

DEVELOPING WRITING TEST FOR EFL LEARNERS

Ary Setya Budhi Ningrum

STAIN Kediri, East Java

Abstract: This paper tries to present a general review of testing writing in English as a foreign language. This paper covers some of the important and relevant aspects of developing writing test such as definition of writing, writing in ESL/EFL classroom, writing assessment, and validating writing test. Special review is addressed to give a brief note of a comprehensive and step-by-step procedure of developing test instrument proposed by Prof. Soenardi Djiwandono.

Keywords: writing test, English as a foreign language

Testing is one of the most influential areas related to teaching as it can be used to measure the effectiveness of teaching learning progress. In fact, if the test used is good it can be used to check the effectiveness of the whole process. Heaton (1988:5) insists on these aspects mentioning that both teaching and testing are so closely interrelated that is virtually impossible to work in either field without being concerned with the other. Hence, learning to teach also requires learning to develop a good test.

There are two major parts of object of the test if we speak about language testing. First part is in the sense of language skills namely speaking, listening, writing, and reading. Secondly are linguistic components such as grammar, vocabulary, and phonetics. In this paper, I will focus on testing language skill namely writing test.

DEFINITION OF WRITING

Writing is a communicative act and a way of sharing observation, information, thoughts, and ideas with others through written language (Cohen et. al., 1989; Troyka 1987). As a communicative act, writing involves both physical and mental process. Cohen et al. (1989) state further that as a physical process, writing means producing graphemes and orthographic symbols in the form of letters or combination of letters that relates to the sounds in spoken language using hands while the eye movements follow over the words or sentences. While physical activity, according to Richards (1990:101), can be captured visually, mental activity, on the other hand, cannot be seen or observed directly. It is the process in the writer's mind that includes making connection between ideas and processing thoughts to be expressed in a meaningful written text by employing linguistics organization.

More specifically to the point, Troyka (1987: 3-4) characterizes writing from its elements. Writing is seen as “a way of communicating a message to a reader for a purpose.” In that case, writing entails four key elements namely, communicating, message, reader, and purpose. Communicating means the act of sending a message from the writer to readers. Message refers to the content, ideas or information carried in the piece of writing. Reader is the receiver of the message. Purpose is the intention of writing. All of elements attain a way of transferring information from the writer, as the encoder of the information, to the reader, as the decoder.

To make the message understandable the text produced in writing process has to be communicatively effective. It means that a writer must have a communicative competence, which includes grammatical, discourse, sociolinguistics, and strategic competences (Savignon, 1983; Eanes, 1997). Savignon (1983); Eanes (1997) further explain that grammatical competence focuses on sentence-level grammar and requires knowledge of lexical items and rules of morphology, syntax, sentence-grammar semantics, and phonology. While discourse competence means the capability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances. In other words, it is concerned with intersentential relationships. Sociolinguistic competence is the knowledge of socio-cultural rules of language and of discourse. It requires an understanding of social context in which the language is used. Finally, strategic competence means the way of manipulating language in order to meet communicative goals.

Furthermore, Brown (2001:343) points out that in order to make the readers, as the addressees, able to interpret the linguistic message produced through the process of writing, a writer needs to acquire the sub skills needed, as shown below:

- 1) Produce graphemes and orthographic patterns of English,
- 2) Produce writing at an efficient rate of speed to suit the purpose,
- 3) Produce an acceptable core of words and use appropriate word order patterns,
- 4) Use acceptable grammatical system (e.g. tense, agreement, pluralization), patterns, and rules,
- 5) Express a particular meaning in different grammatical forms,
- 6) Use cohesive devices in written discourse,
- 7) Use the rhetorical forms and conventions of written discourse,
- 8) Appropriately accomplished the communicative functions of written texts according to form and purpose,
- 9) Convey links and connections between events and communicate such relations as main idea, new information, given information, generalization, and exemplification,
- 10) Distinguish between literal and implied meaning when writing,
- 11) Correctly convey culturally specific references in the context of the written text,
- 12) Develop and use a battery of writing strategies, such as accuracy assessing the audience’s interpretation, using prewriting devices, writing with fluency

in the first drafts, using paraphrases and synonyms, soliciting peers and instructor feedback, and using feedback for revising and editing.

