

ASSESSING ENGLISH LANGUAGE TEACHING

Eka Sustri Harida, M.Pd.¹

email: esha_stainpasid@yahoo.com

Abstract

Asesmen atau kata lain disebut dengan penilaian merupakan upaya pengumpulan informasi terhadap hasil belajar siswa melalui bermacam instrumen. Hal ini bertujuan untuk mengetahui penguasaan siswa terhadap pelajaran yang telah dipelajarinya. Penilaian terhadap pembelajaran bahasa Inggris perlu dilakukan dengan lebih autentik agar apa yang diharapkan dari siswa dapat diketahui dengan benar. Penggunaan instrumen penilaian yang tidak baik atau tidak autentik akan menjadikan guru atau dosen salah dalam menilai siswa atau mahasiswa. Penilaian dalam *reading comprehension* dan *grammar* juga harus dilakukan secara baik dan autentik agar apa yang diharapkan dari pembelajaran tersebut dapat mencapai tujuan dan target yang diharapkan. Penilaian dilakukan harus mengacu kepada silabus dan proses pembelajaran yang sudah dilakukan. Hal ini dilakukan agar guru atau dosen dapat mengetahui secara lebih dalam penguasaan siswa atau mahasiswa terhadap materi yang telah diberikan.

Kata-kata kunci:

Assessment, language teaching, reading, and grammar.

Introduction

Assessment plays an important role in English language teaching; while assessment and learning has emerged as a prominent strand in the level of education. Assessment as judging achievement draws attention to the nature of assessment as the exercise of professional judgment, standing in contrast to misplaced notions of assessment as measurement. Assessment is aimed to know the learners' outcomes in their learning process. In order to discuss about learning assessment, the term of assessment is needed to be explained first. Assessment is the act of judging or deciding the amount, value, quality, or importance of something, or the judgment or decision that is made. Further, assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development. In other word, assessment means "the process of gathering and discussing information from multiple and diverse sources in order to develop a deep understanding

¹A Lecturer of English Department in Faculty of Tarbiyah and Teacher Training, IAIN Padangsidempuan.

of what students know, understand, and can do with their knowledge as a result of their educational experiences; the process culminates when assessment results are used to improve subsequent learning." It can be concluded that assessment is the process of gathering information after teaching and learning process; in developing learners understanding about the materials.

The concepts of assessment, learning and judgement draw together the three core functions of assessment. While assessment can fulfil many functions, three predominate: supporting the process of learning; judging students' achievement in relation to course requirements; and maintaining the standards of the profession or discipline for which students are being prepared. Each of these is important, with each having particular imperatives and each giving rise to particular issues of conceptualization and implementation.

Based the important of assessment, in this paper will explain about assessing reading comprehension as a second language and assessing grammar also in the context of English as a second language.

Assessing Reading Comprehension

Reading comprehension is mental process in which the readers try to understand the meaning in the text by interpreting what have been read in order to find the idea given by the writers. In reading comprehension, the readers should pay attention for some aspects such as how they understand the words and the information from the text, understanding main ideas and details, and be able to make conclusion. All of the aspects will be the indicators of this research on reading comprehension. The result of reading comprehension depends on kinds of texts, background knowledge of the readers, goal of reading, and tools which are used to values interpretation of reading.

In assessing reading in second or foreign language, it is necessary to have a working knowledge of what that process entails. In many ways, reading in a target language is like reading in the native language. In this paper, will discuss about reading in a native language, reading in a second language, assessing reading comprehension ability, types of reading assessed, types of meaning assessed, types of comprehension skills assessed, testing methods, principles for assessing L2/FL reading comprehension.

Reading in the Native Language

The first thing is reading is not a passive activity; it is a psycholinguistic guessing game in which the readers do a process to predict what could come next on the text they read. It is also stated as the process when the readers try to interpret or predict the meaning of the text. Reading is stated as passive process because; 1) when reading, readers may think that remember sentences that they read, 2) they may think that they are reading through a text without regressions, 3) reading in a smooth, linear, word-to-word fashion, their eyes actually jump from one fixation point to another, 4) readers think that they cannot understand a text.

