are the answers based on the grid, the duration is 40 minutes / 10 questions, the difficulty level is 25% |

The evaluation tool developed in this study is a guidebook which contains competency evaluation guidelines for producing short story text based on KD 3.8, KD 3.9, KD 4.8, and KD 4.9 in XIth grade Senior High School. The basic competencies-these basic competencies are then combined with the taxonomy of Rebecca M. Valette and the CLIL approach that creates an evaluation tool development guidebook producing short story text. There are two stages in the development of an evaluation tool guidebook producing short story text, the initial product, and the final product. The initial product consists of two packages of problems, namely the package of knowledge and skills in which there is an attitude of integrity in it. The packages of the questions tested for their validity and reliability. Then the final product is an evaluation tool guidebook produces short story text consisting of five chapters prepared after validity and reliability test. The following will explain both products. The initial product consists of a package of multiple choice questions, a checklist, a description and a skill set of independent tasks (performance/project). Meanwhile, the final product consists of a preliminary covering the foundations of development, goals, objectives, and scope, concepts and theories that include the notion of instructional evaluation, short story text, Rebecca M. Valette's taxonomy and CLIL approach, development tool evaluation principles include the principles of developing evaluation tools with Rebecca M. Valette's taxonomy and CLIL approaches, knowledge evaluation tools that include learning syllabus, lesson plan (RPP), lattice on knowledge and attitude competence, and key answers about knowledge and attitude competence, and skill evaluation tools that include learning syllabus, lesson planning plan (RPP), lattice on skills and attitude competence, skill competence and attitude test, and key answer about skill and attitude competence.

Test validity and reliability of prototype in limited test obtained by data 32 questions of multiple choice, there are 25 problems item stated valid with reliability 0,702 good and enough category, data 8 item about checklist obtained 5 item is valid with reliability 0,393

good and good category once, the data 5 points of the description obtained 5 items of questions declared valid with reliability of 0.431 good category, then data 1 item of skill with 8 aspects of assessment is valid and reliability that is 0,517 good and enough category. The following tables describe the results of a limited trial of validity and reliability.

Based on the results in the first stage there are still items that are not valid and reliable, it is necessary to improve the analysis of validity and reliability in the second stage. Steps that are done is to lower the level of problem difficulty to be moderate or easy and change the differentiation from bad to good so that in the second stage can be obtained the validity and reliability on all items.

Based on the results of the research, it turns out that the evaluation tool needs to produce short story text with taxonomy of Rebecca M. Valette and CLIL approach at XIth grade Senior High School include (1) evaluation tool development principles grouped into four objectives, i.e. language objectives, cultural objectives, literary purposes, (2) developing the content of evaluation tools consisting of parts, that is, the conceptual development of KD materials producing short story text, the development of competency RPP producing short story text, grating development, and instrument development, (3) validity and reliability prototype evaluation tools produce short story text. These three things are in line with previous research on evaluations and assessments made by (Messick, 1996; McNamara, 2001; Burrows, 2004; Norberto Boggino, 2009; Nurgiyantoro, 2011; Eko, 2012; Subyantoro, 2014).

The principles in the development of this evaluation tool are based on four goals taken from Rebecca M. Valette's taxonomy, i.e. language objectives, cultural goals, literary goals, communication goals. These four objectives are in line with the objectives of the 2013 curriculum lesson using the CLIL approach. This can be seen in the syllabus of the Indonesian language subjects (mandatory) of SMA / MA / SMK / MAK 2016. In addition, in the syllabus of Bahasa

Indonesian subjects (mandatory) SMA / MA / SMK / MAK 2016 fourth goal on this evaluation tool plays an important role in the learning of the Indonesian curriculum in 2013 which demands an integrated, interrelated, and mutually supportive scope of matter (language, literature, and literacy) (Zulaeha, 2016).

The same thing is also expressed by Valette (1997) in his article entitled National Standards and the Role of the Imagination in Foreign Language Learning. This article discusses the goals of new standards in language learning, which are proficient in communicating, knowing the culture, knowing the relations of foreign languages with other languages, knowing the comparisons of foreign languages with other languages, and knowing communities that have language. In the process of evaluating this evaluation tool differ from the previous evaluation tools, this evaluation tool uses Rebecca M. Valette's taxonomy and CLIL approach that conforms to the demands of the 2013 curriculum.

Rebecca M. Valette's Taxonomy is the taxonomic of learning advocated by Rebecca M. Valette is a learning taxonomy adapted from Bloom's Bloom's Taxonomy. Taxonomy of Rebecca M. Valette, contains three domains of cognitive domains, affective domains, and psychomotor domains. The specialty of Taxonomy Rebecca M Valette is compared with the bloom taxonomy, which has presented four specific objectives namely, language objectives, cultural objectives, literary goals, and communication objectives. In addition, Rebecca M. Valette's taxonomy has focused on four language skills (listening, reading, speaking, and writing). Therefore, the taxonomy of learning is particularly suitable when applied to language learning (Valette, 1969; Valette, 1997).