WRITING IN EFL CLASSROOM

As one of the four language skills, writing has occupied a place in most English syllabuses. Even so, arguments are sometimes put forward for not teaching students to write because it is felt that a command of the spoken language and of reading is more important. For many students this may be true, but today, given the importance of English as an international language, more and more people need to learn to write in English for occupational or academic purposes. Another thing that makes the ability to write in English important is the fact that now English is the world's major language for the communication of research findings (Swales, 1990). Thus, the ability to write a scientific essay is a major goal of tertiary education. Because of the reason above, instead of being the last skill taught, writing has now become a much more important subject to be taught in second/foreign language curriculum (Leki, 1994).

According to Rivers (1987), writing in language classroom can be classified into four. The first type is writing or noting down which means imitating or copying. An example of this is to copy English sentences or paragraphs. The second is writing in the language or writing practice. The activity might be in the form of writing sample dialogues, uncomplicated translation exercises, dictation, and cloze procedures. The third type is translation which refers to transferring passages or sentences from the native to the target language. The fourth and the highest level of writing is called expressive writing. It deals with the expression of original ideas in the target language. Based on the types of writing above, Rivers (1987) summarizes writing as the expression of ideas in a consecutive way, according to the graphic conventions of the language—the ability to express himself in a polished literary form which requires the utilization of a special vocabulary and certain refinements of structure.

When students start to learn to write, at the beginning they will simply write down English words and might be sentences. Slowly then, the students will learn how to produce a sequence of sentences that are arranged in a particular order and linked together in certain ways. It is not easy to produce such a well organized piece of writing. Learning to write in either a first or a second language is one of the most difficult tasks a learner encounters and one that few people can fully master. Learning to write well is a difficult and lengthy process, one that induces anxiety and frustration in many learners (Richard, 1990).

In EFL context, it is believed that developing writing skills is more complicated than developing other language skills. Unlike speaking, for instance, writing means producing a written message without the presence of the intended readers. In this case the writer has to imagine the situational context and the roles of his readers (Hughes and Porter, 1983). In order to be able to write well, the foreign language learners need to be equipped with early and continued writing experiences. As Byrne (1988:48) suggested that the writing program must be planned carefully in order to develop the mastery of the writing skill which the learner can use for a continually expanding range of tasks. Since the

writing skill will still be guided to large extent, the English teachers are supposed to vary the activities given to the students in order to avoid monotonous approach. Thus, the teacher's task to develop the students' writing skill is more complicated than the other skills.

WRITING TEST

There are so many experts proposed what kinds of aspects should be judged in writing test. First is Inman et al. (1979) who asserted that there are five aspects should be assessed in a composition. They are logic, organization, development, style, and mechanics. While Jacobs (1981) proposed five aspects as well for ESL context, namely content, organization, vocabulary, language use, and mechanics. Harris, (1974), in addition, proposed five general components: content, form, grammar, style, and mechanics. Moreover, Heaton (1988) states five general components or main areas for writing good prose such as language use, mechanics, content, style, and a judgment skill that is ability to write in an appropriate manner for a particular purpose with a particular audience in mind, together with an ability to select, organize and order relevant information. It is in line with Burgess & Head (2005) statement that an answer in writing test that has some errors but achieves its communicative purpose will get a higher mark than an answer that is grammatically accurate but does not meet the task requirement. Other experts supported this mark by saying that effective writing ability can be reached through a combination of sociocultural competence, involving appropriate conventions, register, and (rhetorical) style; discourse competence, involving ideas and their structuring, coherence, and cohesion with an intended audience in the mind; and linguistic competence, involving appropriate and broad lexis, fluent and accuracy syntax, and accurate mechanics.

Furthermore, Langan (1985) has something different in evaluating essay writing. He proposed five areas to be based for assessment in writing. They are unity, support, coherence, and sentence skills. However, in fact, those elements of writing have something in common; that is they are more or less the same. Language, for example, has two elements. They are sentence structure and diction. Mechanics has four points. They are paraphrasing, punctuation, spelling, and capitalization. Style has five aspects. They are economy, simplicity, clarity, congruity, and courtesy. Organization is the rhetorical form, while logic is something to do with content. In short, among several features of composition to be assessed, they have many things in common. They can be grouped into four: content, organization, language, and style.