So, reading in native language is not different with reading in L2. Reading for meaning involves the activation of networks of real-world and rhetorical information for the purpose of interpreting texts. The source of background knowledge related to the *schemata (content schemata, language schemata, and textual schemata)*. Content schemata tells about systems of factual knowledge, values, and cultural conventions, language schemata tells about sentences structure, grammatical inflections, spelling and punctuation, vocabulary, and cohesive structures, and textual schemata tells about the rhetorical structure of different modes of text-for example, recipes, letters, fairy tales, research papers, and science textbooks. In this reading, the readers focus exclusively on what is present in the text itself, on the top-down reading.

Reading in a Second Language

Reading a second language is slower and less successful, it is because the readers are spending more time at each fixation, it is also happening to the eye movement when reading the L2 text eye movement is three times along than reading L1. The low reading in L2 also can be caused of unknown vocabulary words, sentence structure, and other. When reading L2, readers may try to host of busy bottom-up reading activities-e.g. Word analysis, use dictionary, and analysis sentence structure – without coming away to the clues about word, phrase, and sentence, or passage means.

Assessing Reading Comprehension Ability

The areas that could be assessed in reading comprehension ability are 1) awareness and use of cues in text, 2) perception of the author's plan, 3) insight into the

nature of the text, 4) strategies for retrieving information after reading, 5) schemata selection, 6) cognitive monitoring skills, and 7) the knowledge of summarization rules. The assessment can be done by using verbal report as a way to get strategy information. Based on Johnston that the best reading assessment consisted of having teachers observe and interact with students who were reading authentic texts for genuine purposes, and of having teachers see how their students constructed meaning.

Types of Reading Assessed

Types of reading assessed are readers 1) give a lengthy passage to read by skimming, scanning, and searching reading, 2) give a passage to read receptively or intensively, and 3) read responsively.

Types of Meaning Assessed

Types of meaning assessed are *grammatical meaning, informational meaning, discourse meaning, and the meaning conveyed by the writer's tone*. In *assessing grammatical meaning*, there is the role of syntax that is assessed in it. Assessment focused on the syntax of the text and made assumptions about the difficulty of second language text because of its syntax. Research suggested that sentences complexity was likely to be an impediment to grasping special details, while confirming that core syntactic comprehension was enough for getting the gist. It is important to employ a syntactic reading strategy to make easier in understand text. *Assessing informational meaning* of the text, the purpose is to determine what the basic concepts or messages are that the text wishes to convey; it is done by analysis of vocabulary by distinguishing known, possibly known, and unknown words, concepts, relationships, and the like. *Assessing discourse meaning* is to know the meaning from the literature in recent years, especially for the perception of rhetorical functions conveyed by text. *Assessing the meaning conveyed by the writer's tone* is assessing reading based on author's tone-be it sad, cynical, sarcastic, caustic, or humorous-may be signaled by the use of one word or phrase, possibly inducing a play on words.

Types of Comprehension Skills Assessed

There are some numerous taxonomy related to types of comprehension skills assessed, the taxonomies include: 1) recognition of words and phrases of similar and

opposing meaning, 2) identifying or locating of information, 3) discriminating of elements or features within context; as causal, sequential, chronological, hierarchical, 4) interpreting of complex ideas, actions, events, and relationships, 5) inferring-deriving of conclusions and predicting the continuation, 6) synthesis, and 7) evaluation. The taxonomy omits the reader – writer relationship. The other is also about getting the specificity of the information in the text, emphasis on language for specific purposes (LSP), especially at the tertiary education level.

