

MODUL BAHASA INGGRIS UNTUK GURU SD

(PSD 225)

Materi 11

The Principles of Assessment English for Children

Disusun Oleh

Prayogo Hadi Sulistio, M.Pd

UNIVERSITAS ESA UNGGUL

2018

**The Principles of Assessment English for Children**

1. **Introduction**

This module becomes a part of Teaching English for Young Learners in English Education Department. Assessment for children is the fifth topics being discussed of this subject. It discusses how the teacher’s assets the student. This module starts with the definition and the opinion of expert regarding this topic. After that, the module gives explanation of how to assessing and evaluating as well as giving feedback to the student itself. At the end of this lesson, students are expected to be able to give their evaluation to the students.

This module can be used as reference for the students who would like to learn about The principles of Assessment english for children. The students may practice to teach according to what’s written in this module.

1. **Basic Competence**

Students are able to comprehend the basic principles of Assessment English for children

1. **Kemampuan Akhir yang Diharapkan**

* Students are able to applied the principles of Assessment English for children
* an awareness of the characteristics of assessing young learners
* an awareness of research projects which suggest best practice in assessing young learners
* an awareness of some suggestions to try in your practice

1. **Learning Activities 1.**

**The Principles of Assessment English for Children**

Cohen (1994) notes that “The **assessment** tasks are non-threatening and are developmental in nature, allowing the learners ample opportunities to demonstrate what they know and do not know, and providing useful feedback both for the learners and for their teachers” (p. 1).

Similarly, Brown (2004) indicates that a test means a way of measuring an individual’s learning progress, knowledge and performance in a certain field. Without a measurement that reveals the learners’ actual learning effectiveness, the teachers will not be able to access what are essential and reasonable teaching contents for learners. Through the results as well as students’ grades and scores evaluated from testing, the school policy makers would realize what good teaching methodologies are.

Numerous testing specialists argue that **assessment** practices should be guided on the basis of unbiased testing, that is, validity, reliability, practicality, authenticity and washback (Alderson 2001, 2002; Brown, 2004; Hughes, 2003). Hence, an assessment of the English subject should take above five components into consideration, which we think can be what Brown (2004) calls five criteria or principles of English **assessment and evaluation**.

These five components reflecting on the examination and improving the equitability of an examination are the successful key points to explain why a test is necessary and acceptable.

When an examination is designed through displaying the five significant characteristics above,

it can be regarded as a valuable and accurate one that not only reflects students’ real level but also provides information to administrators and teachers how the curriculum should be arranged. Based on the five components, this study investigates and introduces the concepts and the details in the significant five components of a high-quality test.

***Introduction: the Definition of a Test***

Many students tend to unavoidably feel apprehensive when being informed about a schedule of an examination. They might doubt if their performances of taking the test given by a school

instructor would be lower than their classmates. As a matter of fact, a test not only forces students to face their learning result compared with the whole class but also sometimes makes students change their feelings and relationships with their classmates. Through a transcript of an assessment, their talent in learning, acceptance for a certain teacher, and even their **intelligence quotient** (I.Q.) would be measured with a number as well as their score of a test.

No wonder a test has to be designed very cautiously and equitably. In this article, we emphasize that a psychological establishment and teacher’s words of encouragement for students’ preparation before a test is held are significant. It is a teacher’s responsibility to inform the students to treat the grade of an assessment with a relaxing and mature attitude. A score higher than the others means that this student can be a mentor for the rest of the students; in contrast, a lower score is a stimulation that raises the student’s awareness and reminds him/her that he/she should study harder in the future. All individuals should be educated before a test is held.

The goal of a measurement is for teachers and students to self-criticize their ways of teaching and learning. As Cohen (1997) noted:

For many, traditional testing has been viewed as a somewhat unpopular area of language

teaching and learning. Students have sometimes viewed tests as unfair measures of their

language ability and have feared that they would not perform well on them. (p. 2)

We believe that more sophisticated **measurements** of English subjects should not only be designed as a statistical tool for revealing students’ and teachers’ competences of teaching and learning, but also be established as a functional bridge that conveys the students’ and teachers’ effectiveness in learning and teaching. On the basis of a test concerning five crucial elements of **validity**, **reliability**, **practicality**, **authenticity** and **washback**, we believe that a 10-minute communication or brief lecture should be given before students are engaged in a test construction.

