Abstract

This paper examines Purposeful Language Assessment: Selecting the Right Alternative Test. This paper reviews literature and gives a scholarly back-ground to the paper by reviewing some contributions made by various researchers and institutions on testing in English language instruction particularly its usage in language testing. It unveils some views that people and institutions have shared globally on the assessment in English language instruction in education through surveys and other observations. The paper discusses the function of assessment, self-assessment, alternative assessment, and assessing young learners. The major conclusion of the study is that language skills are major aspect of the English language teaching, as it helps learners' understanding of speaking, listening, reading and writing. Hence, attention should be given to its teaching and consequently proper evaluation of its various items. The paper then recommends that language teachers should pay more attention to the careful construction of the test items and phrase each item clearly so that students know exactly what they are requested to do.

Keywords: Language, Assessment, Test, Language Instruction, and Young Learners.


Introduction

Language is used in social interactions to accomplish purposeful tasks (e.g. interacting with another individual in a conversation, writing a text, finding information in a chart or a schedule). Performance is assessed by documenting the successful completion of the task or by using a rubric to assess various dimensions of carrying out the task (e.g. listening comprehension and language complexity in responses to questions in oral interview (Van-Duzer, & Berdan, 2019). Language teachers are often faced with the responsibility of deciding how they intend to measure outcomes and consider what role assessment will play in instruction. Assessment is how to identify the learners' needs, document their progress, and determine how the teachers are doing as teachers and planners (Jerrold, 2012). That being said, how to know the teachers are doing it is right, how to know that the assessment tools are used measure what they intend them to. These are the questions that teachers must continually ask to get the best snapshot
of the progress of the learners and the effectiveness of the programmes (Akintunde, 2023).

Traditionally, the most common way to measure achievement and proficiency in language learning has been the test. Even though, alternatives forms of assessment are growing in popularity, most teachers still use this old standby. And while many teachers may be gifted in classroom, even the best may need some help constructing reliable test items. Carmen (2015) discusses the role of progress testing in the classroom and the importance of matching testing to instruction. He views testing as a tool that can help teachers identify students’ strengths and weaknesses and evaluate the effectiveness of their programmes. In recent years, much has been made of alternative forms of assessment. Whether the teachers want to include students' portfolios or web-based testing in curricula, the focus should always be on gathering information that reflects how well the students have learned what the teachers tried to teach them. Assessment is one of the most difficult and important parts of the teachers' job. Ideally, it should be seen as a means to guide the students on their road to learning, to know how they are progressing and to gauge the effectiveness of our own methodology and materials.

**Testing In English Language Instruction**

Testing is a method of measuring a person's ability, knowledge or performance in a given domain and the method must be explicit and structured, like: Multiple-choice questions with prescribed correct answers; a writing prompt with a scoring rubric; an oral interview based on a question script and a checklist of expected responses to be filled by the administrator (Yesdil.com). Teaching sets up the practice games of language learning: the opportunities for learners to listen, think, take, risk, set goals and process feedback from the teacher (coach) and then recycle through the skills that they are trying to master (Akintunde, 2022). During these practice activities, teachers are indeed observing students' performance and making various evaluations of each learner (Akintunde & Abdallahi, 2022). Then, it can be said that testing and assessment are subsets of teaching. Assessment is related to the learner and his/her achievements, while testing is part of assessment and it measures leaner achievement. In most classrooms today, English is taught through communicative textbooks that provide neither accompanying tests nor any guidance for test construction. Teachers are on their own in constructing tests to measure student progress and performance. The result is they write traditional grammar-based items in a discrete point format that does not fit the communicative orientation of the textbook or the underlying teaching principles. In many cases, teachers have been reluctant to administer regular tests. Stevenson and Riew (2018) give the following reasons for this: a) Teachers consider testing too time-consuming, taking away valuable class time, b) They identify testing with mathematics and statistics, c) Testing goes against humanistic approaches to teaching, d) They have gotten little guidance in constructing tests in either pre-service or in-service training, e) Teachers feel that the time and effort they put into writing and correcting tests is not acknowledged with additional pay or personal praise, f) Testing can be as frightening and frustrating to the teacher as it for the students. Also, one of the important first tasks of any test writer is to determine the purpose of the test. Defining the purpose aids in selection of the right type of test. The table below shows the purpose of many of the common test types.
Table 1. Common Test Types