Testing Methods

Testing methods for reading can be: 1) communicative test; 2) fixed-response formats: multiple choice an alternative; 3) structured-response formats; 4) testing of vocabulary; and 5) computer based testing. *Communicative test* is related to communicative language testing, in which has a feedback potential rewarding with clear, rich, relevant, and generalizable information. It is a communicative assessment deal with *a storyline test*. *Fixed-response formats* are multiple choice and alternative to multiple choice test (summary writing, metacognitive judgment, question selection, multiple acceptable responses, and prior knowledge). *Structured response formats* are like cloze the cloze, the C-test, and recall protocols.

Testing Vocabulary Knowledge

Vocabulary testing may not be very informative if vocabulary tests access a limited sampling of the learner's total vocabulary and if the level is too superficial for the depth of vocabulary development to be determined. The challenge is to assess how ell learners know specific vocabulary words. The approaches to assess vocabulary are four, the three of them are for assessing vocabulary recognition (identifying the meaning by multiple choice, matching, and checking off familiar words), and one is for assessing vocabulary production (combining self-report with performance). *Computer – based testing* which involves the use of the computer as a vehicle for assessment instead of paper and pencil, while computer adaptive testing adds the dimension of having the computer program select test items according to how the respondent has done on previous items on that test. In computer based testing, the use of test is multiple choice test; it can be in the format of *multiple choice, deletion, insertion, correction,*

completion, transformation, reorganization, and matching; it is begun to include efforts at scoring open-ended items.

Principles for Assessing L2/FL Reading Comprehension

The principles of assessing second language or foreign language reading are: 1) choose a text with a familiar topic, that is interesting, has unambiguous intent, and is of an appropriate length; 2) where appropriate, allow students to do some or extensive conceptualizing in their L1; 3) determine the reasoning behind the students' conclusions; 4) design the assessment tasks so that the students can demonstrate the schemata that they have for the content, textual organization, and language of the given text(s); and 5) be flexible about acknowledging individual interpretations that students may have for texts.

Assessing Grammar

Performance on grammar tests can be influenced by many other factors besides grammatical ability. In fact, test scores can vary as a result of the personal attributes of test takers such as their age (Farhady, 1983; Zeidner, 1987), gender (Kunnan, 1990; Sunderland, 1995) and language background (Zeidner, 1986, 1987). They can also fluctuate due to their strategy use (Cohen, 1994; Purpura, 1999), motivation (Gardner, 1985) and level of anxiety (Gardner, Lalonde, Moorcroft and Evans, 1987). However, some of the most important factors that affect grammar-test scores, aside from grammatical ability, are the characteristics of the test itself. In fact, anyone who has ever taken a grammar test, or any test for that matter, knows that the types of questions on the test can severely impact performance.

The goal of every grammar test was to obtain (and provide) information on how well a student knows or can use grammar to convey meaning in some situation where the target language is used. Score-based inferences from grammar tests can be used to make a variety of decisions. The information derived from language tests, of which grammar tests are a subset, can be used to provide test-takers and other test-users with formative and summative evaluations. Formative evaluation relating to grammar assessment supplies information during a course of instruction or learning on how test-takers might increase their knowledge of grammar, or how they might improve their

ability to use grammar in communicative contexts. It also provides teachers with information on how they might modify future instruction or fine-tune the curriculum.

To assess students' grammar ability, the teachers must be considered what is important to be tested. In this sub topic, will be discussed about grammar test tasks.

Describing grammar test tasks

When language teachers consider tasks for grammar tests, they call to mind a large repertoire of task types that have been commonly used in teaching and testing contexts. We now know that these holistic task types constitute collections of task characteristics for eliciting performance and that these holistic task types can vary on a number of dimensions. We also need to remember that the tasks we include on tests should strive to match the types of language-use tasks found in real-life or language instructional domains.

Some have classified tasks according to scoring procedure:

- ✓ objective test tasks (e.g., true-false tasks) are those in which no expert judgment is required to evaluate performance with regard to the criteria for correctness.
- ✓ Subjective test tasks (e.g., essays) are those that require expert judgment to interpret and evaluate performance with regard to the criteria for correctness.