Meanwhile, CLIL approach is an integrated approach to its learning content of 4C filed language as the implementation of CLIL, i.e. content, communication, cognition, culture (community/citizenship). Content is related to what topic (in this case is the topic of natural sciences like an ecosystem). Communication deals with what kind of language is used (e.g.

comparing, reporting). In this section the concept of applied genre, how a particular type of text is composed (text structure) and what form of language is often used in that type of text. Cognition deals with what thinking skills are required regarding the topic (identifying, classifying). Culture deals with the local content of surrounding environments related to the topic, for example, the distinctiveness of plants in the area where students learn, as well as issues of character and language attitudes (McNamara, 2001; Coyle, 2005; Lesca, 2012; Mehisto, 2012; Subyantoro, 2014).

Thus, CLIL learning approaches are suitable for developing the competence of learners in language and literature in the 2013 curriculum (Kemendikbud, 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 goals of this device is the purpose of communication. The purpose of communication is very important to be developed in learning skill producing short story text because it can influence learners to learn result to get maximum result and complete. This is reinforced by articles written by (Valette, 1973; Burrows, 2004).

Meanwhile, the basic competence chosen to be developed into an evaluation tool is the basic competence of producing short story text. Short story text is selected in this research because the text is unique and interesting to improve the language skills of learners. This is reinforced by studies conducted by (Erkaya, 2005; Ibnian, 2010; Pardede, 2011; Al Dersi, 2013; Khatib & Sayyedrezaei, 2013).

Based on the prototype test of validity and reliability of evaluation tools with Rebecca M. Valette taxonomy and CLIL approach, 100% of all questions (multiple choice, checklist, description, and skill) were evaluated and CLIL approaches. Validity and reliability are done twice, that is the first and second stage. The first stage was obtained from 32 items of multiple choice tested to the students of class XI, 7 items that are not valid number 9, 14, 17, 24, 25, 26, 27. The invalidity is due to item 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, for a valid item of 25 items. Then for the reliability of 32 items tested to the learner that is 0.702.

At 8 points about the checklist tested to the students of class XI obtained 3 items that are not valid numbers 33, 37 and 40. The invalidity is due to item 33, 37, and 40 the value of r < value of r table (N = 32 ; 5%) = 0.349. In the meantime, for items that have valid amounted to 5 items. Then for the reliability of the 8 items tested to the students that are 0.393. Furthermore, on the 5 items of the tested description to students of class XI. From 5 items, 2 items were obtained. The invalidity is due to item 2 and 4 the value of r < value of r table (N = 32; 5%) = 0.349. Then for the reliability of the 5 items tested to the learner is 0.431. Finally, in the matter of skills, there are 8 aspects of assessment that are tested to students of XIth grade. Of the 8 aspects are obtained two invalid aspects.

The invalidity is due to aspects 2 and 4 the value of r < value of r table (N = 32; 5%) = 0.349. Then for the reliability of the 8 aspects tested to the learner is 0,517. Based on the results in the first stage there are still items that are not valid and reliable, it is necessary to improve the analysis of validity and reliability in the second stage. The steps taken by the researcher is by lowering the level of difficulty about being moderate or easy and changing the differentiation from bad to good so that in the second stage can be obtained the validity and reliability on all items.

Seeing these results, it can be said that the evaluation tool is effective and feasible to assess the aspects of knowledge, attitudes, and skills that exist in learning Indonesian XIth grade Senior High School 2013 curriculum, especially basic competence producing short story text. This evaluation tool i believed to support the learning of text-based Bahasa Indonesia curriculum 2013 because in this device has been equipped with learning objectives of Indonesian language covering aspects of language, culture, literature, communication in accordance with the demands of the 2013 curriculum (Hidayati & Zulaeha,

2018; Kemendikbud, 2016). In addition, this tool has been detailed with assessments that include aspects (language, culture, literature, communication) in the form of multiple-choice questions, checklist, description, and skills so that it is believed will facilitate the teachers in conducting the assessment of the participant's students.

## CONCLUSION

Based on the results of research and discussion that have been described in chapter IV can be raised the following conclusions. First, the characteristics of the evaluation tool needs to produce short story text with Rebecca M. Valette's taxonomy and CLIL approach based on teachers' perception is basically based on evaluation tool development principles that can be grouped into four purposes, among others: (1) language objectives, (2) cultural objectives, (3) the purpose of literature, and (4) the purpose of communication.