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Main Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placement tests</td>
<td>Place students at appropriate level of instruction within program</td>
</tr>
<tr>
<td>Diagnostic tests</td>
<td>Identify students’ strengths and weaknesses for remediation</td>
</tr>
<tr>
<td>Progress tests or in-course tasks</td>
<td>Provide information about mastery or difficulty with course materials</td>
</tr>
<tr>
<td>Achievement tests</td>
<td>Provide information about students' attainment of course outcomes at end of course or within the program</td>
</tr>
<tr>
<td>Standardized tests</td>
<td>Provide measure of students' proficiency using international benchmarks</td>
</tr>
</tbody>
</table>

A well-planned course should measure the extent to which students have fulfilled course objectives and the progress tests are a central part of the learning process, so, the reasons for testing can be identified:

a) Testing tells teachers what students can or cannot do-in other words, tests show teachers how successful their teaching has been. It provides wash-back for them to adjust and change course content and teaching styles where necessary.

b) Testing tells students how well they are progressing. This may stimulate them to take learning more seriously.

c) By identifying students' strengths and weaknesses, testing can help identify areas for remedial work.

d) Testing will help evaluate the effectiveness of the programme, course books, materials and methods.

Assessment in English Language Instruction

The term assessment usually evokes images of an end-of-course paper and pencil test designed to tell both teachers and students how much material the students don't know or hasn't yet mastered. However, assessment is much more than tests. Assessment includes a broad range of activities and tasks that teachers use to evaluate student progress and growth on a daily basis. Assessment is how to identify the learners' needs, document their progress, and determine how the teachers are doing as teachers and planners (Jerrold, 2012). Language tests are simply instruments or procedures for gathering particular kinds of information, typically, information having to do with students' language abilities. Tests may have variety of formats, lengths, item types, scoring criteria, and media. While language assessment is the process of using language tests to accomplish particular jobs in language classrooms and programmes.

In language assessment, the teacher will first gather information in a systemic way with the help of language testing tools. For example, the teachers may use an oral interview to gather information about students' speaking abilities, and then make interpretations based on that information or make interpretations about students' abilities to perform a range of real-world speaking tasks based on how well students perform in the oral interview. Based on these interpretations, make a decision or take action within the classroom or programme. The teachers may decide that students need more work on oral fluency and should therefore devote more class time to fluency-oriented activities. Language assessment is much more than simply giving a language test; it is the entire process of test use. Indeed, the ultimate goal of language assessment is to use tests to
better inform us on the decisions we make and the actions we take in language education (Norris, 2010). Assessment refers to a variety of ways of collecting information on learner's language ability or achievement. Although, testing and assessment are often used interchangeably, it is an umbrella term for all types of measures used to evaluate student progress. A test is a formal, systematic (usually paper-and-pencil) procedure used to gather information about student's behavior.

**The Function of Assessment**

There are two kinds of the function of an assessment (Carmen, 2015), Formative Assessment and Summative Assessment. Formative assessment is to evaluate students in the process of ‘forming’ their competencies and skills with the goal of helping them to continue that growth process. It provides the ongoing development of the learner's language, for example: when the teachers give a student a comment or a suggestion, or call attention to an error, that feedback is offered to improve the learner's language ability. And virtually, all kinds of informal assessment are formative. While summative assessment aims to measure or summarize what a student has grasped and typically occurs at the end of a course. It does not necessarily point the way to future progress, for example: final exams in a course and general proficiency exams and all tests/formal assessment (quizzes, periodic review tests, midterm exams, and etcetera) are summative.