This should be an essential activity that would contribute to avoid students’ anxiousness, cheating or not trying hard to independently achieve a high score. In the following passages, we will give the readers an overview of what the **validity**, **reliability**, **practicality**, **authenticity** and **washback** refer to in this context.

***Validity***

According to Gronlund (1998), a test with **validity** is one trying to achieve “…the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment.” (p. 226) **Validity** of an assessment means that the extension of assessment results should make teachers’ teaching contents more appropriate, meaningful, and useful after the test. Brown (2004) explains validity in this way.

How is the validity of test established? There is no final, absolute measure of validity, but several different kinds of evidence may be invoked in support. In some cases, it may be appropriate to examine the extent to which a test calls for performance that matches that of the course or unit of study being tested. In other case, we may be concerned with how well a test determines whether or not students have reached an established set of goals or level of

competence. (p. 22)

In fact, Brown (2004) also indicated that validity construction can be demonstrated through five pieces of evidence of **validity**. In the following passages, this paper will introduce these five pieces of evidence that demonstrate how a test can be regarded as a **valid test** from the aspects of different scholars’ theories.

***Content-related Evidence***

The definition of **content-related** proof can be ideally presented by Mousavi (2002) and Hughes (2003). They declare that if an assessment in reality models the subject matter about which grades are to be concluded, and if it requires the examinees to carry out the performance that is being examined, it can assert **content–related** evidence of **validity**. That is to say, the test should reflect the content of the textbook and what teachers have conveyed in the classroom.

Brown (2004) believes that a test can be strengthened with its **content validity** by a good teacher. On the contrary, this also implies that many tests have no satisfactory **content related** evidence because of teachers’ poor control on this issue:

...but classroom teachers have neither the time nor the budget to subject quizzes, midterms, and final exams to the extensive scrutiny of a full construct validation. Therefore, it is critical that teacher hold content-related evidence in high esteem in the process of defending the validity of classroom tests... (p. 24)

In the light of above, teachers should pay more attention to the **content validity** in order to persuade their students to continue to listen to their lectures and follow their instructions in class. If a test can not fully reflect the content of the class progress, the students might not think the test or lecture of the class is worth their attention.

All in all, “**content-related evidence**” can be demonstrated through the test’s content of knowledge, information, and skills. They must reflect in the test with adequate sample items. In order to achieve **content validity**, the test takers can be directly and indirectly tested. They can either really perform the target task or just do a similar task.

***Criterion-related Evidence***

Second, “**criterion-related evidence**” as well as what Brown (2004) has called the “**concurrent validity**” means that the test result must be consistent with the other test’s result in a similar setting.

In the case of teacher-made classroom assessments, criterion-related evidence is best demonstrated through a comparison of result of an assessment with results of some other measure of the same criterion. For example, in a course unit whose objective is for students to be able to orally produce voiced and voiceless stops in all possible phonetic environments, the results of one teacher’s unit test might be compared with an independent assessment-possibly a commercially produced test in a textbook-of the same phonemic proficiency. (p. 24)

Further interpretations for criterion-related evidence that reflect the test validity can be, that is, in two tests under similar parameters such as test time, content, healthy condition ...etc. For one learner, his grades of the two tests should be parallel. Hence, we can also call this type of **validity** “**predictive validity**” because test results should be parallel, predictable and agreeing with future performances.

Consequently, once we find two similar settings for students to test, and the same student has two extreme scores, we would judge that the at least one result of the tests should be doubtful. Through an investigation into the whole components of two assessments, the researchers should find several differences in the parameters of two tests that cause the extreme scores on one student.

***Construct-related Evidence***

Thirdly, we need to observe if there is “**construct-related evidence**” in a measurement. This

means a test needs to tap into theoretical construct such as factors of language testing, including **pragmatic** and **strategic components** that consider a learner’s intelligence and language proficiencies equally important.