Test tasks according to the types of response are selected response tasks, limited production task, and extended production tasks.

The Selected Response Tasks

The multiple-choice (MC) task

The answer or key represents the best, correct or most appropriate, acceptable or natural choice; the other options are the distractors. MC items are well suited for testing many discrete features of grammatical knowledge. While the MC task has many advantages, the items can be difficult and time-consuming to develop. The format encourages guessing, and scores might be inflated due to test-wideness, or the test-taker's knowledge about test taking. This can result in serious questions about the validity of inferences based on these items (Cohen, 1998). Finally, many educators argue that MC tasks are inauthentic language-use tasks.

Multiple-choice error identification task

This task presents test-takers with an item that contains one incorrect, unacceptable, or inappropriate feature in the input. Examinees are required to identify the error. In the context of grammatical assessment, the errors in the input relate to grammatical accuracy and/or meaningfulness. These tasks are often used in editing to identify grammatical errors.

The matching task

This task presents input in the form of two lists of words, phrases or sentences. One list can also be in the form of visual cues. Examinees match one list with the other. To avoid guessing, one list has one or more extra distractors. Matching tasks are designed to test several discrete features of grammatical knowledge within the same task. They are also designed to encourage test-takers to cross-reference and examine the relationships between the two lists so that construct-related associations can be indicated. They are also easy to score.

The discrimination task

This task presents examinees with language and/or non-language input along with two response choices that are polar opposites or that contrast in some way. Some response possibilities include: true-false, right-wrong, same-different, agree-disagree, grammatical-ungrammatical and so forth. As seen in Van Patten's (1996) 'interpretation tasks', the input consists of two contrasting images and one utterance. Discrimination items are designed to measure the differences between two similar areas of grammatical knowledge.

The noticing task

This task presents learners with a wide range of input in the form of language and/or non-language. Examinees are asked to indicate (e.g., by circling, highlighting) that they have identified some specific feature in the language. This item is designed to measure grammatical meaning. It would be scored right or wrong. The noticing task, also referred to as a kind of consciousness-raising (CR) task, is intended to help students process input by getting them to construct a conscious form-meaning representation of the grammatical feature (Ellis, 1997), and for this reason, it seems to be particularly effective in promoting the acquisition of new grammar points.

Grammaticality-judgment task

The grammaticality-judgment task presents learners with sentences that are either well or ill formed. Learners must then decide if these sentences are acceptable, or not.

Limited Production Tasks

Limited-production tasks present input in the form of an item with language and/or non-language information that can vary in length or topic. Different from selected-response tasks, limited production tasks elicit a response embodying a limited amount of language production. Limited-production tasks are intended to assess one or more areas of grammatical knowledge depending on the construct definition.

The gap-filling task:

1. This task presents input in the form of a sentence, passage or dialogue with a number of words deleted. The gaps are specifically selected to test one or more areas of grammatical knowledge. Examinees are required to fill the gap with an appropriate response for the context. Gap-filling tasks are designed to measure the learner's knowledge of grammatical forms and meanings.
2. the cued gap-filling task: the gaps are preceded by one or more lexical items, or cues, which must be transformed in order to fill the gap correctly.
3. the cloze: presents the input as a passage or dialogue in which every fifth, sixth or seventh word is mechanically deleted and replaced by a gap. Examinees have to fill the gap with the best word for the context.

The short-answer task

This task presents input in the form of a question, incomplete sentence or some visual stimulus. Test-takers are expected to produce responses that range in length from a word to a sentence or two. The range of acceptable responses can vary considerably. Short-answer questions can be used to test several areas of grammatical ability, and are usually scored as right or wrong with one or more criteria for correctness or partial credit. Short-answer tasks can also be scored by means of a rating scale.