APPROACHES TO WRITING ASSESSMENT

Experts in writing assessment share something in common that there are two basic approaches in the assessment of writing, namely direct and indirect assessment. The former requires the students to write the actual composition by organizing their ideas into a unified text. While the indirect assessment, on the other hand, does not ask the students to write but it asks them to respond to question about writing or to do a particular task related to writing skill. In other words, they have to perform their knowledge about writing.

Direct Writing Assessment

The most direct way of measuring students' writing ability is to have them write (Harris, 1974). That is why when assessing their writing ability they must be asked to write. In a composition assessment, the students should be presented with a clearly defined problem which motivates them to write. Heaton (1975) says whenever possible, meaningful situation should be given to provide necessary information for writing because: (1) situation determines the register and style to be used in the composition, (2) it gives the possibility to obtain greater degree of reliability in scoring, and (3) it has an excellent backwash effect on the teaching and learning preparatory to the examination.

Indirect Writing Assessment

Writing can be divided into more specific 'discrete' elements, e.g grammar, spelling, vocabulary, punctuation, and orthography and attempts can be made to test these formal elements separately by the use of objective tests. These tests would be indirect ones in that they would only be measuring parts of what we take to be the construct of writing ability. What are tested may be related to proficient writing, but they cannot represent what proficiency writers can do (Weir, 1995). It would be very difficult to generalize from these types of test as to how students might perform on more productive tasks which require construction of a complete text. It would be difficult from these discrete item tests to make direct statements about how good a writer or what he or she can do in writing. Such indirect assessments have an extremely negative washback effect on the teaching that precedes them.

The formats that can be used to assess writing skill can be classified into three, i.e., open-ended, close-ended, and restricted response (Carol, 1980). In an open-ended writing assessment, the students are given an opportunity to initiate, elaborate or modify and to adopt the attitude they consider appropriate in the response to the writing task. The nature of this format is closer to the type of the test that is based on real-life performance and setting, that is writing for academic and social purpose in the real world (Weir, 1995).

PROCEDURES OF SCORING

As mentioned previously, there are two approaches that can be used to measure the students' writing ability, namely the direct and indirect measurement. Considering that each approach result in different tasks, a different procedure in scoring is applied accordingly. In a direct measurement, the score that the students get is derived from the rater's judgment on the basis of the pre-determined criteria stated in the scoring guide. On the other hand, in an indirect measurement, the score is obtained from an objective scoring procedure which is based on an answer key.

Scoring in Direct measurement

There are some proposed scoring guides by some experts deal with how to evaluate writing in direct measurement of writing. First is coming from Jacobs et al (1981) who classified the direct measurement of writing into two major types. They are the holistic scoring and the frequency count marking. In the

holistic scoring, a piece of writing is viewed as a whole and complete idea rather than as a separate element. The rater bases his judgment on his impression of the composition and he might be guided by a holistic scoring guide in scoring the composition. While the frequency count marking, on the other hand, is a procedure in evaluating a piece of writing by tallying and/or enumerating certain elements of the composition, like the number of cohesive devices, spelling errors, grammatical errors, punctuations errors and things like that.

Another classification of direct measurement in writing is introduced by Spandel and Stiggins (1990). They classify writing assessment into three types: primary trait, holistic, and analytic scoring procedure. Primary trait scoring is a procedure in scoring a piece of writing by focusing on the domain trait of the piece, such as on descriptive, narrative, and argumentative writing. Holistic scoring means scoring a piece of writing as a whole where each paper receives only one score. The final score is not the total of sub scores. The difference between primary trait scoring and holistic scoring is on the emphasis. In the first procedure, a different mode of writing has a different scoring guide depending on the types of discourse whereas in holistic scoring, there is no specific emphasis. That is why the holistic scoring can be applied to all types of discourse. Analytic scoring, unlike the first two procedures, scores a piece of writing by referring to a list of features or sub skills on which a rater bases his judgment. The writing quality is shown by the total of the sub scores.