Second, the development of evaluation tools produces short story text with Rebecca M. Valette's taxonomy and CLIL approach to XIth grade Senior High School students consisting of (1) conceptual development of KD materials producing short story text, (2) development of lesson's plan (RPP) competency producing short story text, (3) development of lattice questions, and (4) the development of assessment instruments. In addition, this form of evaluation tool development, an evaluation tool development guide to evaluating the competence of producing short story text in XIth grade Senior High School students based on the taxonomy of Rebecca M. Valette and CLIL approach.

Third, validity and reliability of evaluation tool development produce short story text with Rebecca M. Valette taxonomy and CLIL approach consists of (1) validity and reliability of multiple choice questions, (2) validity and reliability of checklists, (3) validity and reliability of the description question, and (4) validity and reliability of skills.

The researcher also put forward the following suggestions. For teachers, Indonesian

teachers should apply the taxonomy of Rebecca M. Valette and CLIL approach to improve the quality of learning outcomes producing short story text in XIth grade Senior High School. In addition, Rebecca M. Valette's taxonomic model and CLIL approach can be used as an alternative choice in evaluating students' learning outcomes in different classes so that teachers will be familiar with developing evaluation tools, enhancing teacher creativity, and creating creative learning evaluation activities with learners.

For the researchers, it should be possible to conduct further research on the development of evaluation tools with the taxonomy of Rebecca M. Valette and CLIL approach at all levels in elementary, junior or senior high schools. For policymakers, the findings of this research should be used as a reference in the preparation of evaluation tools of Indonesian subjects in all levels of elementary, junior high, or high school. In addition, it should be used as a development of innovative development of evaluation tools of Indonesian subjects, especially the competence of producing short story text.

## REFERENCES

Al Dersi, Z. E. M. (2013). The Use of Short-Stories for Developing Vocabulary of EFL Learners. University of Sebha, Libya: *Journal of English Language & Translation Studies,* 1(1), 72-86. http://www.eltsjournal.org/archive/value1%20issue1/7-1-1-13.pdf

Bloom, B. S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. 1956. *Taxonomy of Educational Objectives: Handbook I:* Cognitive *Domain*. New York: David Mckay.

Boggino, N. (2009). Evaluation as a Strategy for Teaching. Evaluating Processes and Results. Sísifo. *Educational Sciences Journal* 9, 79-86.

Burrows, C. (2004). Searching for Washback: The Impact of Assessment in The Certificate In Spoken and Written English. In G. Brindley & C. Burrows (Eds.), *Studies in Immigrant English Language Assessment* 2. Sydney, Australia: National Center for English Language Teaching and Research.

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*. Romania: The Polytechnic University of Bucharest, World Scientific and Engineering Academy and Society (WSEAS).

Cizek, G. J. (2000). Pockets of Resistance in the Assessment Revolution, Educational Measurement Issues and Practice Journal. *Summer 2000,* 19(2). https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2000.tb00026.x

Coyle, Do. (2005). *CLIL, Planning Tools for Teachers.* Nottingham: University of Nottingham.

Eko, F. (2012). Pengembangan Perangkat Evaluasi dengan Taksonomi *Structure of The Observed Learning Outcome* (*SOLO*) pada Kompetensi Menulis Peserta Didik Kelas VII SMP di Kabupaten Pekalongan. *Thesis.* Pascasarjana UNNES.

Erkaya, O. R. (2005). Benefits of Using Short Stories in the EFL Context. International IATEFL Annual Conference in Brighton: *Journal Article in Teaching Articles for the Profession*, Asian EFL Journal V.

Ibnian, S. S. K. (2010). The Effect of Using the Story-Mapping Technique on Developing Tenth Grade Students' Short Story Writing Skills in EFL. Middle East University for Graduate Studies. *Journal of English Language Teaching,* 3(4). https://files.eric.ed.gov/fulltext/EJ1081969.pdf

Kementrian Pendidikan dan Kebudayaan. (2016). *Silabus Mata Pelajaran SMA/MA/SMK/MAK.* Jakarta: Kementrian Pendidikan dan Kebudayaan.

Khatib, M. & Seyyed, H. S. (2013). Short Story Based Language Teaching (SSBLT): A Literature-Based Language Teaching Method. Tabatabai University, Islamic Azad University, Iran: *International Journal of Basic Sciences & Applied Research*, 2(2), 155-159. https://docplayer.net/22703099-Short-story-based-language-teaching-ssblt-a-literature-based-language-teaching-method.html

Kusumaningsih, D. (2013). Indonesian Text Role as Draft Science in Curriculum 2013: Assessment in introduction Text Structure Strategies in an Indonesia Book. *Asian Journal Sciences and Humanities,* 2(4). http://www.ajssh.leena-luna.co.jp/AJSSHPDFs/Vol.2(4)/AJSSH2013(2.4-28).pdf

Lesca, U. (2012). *An Introduction to CLIL.* Oxford: Quintino Sella.