The dialogue (or discourse) completion task (DCT)

The DCT presents input in the form of a short exchange or dialogue with an entire turn or part of a turn deleted. Examinees are expected to complete the exchange with a

response that is grammatically accurate and meaningful. DCTs are intended to measure the students' capacity to use grammatical forms to express a variety of literal or grammatical meanings (e.g., request), where the relationship between the form and the meaning is relatively direct. DCTs can also be used as a measure of pragmatic knowledge, in which case they could also be scored for sociolinguistic or sociocultural appropriateness, contextual acceptability, or naturalness depending on the purpose of the test and the construct(s) being measured. DCTs have been used extensively in applied linguistics research to investigate the use of semantic formulas and other linguistic devices to express a wide range of literal and implied contextual meanings (e.g., refusals, apologies, compliments). They have also been used to examine sociolinguistic and sociocultural meanings (social distance, power, register) associated with these contexts.

Extended Production Tasks

Extended-production tasks present input in the form of a prompt instead of an item. Extended-production tasks aim to elicit large amounts of data of which the quality and quantity can vary greatly for each test-taker. The quality of the extended-production task responses is judged (1) with reference to the theoretical construct(s) being measured and (2) in terms of different levels of grammatical ability or mastery.

The information-gap task (info-gap)

This task presents input in the form of two or more sets of partially complete information. Test-takers are instructed to ask each other questions to obtain one complete set of information. Being reciprocal in nature, info-gap tasks are intended to elicit data involving negotiated interaction and feedback, which can be used to measure the test-takers' ability to use grammatical forms to convey a range of literal functional meanings. Info-gap tasks are scored by means of the rating-scale method.

Story-telling and reporting tasks

These tasks present test-takers with prompts that require them to use information from their own experience or imagination to tell a story or report information. These tasks can be used to measure the test-takers' ability to use grammatical forms to convey several meanings – both literal and implied. This task aims to measure grammatical form and meaning in terms of the present and past modals used to express speculations

about the crime and the cohesive forms and meanings used in connecting sentences. Finally, this task measures pragmatic knowledge in terms to the examinees' ability to organize the sentences into a coherent report (rhetorical control).

The role-play and simulation tasks

These tasks present test-takers with a prompt in which two or more examinees are asked to assume a role in order to solve a problem collaboratively, make a decision or perform some transaction. The input can be language and/or non-language, and it can contain varying amounts of information. The purpose of the test and the construct definition will determine what will be scored. This task is intended to measure the test-takers' ability to understand the prompt, assume one of the roles and use their knowledge of grammatical forms and meanings to participate in the discussion.

Conclusion

For assessment, the process of operationalizing test constructs and the specification of test tasks are extremely important. They provide a means of controlling what is being measured, what evidence needs to be observed to support the measurement claims, what specific features can be manipulated to elicit the evidence of performance, and finally how the performance should be scored. In assessing reading comprehension, the types of meaning that are involved in reading comprehension, the comprehension skills involved, and testing methods. The discussion is also about communicative test reading comprehension, the storyline method, and others. Assessment of reading must be done in different kinds of methods or appropriate types that are suitable for the students to be measured to get the better result in students' assessment. Framework for characterizing test tasks and showed how it could be used to characterize SL grammar tasks. This framework allows us to examine tasks that are currently in use, and more interestingly, it allows us to show how variations in task characteristics can be used to create new task types that might better serve our educational needs and goals.

Bibliography

Coles, Martin and Jenkins, Rhonda. *Assessing Reading 2: Changing Practice in Classrooms*. Canada: the Taylor & Francis e-Library, 2002.

Cohen, Andrew D. 1994. *Assessing Language Ability in the Classroom*. Boston: Heinle & Heinle Publishers.

Merriam-Webster Dictionary assessed from http://medsci.indiana.edu/m620/reserves/def_assess.pdf

Purpura, James. 2004. *Assessing Grammar*. New York: Cambridge University Press.

G. Joughin (ed.), *Assessment, Learning and Judgement in Higher Education*.

<http://dictionary.cambridge.org/dictionary/british/assessment>

http://medsci.indiana.edu/m620/reserves/def_assess.pdf