Still another classification of direct measurement in writing is given by White (1985), and Brown (2005), who classify evaluation on writing into two basic scoring procedures. The first is holistic scoring and the second is analytic scoring. In holistic scoring, a rater judges a piece of writing as a whole without any separable aspects and their sub-scores. The holistic evaluation must come up with a single score which does not result from summing up the sub-scores. Analytic scoring, in contrast, come up with a single score resulting from summing up of the sub-scores which are derived from the scoring of the features, or aspects of the piece.

In short, there are two types in common for writing scoring. If the procedure of scoring is based on the analysis of features, it is called analytic. When the scoring is based on the judgment of rater(s) as a whole without separating features or aspects, it is called holistic.

Scoring in Indirect Measurement

Indirect measurement of writing ability as stated earlier is basically a measurement of knowledge about writing or sometimes it is called a measurement of editorial skills. This is due to the fact that the students are required to respond to questions about writing in an objective-type test, such as multiple choices, or a completion test depending on the purpose or on what aspects of writing to be measured. The type of task varies such as choosing the most suitable topic sentences from a given option, identifying irrelevant sentences in a paragraph, filling in the missing transitional signals, and many others.

Actually in term of reliability, indirect measurement has higher reliability coefficient than its counterpart as it is an objective-type test. This is

due to the fact that indirect measurement uses an objective scoring system with definite answer, so there will be no subjective judgment. A direct measurement, in contrast, is usually difficult to obtain high reliability coefficient for the scores depend on the raters' judgment. Avoiding subjectivity is extremely difficult. To overcome this problem, at least two raters, even more, are needed. A third rater is required in case the scores from the two raters are more than pre-determined maximum acceptable difference of scores.

DEVELOPING PROMPTS FOR WRITING TEST

Developing prompts for writing test should be done in carefully way since the prompts used in the test will determine the successfully of the test itself. In selecting the appropriate writing prompts, there are some aspects to be taken into consideration. California Assessment Program (in O'Malley and pierce, 1996) proposed that each prompt used in writing assessment should meet the following criteria. First, the writing prompt invites the desired type of writing or genre. Second, it engages the thinking, problem solving, composing, and text-making processes central to the type of writing. Third, it is challenging for many students and accessible to all. Fourth, it provides equitable opportunities for all students to respond. Fifth, it produces interesting, not just proficient writing. Sixth, it is liked by many students.

Weir (1995) emphasized the importance of considering the students' background knowledge when developing the test. This is due to the fact that students write best when they find something they can write on. Concerning the topics presented in a writing test, Reid (1993) asserted that effective writing topics are the ones that (1) interest both the students and the readers, (2) are accessible to all students, (3) involve the students in the topic, (4) are comprehensible to the students, (5) are unbiased with regard to the culture of the students, (6) are not culture bound (that is, they do not require intimate knowledge of a particular culture, especially if the test is designed for the students who come from different cultural backgrounds), and (7) allow the students to learn while they write.

Based on the experts' argument above, the prompt used in this test writing are developed in such way that they have the main factors as following; (1) the prompt do not require an ambiguous task so that the students do not waste their time trying to figure out what is called for, (2) the prompt specify a particular purpose mode of writing that is required only argumentation, (3) the prompt offers interest to all the students so that they wrote with some genuine concern, (4) the prompt specify the amount of time the students will have to complete the writing, (5) the prompt specify the length of the text produced by the students, (6) the last part of prompt tell the students what will be valued in the writing.

VALIDATING WRITING TEST

A test can be used as a valuable device in giving overall impression about teaching learning process has been done. By administering testing, the teacher can evaluate the effectiveness of the syllabus as well as the methods and materials (Heaton, 1988:7). However, to do so the test should fulfill some criteria

of being a good test. A good language test should possess three qualities i.e. validity, reliability, and practicality (Harris, 1974).

Validity

Every test, whether it be a short, informal classroom test or a public examination should be as valid as the constructor can make it (Heaton, 1988:159). Validity, according to Ebel & Frisbie, 1986 in Latief, 2000:98) refers to the appropriateness of making specific inferences or of making certain decision on the basis of scores from a test. In other words, the test must aim to provide a true measure of the particular skill which is intended to measure: to the extent that measures external knowledge and other skills at the same time, it will not be a valid test (Heaton, 1988:159).