**Self-Assessment**

Until the 1980s, references to self-assessment were rare but since then interest in self-assessment has increased. This increase can at least in part be attributed to an increased interest in involving the learner in all phases of the learning process and in encouraging learner autonomy and decision making in (and outside) the language classroom (e.g., Carmen, 2015). The introduction of self-assessment was viewed as promising by many, especially in formative assessment contexts (Oscarson, 2019). It was considered to encourage increasing sophistication in learner awareness, helping learners to: gain confidence in their own judgement; acquire a view of evaluation that covers the whole learning process; and see errors as something helpful. It was also seen to be potentially useful to teachers, providing information on learning styles, on areas needing remediation and feedback on teaching (Carmen, 2015). However, self-assessment also meets with considerable scepticism, largely due to concerns about the ability of learners to provide accurate judgements of their achievement and proficiency. For instance, Hancock, (2013), while acknowledging that self-assessment is an important element in self-directed learning and that learners can play an active role in the assessment of their own language learning, argues that learners cannot self-assess unaided. Taking self-assessment data gathered from students on a pre-sessional EAP programme, he reports a poor correlation between teachers' assessments of the students and their own self-assessments. He also shows that in multicultural groups such as those typical of pre-sessional EAP courses, overestimates of language proficiency are more common than underestimates. He argues that learners' lack of familiarity with metalanguage and with the practice of discussing language proficiency in terms of its composite skills impairs their capacity for identifying their precise language learning needs.

Such concerns, however, did not dampen enthusiasm for investigations in this area and research in the 1980s. It was concerned with the development of self-assessment instruments and their validation (e.g., Oscarson, 2019; Carmen, 2015). Consequently,
a variety of approaches were developed including pupil progress cards, learning diaries, log books, rating scales and questionnaires (Oscarson, 2019). In the last decade the research focus has shifted towards enhancing our understanding of the evaluation techniques that were already in existence through continued validation exercises and by applying self-assessment in new contexts or in new ways. For instance, Carmen (2015) uses standardised achievement and oral proficiency tests both for testing and for self-assessment purposes. He argues that this approach helps to circumvent the problems of training that are associated with self-assessment questionnaires. Carmen (2015) documents the use of a 'do-it-yourself' instrument for placement purposes, reporting that it results in much the same placement levels as suggested by a traditional multiple-choice test. Shohamy (2011) argues that placement testing for large numbers in her context has resulted in the implementation of a traditional multiple-choice grammar-based placement test and a consequent emphasis on teaching analytic grammar skills. She believes that the 'do-it-yourself-placement' instrument might help to redress the emphasis on grammar and stem the neglect of reading and writing skills in the classroom. Hancock, (2013) discusses how self-assessment can become part of the learning process. He describes his use of questionnaires to encourage learners to reflect on their learning objectives and preferred modes of learning. He also presents an approach to monitoring learning that involves the learners in devising their own criteria; an approach that he argues helps learners to become more aware of their own cognitive processes.

A typical approach to validating self-assessment instruments has been to obtain concurrent validity statistics by correlating the self-assessment measure with one or more external measures of student performance (e.g., Shohamy, 2011; Wiggins, 2010). Other approaches have included the use of Multi-Trait Multi-Method (MTMM) designs and factor analysis (Hancock, 2013) and a split-ballot technique (Genesee & John, 2016). In general, these studies have found self-assessment to be a robust method for gathering information about learner proficiency and that the risk of cheating is low (Carmen, 2015). However, they also indicate that some approaches to gathering self-assessment data are more effective than others. Hancock, (2013) reports that learners were able to identify what they found difficult to do in a language than what they found easy to do. Therefore, 'Can-do' questions were the least effective question type of the three they used in their MTMM study, while the most effective question type appeared to be that which asked about the learners' perceived difficulties with aspects of the language.