McNamara, T. F. (2001). The Effect of Interlocutor and Assessment Mode Variables in Overseas Assessments of Speaking Skills in Occupational I Settings. *Language Testing,* 14(2), 140-156. http://journals.sagepub.com/doi/10.1177/026553229701400202

Mehisto, P. (2012). Criteria for Producing CLIL Learning Material. *Encuentro Journal,* 21(5), 15-33. https://eric.ed.gov/?id=ED539729

Messick, S. (1996). Validity and Washback in Language Testing. *Language Testing Journal,* 13, 241-256. http://journals.sagepub.com/doi/abs/10.1177/026553229601300302

Nurgiyantoro, B. & Pujiyati, S. (2011). Model Penilaian Otentik dalam Pembelajaran Bahasa. *Litera Jurnal Penelitian Bahasa, Sastra dan Pengajarannya,* 10(2). https://journal.uny.ac.id/index.php/litera/article/view/1157

Pardede, P. (2011). Using Short Stories to Teach Language Skills. *Journal of English Teaching.* 1(1). https://www.researchgate.net/publication/283724743_Using_Short_Stories_to_Teach_Language_Skills

Pastor, Victor M. Lopez, Juan Carlos Manrique Arribas, Roberto Monjas Aguado, & Juan Manuel Gea Fernandez. (2010). Formative Assessment in Project-Oriented Learning to Improve Academic Performance. *The Assessment, Learning and Teaching Journal* 9. Leeds Metropolitan University.

Pujianto, A., Nuryatin, A. & Subyantoro. (2015). Keefektifan Pembelajaran Keterampilan Menulis Cerpen dengan Model Investigasi Kelompok dan Model Stad Berdasarkan Tipe Kepribadian Peserta Didik Kelas VII. *Seloka, Jurnal Pendidikan Bahasa dan Sastra Indonesia,* 4(2). https://journal.unnes.ac.id/sju/index.php/seloka/article/view/9871

Dewi, K. C., & Zulaeha, I. (2017). Pengembangan Perangkat Pembelajaran Teks Anekdot Tema Konflik Sosial di Kalangan Remaja dengan Pendekatan CLIL dan Model Berbasis Projek". *Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia,* 5(2), 172-176. https://journal.unnes.ac.id/sju/index.php/seloka/article/view/13079

Hidayati, N. & Zulaeha, I. (2018). The Effectiveness of Poetry Reading Learning using Draladater Model on Extrovert and Introvert Senior High School Student. *Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia,* 7(1), 1-8. https://journal.unnes.ac.id/sju/index.php/seloka/article/view/22057

Pocinho, A. S. & Renato, G. C. (2014). *Learning to Write and Get School Success: a Quasi-Experimental Study among 9th Formers.* Portugal: University of Madeira.

Saif. (2006). Aiming for Positive Washback: a Case Study of International Teaching Assistants. *Language Testing Journal*, 23(1). https://eric.ed.gov/?id=EJ752013

Smedt, F. D. & Hilde, V. K. (2013). A Research Synthesis on Effective Writing Instruction in Primary Education. *International Conference on Education & Educational Psychology*. Belgium: Ghent University, Dunantlaan.

Subyantoro. (2014). Pengembangan Perangkat Evaluasi Berdasarkan Taksonomi *The Structure of Observed Learning Outcome.* Jurnal LITERA 13(1). http://download.portalgaruda.org/article.php?article=163043&val=486&title=pengembangan%20evaluasi%20pembelajaran%20bahasa%20indonesia%20berbasis%20taksonomi%20structure%20of%20observed%20learning%20outcome

Subyantoro. (2014). *Teori Pembelajaran Bahasa: Implementasi Psikolinguistik Pendidikan.* Semarang: UNNES Press.

Valette, R. M. (1969). *Modern Language Testing: Second Edition*. New York: Harcourt Brace Jovanovich, INC.

Valette, R. M. (1973). *Developing and Evaluating Communication Skills in the Classroom*. *TESOL Quarterly,* 7(4), 407-424. https://www.jstor.org/stable/3585871

Valette, R. M. (1977). *Modern Language Testing*. USA: Harcourt Brace Jovanovic Inc.

Valette, R. M. (1997). National Standards and the Role of the Imagination in Foreign Language Learning. *Journal of the Imagination in Language Learning*. Jersey City: Hepburn Hall, NJ 07305-1597.

Zabihi, R. & Maryam, T. (2011). Teachers' Evaluation and Use of Teacher's Guides in Foreign Language Classes. *Willed Journals Continental J. Arts and Humanities.*

Zulaeha, I. (2016). *Teori, Model, dan Implementasi Pembelajaran Menulis Kreatif.* Semarang: UNNES Press.