Differ with conventional concept that claims there are so many kinds of validity; Djiwandono (2008:165) point out that validity is unitary concept. It means that validity conceptually is only one kind, the variousness are lies on how to prove the validity. They are mainly three ways to give a support or evidence to validity, and we can choose one of the most appropriate one to get evidence of our test validity. They are content validity, criterion-related validity, and construct validity.

Content validity is a kind of validity that depends on a careful analysis of the language being tested and of the particular course objectives (Heaton, 1988:160). It means that a test should contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent. Furthermore, Gronlund (1985) in Latief (2000:1) states that content validation is the process of determining the extent to which a set of test tasks provides a relevant and representative sample of the domain of tasks under consideration. Heaton (1988:160) stated that if we want to use content validity to give evidence to our test, the test writer should first draw up a table of test specifications, describing in very clear and precise terms the particular language skills and areas to be included in the test. Then the important point is that the test writer has attempted to quantify and balance the test components, assigning a certain value to indicate the importance of each component in relation to the other components in the test. By so doing, the test will achieve content validity and reflect the component skills and areas which test writer wishes to include in the assessment.

As cited from Djiwandono (2008:165), criterion-related validity can be proved by comparing between the students' score obtained and the scores from the similar test which has been marked as a good test at the same time, we speak of 'concurrent validity'. For instance, we compare the students' achievement in English course and the students' scores in TOEFL. When the correlation between the students' achievement in the course and their achievement in TOEFL test is high, the result of the test/measurement in the English course has strong criterion validity evidence. Therefore, the result of the English test is believed to have high concurrent validity. On the contrary, when the comparison of the two test results makes low correlation, the result of the test in the English course is said to have weak or low concurrent validity evidence. If it concerns the degree to which a test can predict the test takers'

performance in the future, it is called 'predictive validity'. For instance, with placement test: once courses are under way, we can check validation by establishing the proportion of the students who were thought to be misplaced (Hughes, 1989). This type of validity has also been called 'empirical', 'pragmatic', or 'statistical' (Bell, 1981:198).

Whereas a test will be said has construct validity if it is capable of measuring certain specific characteristics in accordance with a theory of language behavior and learning (Heaton, 1988:161). This type of validity assumes the existence of certain learning theories or constructs underlying the acquisition of abilities and skills. For example, if the assumption is held that systematic language habits are best acquired at the elementary level by means of the structural approach, then a test which emphasizes the communicative aspects of the language will have low construct validity. Conversely, if a communicative approach to language teaching and learning has been adopted throughout a course, a test comprising chiefly multiple choice items will lack construct validity.

In Latief (2000:100), the aspects of the test which can be used as the construct validity evidence are the tasks or activities required by the test that the students should perform. A test should always ask the test takers (students) to perform particular tasks or activities. Based on the tasks or activities that the students should perform in the test is the basis of scoring. They are the construct validity evidence. When the tasks or activity that the students should perform and they are the basis of scoring reflect the students' behavior which are being measured, the result of the measurement or the test has a strong construct validity evidence. Therefore, the result of the test is believed to have a high validity. On the contrary, when the tasks or activities that the students should perform in the test and they become the basis of scoring do not reflect the students' behaviors which are being measured, the result of the measurement or the test does not have strong construct validity evidence. Therefore, the result of the test is believed to have weak construct validity.

Reliability

Reliability of the result of language test refers to the preciseness of the result in presenting the actual level of the language proficiency of the examinees (students) (Latief, 2001:214). If the test is administered to the same candidates on different occasions, then, to the extent that it produces differing results, it is not reliable (Heaton, 1988:162). Reliability measured in this way is commonly referred to as test/re-retest reliability to distinguish it from mark/re-mark reliability. Another common reliability denotes the extent to which the same marks or grades are awarded if the same test papers are marked by (i) two or more different examiners or (ii) the same examiner on different occasion. In short, in order to be reliable, a test must be consistent in its measurements.