Additionally, learner experience of the self-assessment procedure and/or the language skill being assessed has been found to affect self-assessments. Genesee and John (2016) in a study of the role of response effects, reports both an acquiescence effect (the tendency to respond positively to an item regardless of its content) and a tendency to overestimate ability, these tendencies being more marked among less experienced learners. Wiggins (2010) has found that the reliability of learners' self-assessments is affected by their experience of the skill being assessed. He suggests that when learners do not have memory of a criterion, they resort to recollections of their general proficiency in order to make their judgement. This process is more likely to be affected by the method of the self-assessment instrument and by factors such as self-flattery. He argues, therefore, for the design of instruments that are cast in terms which offer learners a reference point such as specific curricular content. In a similar finding Van-Duzer, and Berdan, (2019) report that respondents' self-assessments of their oral proficiency in Fijian Hindi are less reliable at the highest levels of the self-assessment scale like Wiggins (2010) they attribute this slip in accuracy to the respondents' lack of familiarity with the criterion measure. Oscarson (2019) sums up progress in this area.
by reminding us that research in self-assessment is still relatively new. He acknowledges that conundrums remain. For instance, learner goals and interpretations need to be reconciled with external imperatives. Also, self-assessment is not self-explanatory; it must be introduced slowly and learners need to be guided and supported in their use of the instruments. Furthermore, particularly when using self-assessment in multicultural groups, it is important to consider the cultural influences on self-assessment. Nevertheless, he considers the research so far to be promising. Despite residual concerns about the accuracy of self-assessment, the majority of studies report favourable results and we have already learned a great deal about the appropriate methodology to use for capturing self-assessments. However, as Oscarson (2019) points out, more work is needed, both in the study of factors that influence self-assessment ratings in various contexts and in the selection and design of materials and methods for self-assessment.

Alternative Assessment

Self-assessment is one example of what is increasingly called 'alternative assessment'. 'Alternative assessment' is usually taken to mean assessment procedures which are less formal than traditional testing, which are gathered over a period of time rather than being taken at one point in time, which are usually formative rather than summative in function, are often low-stakes in terms of consequences, and are claimed to have beneficial wash-back effects. Although, such procedures may be time-consuming and not very easy to administer and score, their claimed advantages are that they provide easily understood information, they are more integrative than traditional tests and they are more easily integrated into the classroom. Norris (2010) makes the point that alternative assessment procedures are often developed in an attempt to make testing and assessment more responsive and accountable to individual learners, to promote learning and to enhance access and equity in education.

On language testing, Norris (2010) reports on a symposium to discuss challenges to the current mainstream in language testing research, covering issues like assessment as social practice, democratic assessment, the use of outcomes based assessment and processes of classroom assessment. Such discussions of alternative perspectives are closely linked to so-called critical perspectives (what Shohamy (2011) calls critical language testing). The alternative assessment movement, if it may be termed such, probably began in writing assessment, where the limitations of a one-off impromptu single writing task are apparent. Students are usually given only on, or at most two tasks, yet generalisations about writing ability across a range of genres are often made. Moreover, it is evidently the case that most writing, certainly for academic purposes but also in business settings, takes place over time, involves much planning, editing, revising and redrafting, and usually involves the integration of input from a variety of (usually written) sources. This is in clear contrast with the traditional essay which usually has a short prompt, gives students minimal input, minimal time for planning and virtually no opportunity to redraft or revise what they have produced under often stressful, time-bound circumstances. In such situations, the advocacy of portfolios of pieces of writing became a commonplace, and a whole portfolio assessment movement has developed, especially in the USA for first language writing (Norris, 2010) but also increasingly for ESL writing assessment (Phillips, 2016) and also for the assessment of foreign languages (French, Spanish, German, etc.) writing assessment.