In an assessment considering **construct-related evidence**, a teacher should take both students’ linguistic constructs and psychological constructs into evaluation. That is to say, students’ **communicative competence** and proficiency (Linguistic portion) and **self-esteem** and **motivation** (Psychological portion) are evenly decisive for the individual’s score.

Therefore, Davison, Hudson and Lynch (1985) declare that tests are, in a manner of speaking, operational definitions of constructs in that they ‘operationally’ the entity that is being measured. Take **Test of English as an Foreign Language** (TOEFL) for an example. Although the oral component in TOEFL is omitted ten years ago, it is assumed by the test designers that the other parts, such as listening and writing, have already made positive correlation to the omitted speaking part.

Another example that we would like to offer for explaining **construct-related evidence** is an evaluation through an interview after a communication training course. If students’ contents of speeches in the test are more positive, confident and aggressive than the others, the evaluators should add more points to these students’ final scores and appraise the students’ abilities in pronunciation, grammatical, lexis, and strategy applications.

***Consequential Validity***

The fourth component of language test’s **validity** is named **consequential validity** that is associated with the accuracy in measuring intended criteria, the impact on the preparation of test-takers, and the social influences of a test’s interpretation and use (Brindley, 2001; Gronlund, 1998; McNamara, 2000).

Because of the consequence of a test’s interpretation and use taking place on the test takers, the students’ motivation, study habits, and attitudes of learning would be altered after they take the examination. Several scholars regard the **consequential validity** as really significant in a test because it could stimulate a learners’ motivation of learning due to the fact that a test’s interpretation could affect the learner’s future careers from the realistic social perspectives.

Reckase (1998) thinks that a responsible test developer would do to gain information to support the consequential basis of **validity** for a test early in the development. Also, he implies that the consequential basis of **validity** of evaluation would be monitored and reported during the life of an educational program by the test developers.

His statements actually reflect the present policies of many universities, in Taiwan, where chairs of English departments usually designed students’ mid-term and final examinations similar to TOEFL or **General English Proficiency Test** (GEPT) due to their possible contributions to students’ future careers from a practical aspect of students’ job search or studying abroad.

These administrators actually tend to apply the consequential validity of above two examinations in school and assist their students to finally achieve higher scores in TOEFL and

GEPT, which are regarded as criteria of entering a distinguished graduate school or a corporation promising an infinite bright future. Through the same logic, we can observe that many university administrators would allow their students to submit a certain high scores to achieve in the TOEFL and GEPT in order to wave some of their required Basic English general courses.

***Face Validity***

Lastly, the **face validity** should be observed if the following characteristics can be carried out

in an assessment.

1. The test should be a little bit difficult and reasonably challenging which would help students think and learn when taking the test.

2. The test should “…refer to the degree to which a test looks right and appears to measure judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers” (Mousavi, 2002, p. 244)

3. Face validity makes “… students viewing this assessment as fair, relevant, and useful for improving learning” (Gronlund, 1998, p. 210). It can be high if it includes well-designed time limit, directions, and format.

Brown (2004) sets six principles for teachers to follow and establish **face validity i**n their

test. His principles of face validity are:

A well-constructed, expected format with familiar tasks, a test that is clearly doable within the allotted time limit, items that are clear and uncomplicated, directions that are crystal clear, tasks that relate to their course work, and a difficulty level that presents a reasonable challenge. (p. 27)

**Face validity** can’t be easily tested by an administrator or even by an assessment expert. Based on the moral spirit of the test designers, a valid **assessment** should be established and the test takers’ anxieties could be decreased and their time spent could make them learn more from the procedure of the assessment. It is not meaningful if a test is too complicated and stressful, which might frustrate our students.

On the other hand, if a test is always trouble-free, the students might not admire the teacher and the subject matter. They might assume the knowledge gained in the class is disreputable, so that they would become proud. As a conscientious and reputable teacher and a test designer, we can not ignore the issue of face validity.