Since there are so many kinds of test for language proficiency, so there are many ways as well in calculating the level of reliability (see Djwandono, 2008:171). The calculation of reliability level always required two sets of scores to measure the consistency of the test. Correlation coefficient as a measurement of the consistency of the test can be got by calculation method. There are so

many methods of estimating the reliability test. Based on the score obtained to calculate correlation coefficient, there are eight kinds of reliability. They are (1) test-retest reliability, (2) equivalent-forms reliability or alternate-forms reliability, (3) split-half reliability, (4) cronbach alpha reliability, (5) cronbach alpha for writing ability, (6) Kuder-Richardson (KR) reliability, (7) scorer reliability or rater reliability, and (8) estimated reliability.

The approval of reliability is completely an empirical matter in which it involves statistical analysis. The statistical analysis is used to show the correlation in a various level. It is expressed in the form of correlation coefficient. Since the reliability is a correlation in a various level, therefore, reliability is actually a spectrum of level and is not dicotomically reliable and unreliable. The reliability is spread from the highest to the lowest with some levels in between. It is a form of a continuum of coefficient. The highest reliability is statistically figured as 1.00. The reliability with value 1.00 means the score has absolute consistency without any deviation at all. This kind of reliability is theoretical, because in the reality almost there is no result of measurement which is absolutely consistent without any difference at all, moreover in the measurement of a multi-aspect subject such as language teaching. In practicing, the level of reliability is usually found to be lower than the absolute correlation coefficient (1.00) that is 0.95, 0.90, 0.70, and so forth.

Practicality

There are so many experts in language testing discuss about the practicality of the test. According to Djiwandono (2008: 190), practicality of a test does not have relationship with something abstract or theoretical, but it is something done with its application, mainly in 1) the practicality of administering the test, and 2) the financial aspect. It is line with Bell (1981:200) that involved two parameters in determining practicality that is economy (in terms of money and time) and ease. In addition, Harris (1974) asserted that a test is said to be practical if it is economical in terms of cost and time, easy to administer, score, and interpret.

TEST DEVELOPMENT PROJECT (TDP)

Since testing or assessment can not be separated from the language teaching practice, we as a practitioner (a lecturer) must be able to make a good test. However, it is not as easy as we can imagine. To make a test everybody can, but to make a good one? No one can guarantee that every teacher/lecturer is able to make a good one. Fortunately, Prof. Dr. M. Soenardi Djiwandono, a professor of language assessment of State University of Malang, provides us systematic, comprehensive step by step procedures to develop a good test called Test Development Project (TDP).

I can say that this Test Development Project (TDP) really works as it guides us through systematic, comprehensive, step by step procedures. It is systematic as we can not go on to the next step before the first step is truly done. It is so, as the first step is a foundation of the next step. While it is said comprehensive since the TDP comprises every aspect or component of any language test. In the TDP proposed and developed by Prof. Dr. M. Soenardi

Djiwandono, there are five stages of developing any language test—language skills or language component as well. They are (1) preparation, (2) test plan, (3) test development, (4) test validation, and (5) test try-out.

The first stage in TDP is preparation. In this stage, we must identify the prospective test takers such as the institution, the level, number of the students, and other related information to test writing. It is very important as different tests are intended for different institutions or schools, or groups of test takers, levels, and the like.

The second step is test plan. Finishing first step, we go to plan the test. Starting from the available general and specific course objectives, we collect the data. If the available course objectives are not appropriate, we proposed the appropriate one and of course with justification. Then we determine the general and specific test objectives. The general test objective must be in line with the general course objective and specific test objectives must be in line with the specific course objectives. Then we determine the kind and format of the test we developed. Determining those two aspects are very crucial as it is a reflection whether the test we developed is appropriate. Moreover, every kind of language testing has its format. For example, you can not use multiple objective tests for speaking test. After we decide the most suitable format for our test, we also have to make Table of Specification (TOS). TOS covers specification of what we will test and its proportion. Then we need to identify the source of the test materials, the text used in the test, and the test item.

After completing the second stage, the next step is test development. Here we develop the test draft, determine the answer key or scoring guide we will use to check the students' answer, and determine how the way to score and how scores are converted to grades.