Although, portfolio assessment in other subject areas (art, graphic design, architecture, music) is not new, in foreign language education, portfolios have been hailed as a major
innovation, supposedly overcoming the drawbacks of traditional assessment. A typical example is Oscarson (2019) who describe the design and implementation of portfolio assessment in Japanese, Chinese, Korean and Russian, to assess growth in foreign language proficiency. He makes a number of practical recommendations to assist teachers wishing to use portfolios in progress assessment. Norris (2010) describes how portfolio assessment was integrated with criterion-referenced grading in pre-university English for academic purposes programme, together with the use of contract grading and collaborative revision of grading criteria. It is claimed that such an assessment scheme encourages learner control whilst maintaining standards of performance. Carmen (2015) discusses the need for better assessment models for instruction where content and language instruction are integrated. He describes examples of the implementation of a number of alternative assessment measures, such as checklists, portfolios, interviews and performance-tasks, in elementary and secondary school integrated content and language classes. Norris (2010) describes a number of alternative procedures for assessing reading, including checklists, teacher-pupil conferences, learner diaries and journals, informal reading inventories, classroom reading aloud sessions, portfolios of books read, self-assessments of progress in reading, and the like.

Many of the accounts of alternative assessment are for classroom-based assessment, often, for assessing progress through a programme of instruction. Jerrold (2012) gives an account of the use of process assessment in an ESP course; Oscarson (2019) describes the use of continuous assessment over a full school year in Spain, to measure achievement of objectives and learner progress. Norris (2010) describes ways he has successfully used a video camera and task-based activities to make classroom-based oral testing more communicative and realistic, less time-consuming for the teacher, and more enjoyable and less stressful for students. Wiggins (2010) describes an experimental system of peer evaluation using questionnaires in a pre-sessional EAP summer programme, to assess speaking abilities. He concludes that this form of evaluation had a marked effect on the extent to which speakers took their audience into account. Norris (2010) discusses how assessment can be integrated with the learning process, illustrating his argument with an example where pupils prepare, practise and perform a set task in Spanish together. He offers practical tips for how teachers can reduce the amount of paperwork involved in classroom assessment of this sort. Phillips (2016) discusses the difficulties of monitoring learning with large groups of students (in contrast with that of individuals) and describes the use, with 200 learners of Dutch, of a simple monitoring tool (a personal computer) to keep track of the performance of individual learners on a variety of learning tasks.

Typical of these accounts, however, is the fact that they are descriptive and persuasive, rather than research-based, or empirical studies of the advantages and disadvantages of 'alternative assessment'. Oscarson (2019) presents a critical overview of such approaches, criticising the evangelical way in which advocates assert the value and indeed validity of their procedures without any evidence to support their assertions. They point out that there is no such thing as automatic validity, a claim all too often made by the advocates of alternative assessment. Instead of 'alternative assessment', they propose the term 'alternatives in assessment', pointing out that there are many different testing methods available for assessing student learning and achievement. They present a description of these methods, including selected response techniques, constructed-response techniques and personal-response techniques. Portfolio and other forms of 'alternative assessment' are classified under the latter category, but Wiggins (2010) emphasises that they should be subject to the same criteria of reliability, validity and practicality as any other assessment procedure, and
should be critically evaluated for their 'fitness for purpose', what Hancock, (2013) called 'usefulness'. Jerrold, (2012) concludes that portfolio scoring is less reliable than traditional writing rating; little training is given and raters may be judging the writer as much as the writing. Stevenson and Riew (2018) emphasise that decisions for use of any assessment procedure should be informed by considerations of consequences (wash-back), the significance and need for, and value of, feedback based on the assessment results, and the importance of using multiple sources of information when making decisions based on assessment information. Jerrold, (2012) makes the point that many alternative assessment procedures are not pre-tested and trialled, their tasks and mark schemes are therefore of unknown or even dubious quality, and despite face validity, they may not tell the user very much at all about learners' abilities. In Carmen (2015) admits, alternative assessment procedures have yet to 'come of age', not only in terms of demonstrating beyond doubt their usefulness, in Hancock, (2013) terms, but also in terms of being implemented in mainstream assessment, rather than in informal class-based assessment. He argues that consistency in the application of alternative assessment is still a problem, that mechanisms for thorough self-criticism and evaluation of alternative assessment procedures are lacking, that some degree of standardisation of such procedures will be needed if they are to be used for high-stakes assessment, and that the financial and logistic viability of such procedures remains to be demonstrated (Stevenson & Riew, 2018).