**Reliability**

According to Brown (2004), “A reliable test should be consistent and dependable” (p. 20) when the same tests in two different occasions are given to the same student, he or she should havea similar result. The **reliability** can be raised to be the highest if the factors that influence the results can all be carefully paid attention. Based on Cheng’s (2005) study, “To enhance the reliability of the data collection and analysis, the researcher employed two approaches:

(a) cross checking with existing data,

(b) inviting inter-coders.” (p.110)

The most significant four factors detrimentally affecting the learners’ grades are **student-related reliability**, **rater-reliability**, **administration reliability** and **test reliability**.

***Student-related Reliability***

First, **student-related reliability** can be originated from the examinees’ short-term illness, exhaustion, a bad luck, test takers’ physical or psychological weakness and disadvantaged test strategies applied in the assessment. Actually, it can be also related to their “like” or “not like” the teacher who teaches the subject matter. When the students like the teacher, they pay more attention and effort when taking the test.

On the contrary, if they do not care about how their teacher looks at them, they can pay no heed to the items of the test paper. Therefore, sometimes an outstanding student’s grade could suddenly become very low because of the above extraordinary causes. Hence, a stably commanding condition should be maintained by the test takers who desire to always achieve high scores.

The factors as **test-taker’s** “test wiseness” or their application of examination strategies for efficient test taking would reflect the definition of **student-related reliability** (Mousavi, 2002). We can find many references that teach our students how to avoid the weakness during the period of taking the best, and how to raise the effectiveness of preparing for the test, answering the items and achieving the highest score. In fact, many famous teachers in Bushiban (cram schools) would talk a lot about testing strategies in their lecture, which might be useful in achieving a high grade in the competitive examinations.

***Rater-reliability***

Second, the **rater-reliability** is an issue that may contribute to the unreliability of a test. Rater-reliability contains **inter-rater reliability** and **intra-rater reliability**. **Inter-rater reliability** occurs when two or more scores yield very incompatible scores of the identical test taker, which can happen because the raters’ lack of concentration to scoring criteria, immaturity in grading skills, or inflexible biases.

In contrast, intra-rater reliability is another frequent event for classroom teachers who tend to give scores based on their own belief. This means that the teacher would give scores without really looking into the examinees ability, but personally and independently give grade according to their own impressions toward students. Sometimes, in Taiwan, we call this is Impressionism score, which means that the teacher might give score to the students based on their “like” and “do not like” toward the students.

Besides, intra-rater bias can also be caused by the teachers’’ absent-mindedness, which

means the teacher does not pay enough attention to the students’ proficiency and do not have any concept about the students’ competences and capability. Also, the teachers might be too busy doing research on their own and forget their students’ study status and progress. **Intra-rater** bias also can take place because of the raters’ indistinguishable grading criterion, tiredness, prejudices on certain students, and stubborn impressions on particular excellent or awful students.

Brown (2004) described **Inter-rater reliability** as follows.

Human error, subjectively, and bias may enter into the scoring process. Inter-rater reliability occurs when two or more scorers yield inconsistent scores of the same test, possibly for lack of attention to scoring criteria, inexperience, inattention, or even preconceived biases….”

(p.21)

He also stated what the **Intra-rater reliability** is. “Intra-rater reliability is a common occurrence for classroom teachers because of unclear scoring criteria, fatigue, bias toward particular good and bad students, or simple carelessness.” (Brown, 2004, p. 21)

***Test Administration Reliability***

The third element that might influence students’ grade is **test administration reliability**, which depends on the situations where the examination is conducted, controlled and administrated. The controlling for silence, peace atmosphere and relaxing temperature of the test surroundings can be the keys to attaining high test administration reliability. Moreover, the stipulations of desks, chairs and equipments of testing, like notebooks or tape-recorders are all very influential on students’ score distributions. Brown explained test **administration reliability** as follows.

Unreliability may also result from the conditions in which the test is administrated. I once witnessed the administration of a test of aural comprehension in which a tape recorder played items for comprehension, but because of street noise outside the building, students sitting next to windows could not hear the tape accurately. This was a clear case of unreliability cause by the conditions of the best admonition. (p. 21)

Indeed, taking a test is usually not a calm activity from the aspect of the examinees. This study suggests that the **test administrators** should show their sympathetic attitude toward using appropriate equipments of testing and reconfirm for the examinees that the air of the testing room is clean and fresh, the temperature is comfortable, and the surrounding of the room is settled down to be silent.