Finishing completing the third stage, the next is test validation. To validate our test, Prof. Soenardi proposed three steps. First, we give our test draft to the expert. It means we seek for the expert validation. The expert will review of the entire test we made and we will get a feedback from the expert concerning to our test draft's content. Then we revise our test draft based on his/her feedback. Next, we determine what kind of validity we will use and how to apply it. As what Prof. Soenardi stated that validity is unitary concept that means we can choose the appropriate one to validate our test. For reliability, the last step in TDP is conducting the try-out. After we do the try out for our test, we analyzed the result of our test. We identify the problems we found when conducting the try out and the weaknesses of our test. Then we improve and revise our test to be a better one.

CONCLUSION

Every teaching and learning process needs an evaluation. In this paper I have briefly touched upon several issues related to assessment that writing teacher should be aware of. A solid understanding of assessment issues should be part of every teacher's knowledge base, and teacher should be encouraged to equip themselves with this knowledge as part of their ongoing professional development.

REFERENCES

- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: O.U.P.
- Bell, R.T. 1981. *An Introduction to Applied Linguistics*. London: Batsford Academic and Educational Ltd. (Appendix C, "Language Testing").
- Brown, H. D. 2001. *Teaching by Principles. An Interactive Approach to Language Pedagogy* (2nd ed). White Plains, New York: Pearson Education.
- Brown, J.D. 2005. *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. New York: McGraw-Hill.
- Burgess, Sally & Head, Katie. 2005. *How to Teach for Exams*. Jeremy Harmer (Ed). Harlow: Pearson Education Limited.
- Byrne, D. 1988. *Teaching Writing Skills: Handbook for Language Teachers*. London: Longman Group UK Limited Company.
- Cohen, M., and Margaret, R. 1989. *The Effect of distance on Students' Writing*. American Educational Research Journal Vol.26, no.2, pp.143-159.
- Das, K Birkam (Ed). 1989. *Principles of Language Learning and Teaching*. Singapore: SEAMO Report Language Center.
- Djiwandono, M. Soenardi. 2008. *Tes Bahasa: Pegangan Bagi Pengajar Bahasa*. Jakarta: PT. Indeks.
- Eanes, R. 1983. *Content Area Literacy: Teaching for Today and Tommorrow*. Albany: Delimar publisher.
- Harris, D.D. 1974. *Testing English as a Second Language*. New York: McGraw-Hill.
- Heaton, J.B. 1988. *Writing English Language Test*. New York: Longman, Inc.
- Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: C.U.P.
- Inman, B.A & Ruth, G. 1979. *Aspects of Composition*. Second edition. New York: Harcourt Brace Jovanovich, Inc.
- Jacobs, H.L., Zinkgraf, S.A., Wormuth, D.R., Hartfiel, V, F. & Hughey, J.B. 1981. *Testing ESL Composition: A Practical Approach*. Massachusetts: Newbury House Publishers.
- Langan, J. 1985. *College Writing Skills with Reading*. New York: McGraw-Hill.
- Latief, M.A. 2000. *Validitas Hasil Pengukuran. Bahasa dan Seni*, 28 (1): 95-104.
- Leki, I. 1994. *Teaching Second-Language Writing: Where We Seem to Be*. In Karl, T. (Ed.), *Teacher Development: Making the Right Moves* (pp. 170-178). Washington, Dc: USIA.
- Reid, J. 1993. *Testing ESL Writing*. New York: Prentice Hall.
- Richards, J.C. 1990. *The Language Teaching Matrix*. New York: Cambridge University Press.
- Rivers, W. M. 1987. *Interactive Language Teaching*. Cambridge: Cambridge University Press.
- Savignon, Sandra J. 1983. *Communicative Competence: Theory and Classroom Practice*. Massachusetts: Addison-Wesley Publishing Company.
- Swales, J.M. 1990. *Genre Analysis: English an Academic and Research Setting*. Cambridge: Cambridge University Press.
- Troika, L.Q. 1987. *Handbook for Writers*. New Jersey: Prentice Hall Inc.
- Weir, C.J. 1995. *Understanding and Developing Language Test*. Singapore: Phoenix.