Assessing Young Learners

Young learners are typically considered to apply to the assessment of children between the ages of 5 and 12 (but also including much younger and slightly older children), the assessment of young learners’ dates back to the 1960s. However, research interest in this area is relatively new. This trend can be largely attributed to three factors. Firstly, second language teaching (particularly English) to children in the pre-primary and primary age groups both within mainstream education and by commercial organisations, has mushroomed. Secondly, it is recognised that classrooms have become increasingly multi-cultural and, particularly in the context of Australia, Canada, the United States and the UK, many learners are speakers of English as an additional/second language (rather than heritage speakers of English). Thirdly, the decade has seen an increased proliferation, within mainstream education, of teaching and learning standards (e.g., the National Curriculum Guidelines in England and Wales) and demands for accountability to stakeholders. The research that has resulted falls broadly into three areas: the assessment of language delay and/or impairment, the assessment of young learners with English as an additional/second language, and the assessment of foreign languages in primary/elementary school (Stevenson & Riew, 2018).

Changes in the measurement of language delay and/or impairment have been attributed to theoretical and practical advances in speech and language therapy. It is claimed that these advances have, in turn, wrought changes in the scope of what is involved in language assessment and in the methods by which it takes place (Carmen, 2015). Resulting research has included reflection on the predictive validity of tests involving language production that are used as standard screening for language delay in children as young as 18 months (particularly in the light of research evidence that production and comprehension are not functionally discrete before 28 months) (Phillips, 2016). Other research, however, has looked at the nature of the language disorder. Oscarson (2019) investigates the effect of semantic inconsistency on sentence grammaticality judgements for children with and without language-learning disabilities.
(LD), finding that children with LD differed most from their chronological age-group peers in the identification of ungrammatical sentences and that it is important to consider the effect on performance of competing linguistic information in the task. Jerrold (2012) develops a phonological assessment procedure for bilingual children, using this assessment to describe the phonological development, in each language, of normally developing bilingual children as well as of two bilingual children with speech disorders. He concludes that the normal phonological development of bilingual children differs from monolingual development in each of the languages and that the phonological output of bilingual children with speech disorders reflects a single underlying deficit. The findings of these studies have implications for the design of assessment tools as well as for the need to identify appropriate norms against which to measure performance on the assessments. Such issues, particularly the identification of appropriate norms of performance, are also important in studies of young learners' readiness to access mainstream education in a language other than their heritage language. Recent research involving learners of English as an additional or second language (EAL/ESL) has benefited from work in the 1980s (Carmen, 2015; Jerrold, 2012) which problematised the use of standardised tests that had been normed on monolingual learners of English. The equity considerations they raised, particularly the **false positive** diagnosis of EAL/ESL learners as having learning disabilities, has resulted in the development of EAL/ESL learner 'profiles' (also called standards/benchmarks/scales) (Oscarson, 2019). Research has also focused on the provision of guidance for teachers when monitoring and reporting on learner progress (Genesee & John, 2016; Phillips, 2016). Curriculum-based age-level tasks have also been developed to help teachers observe performance and place learners on a common framework/standard (Jerrold, 2012).