***Test Reliability***

The fourth element that influences test takers’ grade for a great deal is the test reliability, which includes inappropriate time factor and bad written test items in an examination. This can be resolved only when the test givers and the teachers raise their awareness of counting testing time and fully apply their expertise and knowledge in the **assessment** design.

According to Rexrode, Kathryn, Petersen and O’Toole (2008), “…**reliability** has been assumed to be consistent as an attribute of the test.” (p. 262) Based on Weir’s statement, “…one obvious way of demonstrating that a test is measuring ability consistently is to give the same test twice to the same group of learners.” (p. 25) It is not only very significant but also necessary to ensure the **reliability** of a test when the results are doubted by the test takers.

We believe that an assessment’s result can be reexamined and investigated by utilizing the survey questionnaires. An unreliable test can be concluded if a test is doubted by a high percentage of test takers in a whole class or a school. The score of an assessment would be reliable if most of the students regard all the parameters in the design of assessment are conducted in an appropriate and acceptable fashion.

***Practicality***

A practical test does not take too much time or money. It should be easy to conduct and have an efficient procedure of scoring. In addition, it is impractical to design an expensive assessment.

Practicality of an assessment means that the test should not be very expensive but affordable to most of the students. For example, a student may have an opportunity to achieve a high score through a costly assessment conducted by institutions with a commercial atmosphere. However, it would not be possible for those students who do not have sufficient budget to take the same test.

Second, a test has to be within appropriate time constraints. Either five minutes or five hours may not be the proper time length for the examiners and teachers to discover the learning effects. Namely, a too long or a too short period of test time should not take place since it could not provide the effective results of the learners’ learning. It is meaningless to conduct an unpractical test, which would cause waste of time for students and face-losing for teachers.

Third, an assessment has to be relatively easy to administer and it must contain a scoring/evaluation procedure that is time efficient. That is, the feasibility of the administration in an examination is necessary and significant. If an examination is economical, not too short or too long, easy to score, the teacher and students would welcome the tests because they might improve students’ learning in an efficient way.

Also, the assessments would provide a quick access to students’ learning result for the teachers. Based on the outcomes offered to teachers in a very short time through scoring technique by computers or multiple choice questionnaires, the teachers can efficiently arrange their ways of teaching in order to fit the students’ progress of learning.

Above three deliberations from viewpoints of finances, time-consuming, and the scoring method indicate that a practical test might not be easy to achieve without cooperation with school and students. If a teacher can not get access to technical support of school’s computing system and students’ support to take a 30-minute, or one-hour long test, then the practicality of the assessment could not be carried out.

Conclusively, a practical test could occur only when the assessment is held in a good school with the computer-scoring system and patient students who can be required to take a test longer than half an hour.

***Authenticity***

**Authenticity** of an assessment is related to the target task assigned in the curriculum or the course syllabus and the knowledge that we can apply in our real world. In the second millennium, everybody treats time as money. Therefore, a test must fit the learners’ learned contents as well as the knowledge they really need to utilize in the real world.

Bachman and Palmer (1996) have defined authenticity in the language assessment as “...the degree of correspondence of the characteristics of a given language test task to the features of a target language task.” (p. 23) Usually, the target language test should be associated with the real world and have strong relationship with the target purpose of why the language learners need to learn the language.

According to Cohen (1998), “…authenticity is seen to comprise (1) the relationship between test method characteristics and features of a specific language use situation and (2) the degree to which the test method invokes the test taker’s language ability.” (p.19) Consequently, the use of technical terms and topics from Shakespeare’s play may raise the authenticity of a test for specialists in English literature.

Bachman (1991) points out, since test takers many have perceptions that differ from those of

the test developer on the relevance of the test task to their target language use situations, the fit between the test method and language use situation needs to be assessed from different perspectives.

Hence, we think if test givers could previously specify particular tasks, contextualized items or

thematic organizations of a story or episode, then the form of authenticity can be increased by

adjusting the difficulty and clearly assigning tasks for an assessment. Brown (2004) presents five principles for demonstrating what an authentic test is.

The language in the best is as natural as possible. Items are contextualized rather than isolated. Topics are meaningful (relevant, interesting) for the learner. Some thematic organization to items is provided, such as through a story line or episode. Tasks represent, or closely approximate, real-world tasks. (p. 28)

Namely, an assessment should suggest the learners to perform within a specific usual and significant area, through certain types of terms in order to carry out the authenticity in the examinations. For example, the **authenticity** of writing assessment would be improved by means of providing test takers an explicit criterion for grading. Moreover, the examinees should have a basic conception how well they need to achieve in the test, and what direction they can prepare for the test, based on the requirement of the institution that would need the evaluation result of the test taken.

The test takers should be suggested by the test descriptions beforehand. For instance, what specific consequential and professional elements should be included in the writing, so that we can call this test as an authentic test.

***Washback***

The concern of how assessment instruments affect educational practices and beliefs has also been referred to in the literature as the **washback** or the backwash effect (Hughes, 2003 ; Morrow, 1986; Weir, 1998), so we can also call it washback as the test can have an influence on teaching and learning.

In fact, “the study of washback is also concerned with the political use of tests to implement changes in classrooms that are seen as improvements by governments.” (Fulcher & Davidson, 2007, p.74) Hence, without doing research and taking the element of washback into consideration, a research can not be well designed. It is thus argued that considering the procedure of **washback** is significant and unavoidable.

Cohen (1998) addresses that truly considering the effect of the test on teaching calls for looking at:

1. The effect of the test on the instinctual content, 2. The process whereby the content is taught, 3. The rate at which the material is covered, 4. The nature of the sequencing of material, 5. The quantity of material that the students actually learn and the depth at which they learn the given items or objectives. (p. 41)

This lesson of language assessment of this book emphasizes that the **washback** of evaluation is indispensable and noteworthy because it recommends teaching and learning community to discover how sound the teachers instruct and how much the students prepare for the measurement. Brown (2004) interprets **washback** as follows.

…**washback** also implies that students have ready access to you to discuss the feedback and evaluation you have given. While you almost certainly have known teachers with whom you wouldn’t dare argue about a grade, an interactive, cooperative, collaborative classroom nevertheless can promote an atmosphere of dialogue between students and teachers regarding evaluate judgments. (p. 30)

The **washback** would be a confirmation that the teachers do worry about each student’s learning consequence and development. The teachers pay interest to a certain group of students and appraise their learning condition with details. Through **washback** and warming dialogues between the teachers and each learner, the atmosphere of teaching in the classroom can be peaceful and tolerable.

Usually, through a meeting or a portfolio for **washback**, the details and explanations in the interactive dialogues between teachers and the students could be accomplished for both sides, such as how or whether the further course should be conducted and from which characteristic the teachers should reimburse for their previous lessons that might close the eyes to some important parts in the textbooks.

***Significance of Language Assessment***

According to Liying, Todd and Huigin (2004), “Teachers devote a large of their preparation time to creating instruments and observation procedures, marking, recording, and synthesizing results in informal and formal reports in their daily teaching (p. 360).” Through the execution of language measurement, the teachers would be able to assess a way of observing students’ learning effectiveness, marking and recording students learning achievement and failure.

Based on Lin and Chien’s (2009) study, they discover that during the process of language evaluation, students would gain an entry to learn more if their thoughts could be accurately associated with language usages. More notably, they were diagnosed by the trainers with a

professional manner in order to make compensation for their weakness of learning.

**General English tests of English Languag**e Learners of English are often dedicated to getting accreditation and a number of exams are known internationally. For example, governments in Taiwan provided a test title **General English Proficiency Test** (GEPT) which four kinds of English proficiencies, Listening, Reading, Speaking and Writing are tested with two sessions of testing.

Also, Trinity College London ESOL similarly offers Integrated Skills in English (ISE), series of 5 tests that assess Reading, Writing, Speaking and Listening accepted by academic institutions in the United Kingdom. They at the same time offer a test titled Graded Examinations in Spoken English (GESE), series of 12 periods of evaluations that assess Speaking and Listening and ESOL Skills for Life and ESOL for Work exams in the UK only.