However, these directions, though productive, have not been unproblematic, not least because they imply (and indeed encourage) differential assessment for EAL/ESL learners in order for individual students' needs to be identified and addressed. This can result in tension between the concerns of the educational system for ease of administration, appearances of equity and accountability and those of teachers for support in teaching and learning (Jerrold, 2012). Indeed, Australia and England and Wales have now introduced standardised testing for all learners regardless of language background. The latter two countries are purportedly following a policy of entitlement for all but, as Carmen (2015) argues; their motives are far more likely to be to simplify/rationalise reporting in order to make comparisons across schools and on which to predicate funding. Furthermore, and somewhat paradoxically, as Shohamy (2011) have established, the use of standardised attainment targets does not result in more equitable treatment of learners, because teachers implicitly apply native-speaker norms in making judgements of EAL/ESL learner performances. Latterly, research has focused on classroom-based teacher assessment, looking, in the case of Wiggins, (2010), at the constructs underlying formative and summative assessment and, in the case of Carmen (2015), at the epistemic and practical challenges for alternative assessment. The overriding conclusion of both studies is that 'insufficient research has been done to establish what, if any, elements of assessment for learning and assessment as measurement are compatible' (Van-Duzer, & Berdan, 2019), a concern no doubt shared by researchers studying the introduction of assessment of foreign languages in primary/elementary schools.

Indeed, the growing tendency to introduce a foreign language at the primary school level has resulted in a parallel growth in interest in how this early learning might be assessed. This research focuses on both formative (e.g., Stevenson & Riew, 2018) and summative assessment (Phillips, 2016) and is primarily concerned with how young
learners' foreign language skills might be assessed, with an emphasis on identifying what learners can do. Motivated in many cases by a need to evaluate the effectiveness of language programmes (e.g., Jerrold, 2012; Van-Duzer, & Berdan, 2019), these studies document the challenges of designing tests for young learners. In doing so they cite, among other factors: the learners' need for fantasy and fun, the potentially detrimental effect of perceived 'failure' on future language learning, the need to design tasks that are developmentally appropriate and comparable for children of different language abilities who have studied in different schools/language programmes and the potential problem inherent in tasks which encourage children to interact with an unfamiliar adult in the test situation (Phillips, 2016). The studies also reflect a desire to understand how teachers implement assessment (Shohamy, 2011) as well as a need for inducting teachers into assessment practices in contexts where there is no tradition of assessment (Jerrold, 2012). Recent years have also seen a phenomenal increase in the number of commercial language classes for young learners with a consequent market for certification of progress. The latest additions to the certificates available are the Saxoncourt Tests for Young Learners of English (STYLE) (http://www.saxon-court.com/publishing.htm) and a suite of tests for young learners developed by the University of Cambridge Local Examinations Syndicate (UCLES): Starters, Movers and Flyers (http://www.Cambridge-efl.org/exam/young/bg_yle.htm)

In the development of the latter, the cognitive development of young learners has purportedly been taken into account and though certificates are issued, these are intended to reward young learners for what they can do. By adopting this approach, it is hoped that the tests will be used to find out what the learners already know/have learned and to check if teaching objectives have been achieved (Phillips, 2016). It is clear that, despite an avowed preference for teacher-based formative assessment, recent research on assessing young learners documents a growth in formal assessment and on-going research exemplifies the movement towards greater standardisation of assessment activities and measures of attainment.

Conclusion

The teaching and learning of English language skills is not complete without appropriate evaluation of all aspects. Therefore, teachers should ensure that all the skills items are appropriately selected and taught separately for better understanding of the basic skills required for each item. This discussion has therefore shown that language skills are major aspect of the English language teaching, as it helps learners' understanding of speaking, listening, reading and writing. Hence, attention should be given to its teaching and consequently proper evaluation of its various items.

Recommendations

The language teacher should pay more attention to the careful construction of the test items. Phrase each item clearly so that students know exactly what they are requested to do. Try to write items that discriminate among good and poor students and are of an appropriate difficulty level. The questions should neither be too easy nor too difficult for the students. Longer tests tend to reduce the influence of chance factors such as guessing. So, essay questions are preferable to multiple choice questions because the latter requires more time to write than the former. Setting longer tests improves reliability only when the additional items are of good quality and as reliable as the original ones. Adding poor quality items induces error and lowers reliability.
References