Moreover, University of Cambridge ESOL Examinations provides a suite of five examinations, including First Certificate in English (FCE), Certificate in Advanced English (CAE) and Certificate of Proficiency in English (CPE). Besides, IELTS (International English Language Testing System), accepted by academic institutions in the UK, Australia, New Zealand and Canada, and by many in the USA, also belong to one of the most significant English proficiency tests in the world nowadays.

Very importantly, **Test of English as a Foreign Language** (TOEFL), an Educational Testing

Service product, developed and applied primarily for universities and colleges in America, and is now widely accepted in tertiary institutions in Canada, New Zealand, Australia, the UK, and Ireland.

The current test of TOEFL is Internet based, and is known as the TOEFL iBT (Internet Based Test) used as a proxy for English for Academic Purposes.

* **TOEIC (Test of English for International Communication)**, an Educational Testing Service product for Business English
* **TSE** - **Test of Spoken English,** grade transcript usually for job searching and entering an institution of studying
* **TWE** - **Test of Written English,** grade transcript usually for job searching and entering an institution of study

Many countries also have their own exams. ESOL learners in England, Wales and Northern Ireland usually take the national Skills for Life qualifications, which are offered by several exam

boards. EFL learners in China may take the College English Test. In Greece English students may take the **PanHellenic Association of Language School Owners (**PALSO) exams.

***The General Overview***

**How are young learners being assessed?**

* How do you assess the young learners you teach?
* Make a list of all the ways you assess young learners.
* Think about both classroom tests and those created by an external authority.

**Two researchers asked teachers how they assessed their young learners (ReaDickins and Rixon, 1999).**

* What methods do you think teachers used?
* Check your ideas against those found by the researchers.

**Teachers said they used:**

* Grammar and vocabulary tests
* Single sentence exercises
* Gap-filling
* Vocabulary matching
* Restricted dialogues to test speaking
* Listening skills were not mentioned

**Purposes of classroom assessment (McKay 2006)**

* We can assess learners at the start of the year to identify our students’ strengths and weaknesses.
* During the year we can use the results of tests to help us make decisions about what to teach next and what we need to revise.
* As teachers of young learners, we collect information about children to share with parents and, of course, with the children themselves.
* Teachers use assessments to provide evidence of student progress.
* These assessments can be required by local authorities and teachers must base their assessments on a local or national curriculum.
* Classroom assessment can also be summative. Children can be given a mark or a grade at the end of the school year.

***Assessing the four language skills***

We often think about assessing the four language skills – speaking, listening, reading, writing.

**Writing criteria**

* When assessing writing you need to think about the criteria you use. The criteria may be given to you by your school or local authority.
* Do you have clear criteria for how to assess your pupils’ writing?
* Do you share these with your pupils and their parents?
* Do you give writing tasks that let your pupils demonstrate that they have the abilities you are assessing them on? • Does your feedback reflect these criteria?
* Do you give corrective feedback that pupils can really benefit from?

**Reading Assessment**

* Do your pupils have the opportunity to read texts other than those included in the textbook?
* Have you tried assessing the reading of the children you teach? If so, what methods have you used?
* Could the methods suggested here broaden your assessment practice?

**Oral Assessment**

Think about the following points:

* Level of language proficiency
* Level of confidence with the language
* Topics which have been recently covered in class
* How long the task should take • How will the task be marked?
* Will the task be completed in groups or individually?
* How will the results of the assessment be used?
* What scaffolding (support) will the children need to complete the task?

**Assessing Listening**

Think about the following points:

* Level of language proficiency
* Level of confidence with the language
* The source of the listening material
* How long the task should take
* How will the task be marked?
* Will the task be completed in groups or individually?
* How will the results of the assessment be used?
* What scaffolding will the children need to complete the task?
* What type of listening skill is the target of the test?
* Does the test use write or physical response?

**Other ways to assets**

* Observation
* Self-Assessment
* Portfolio
* Project Work

**Exercise**

* 1. How many basic competence skills in language learning?

1. One
2. Two
3. Four
4. Eight
   1. How many assessments that exist in English Assessment?
5. Eight
6. Six
7. Nine
8. Four
   1. The list below are the test that have been acknowledge by the world, except
9. TOEFL
10. TOEIC
11. VIERRA
12. IELTS
    1. How many types of validity that exist in English Assessment?
13. Six
14. Three
15. Four
16. Eight
    1. The word ‘restricted’ is similar with
17. Forbidden
18. Secret
19. Controllable
20. Chamber
    1. Umpan Balik dan Tindak Lanjut

Please match your answers above with answer key of formative test 1 which is located in the end of the module. Measure your topic mastery of learning activity 1 with formula given below:

Level of mastery= (total of right answers: 5) x 100%

Vey good = 90-100%

Good = 80 - 89%

Fair = 70 – 79%

Poor = 0 – 69 %

If level of mastery of the topic is more than 80%, you can continue to learning activity 2 . If level of mastery is less than 80% you need to re-do learning activity 1 especially from you have not understood part.

1. **Kegiatan Belajar 2**
   1. Uraian dan contoh

text

* 1. Latihan

text

* 1. Rangkuman

text

* 1. Tes Formatif

text

* 1. Umpan Balik dan Tindak Lanjut

text

1. **Kegiatan Belajar 3**
   1. Uraian dan contoh

text

* 1. Latihan

text

* 1. Rangkuman

text

* 1. Tes Formatif

text

* 1. Umpan Balik dan Tindak Lanjut

text

1. **Kunci Jawaban**
   1. Tes formatif 1

1. D

2. A

3. C

4. C

5. A

* 1. Tes formatif 2

text

* 1. Tes formatif 3

text

**Daftar Pustaka**

Alderson, J. C. (2001). Language testing and assessment (Part 1). *Language Teaching, 34,* 213-36.

Alderson, J. C. (2002). Language testing and assessment (Part 2). *Language Teaching, 35,* 79-113.

Bachman, I. F. (1991). What does language testing have to offer? *TESOL Quarterly, 25*(4)*,* 671-704.

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice.* New York: Oxford University

Press.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices,* NY: Pearson

Education, Inc.

Cohen, A. D. (1994). Assessing language ability in the classroom, Boston: Heinle & Heinle

Publishers.

Cheng, L. (2005). *Studies in language testing: Changing language teaching through language.*

Cambridge, Cambridge University Press.

Davidson, F., Hudson, T., & Lynch, B. (1985) *Language testing: Operationalization in classroom*

*measurement and L2 research.* In Marianne Celce-Murcia (Ed.). Beyond basics: Issues and

research in TESOL. Rowley, MA: Newbury House.

Fred, D., Fulcher, G. (2007). *Language Testing and Assessment, An advanced resource book.* New

York, Routledge Applied Linguistics.

Gronlund, N. E. (1998). *Assessment of student achievement.* (6th Ed.). Boston: Allyn and Bacon.

Hughes, A. (2003). *Testing for language teachers.* 2nd Edition. Cambridge: Cambridge University

Press.

Lin, G. H. C. & Chien, P. S. C. (2009, Nov. 6) An Exploration into Equitable Language Assessment,

The 2009 AE Conference Committee. Iso University, Kaohsiung, P. 140-156

Liying, C., Todd, R., & Huigin, H. (2004) ESL/EFL instructors’ classroom assessment practices:

Purposes, methods, and procedures, *Language Testing, 21*(3)*,* 360-389.

Morrow, K. E. (1986). *The evaluation of tests of communicative performance.* In M. Portal (Ed.),

Innovations in language testing. Berkshire, England: EFER-Nelson.

Mousavi, S. A. (2002). *An encyclopedic dictionary of language testing.* (3rd Ed.). Taiwan: Tunghua

Book Company.

Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational*

*Measurement: Issues and Practice, 17*(2), 13-16.

Rexrode, K. R., Petersen, S., O’Toole, S. (2008). The ways of coping scale: A reliability

generalization study. *Educational and Psychological Measurement, 68*(2)*,* 262-280.

Weir, C. J. (1998). *Communicative language testing.* Exeter: University of Exeter Press.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Hampshire: Palgrave Macmillan.