

ISSN 2071 - 1964

**Revue interafricaine de littérature,
linguistique et philosophie**

Particip'Action

**Revue semestrielle. Volume 14, N°1 – Janvier 2022
Lomé – Togo**

ADMINISTRATION DE LA REVUE PARTICIP'ACTION

Directeur de publication	: Pr Komla Messan NUBUKPO
Coordinateurs de rédaction	: Pr Martin Dossou GBENOUGA : Pr Kodjo AFAGLA
Secrétariat	: Dr Ebony Kpalambo AGBOH : Dr Komi BAFANA : Dr Kokouvi M. d'ALMEIDA : Dr Isidore K. E. GUELLY

COMITE SCIENTIFIQUE ET DE RELECTURE

Président : Serge GLITHO, Professeur titulaire (Togo)

Membres :

Pr Augustin AÏNAMON (Bénin), Pr Kofi ANYIDOHO (Ghana), Pr Zadi GREKOU (Côte d'Ivoire), Pr Akanni Mamoud IGUE, (Bénin), Pr Mamadou KANDJI (Sénégal), Pr Taofiki KOUMAKPAÏ (Bénin), Pr Guy Ossito MIDIOHOUAN (Bénin), Pr Bernard NGANGA (Congo Brazzaville), Pr Norbert NIKIEMA (Burkina Faso), Pr Adjai Paulin OLOUKPONA-YINNON (Togo), Pr Issa TAKASSI (Togo), Pr Simon Agbéko AMEGBLEAME (Togo), Pr Marie-Laurence NGORAN-POAME (Côte d'Ivoire), Pr Kazaro TASSOU (Togo), Pr Ambroise C. MEDEGAN (Bénin), Pr Médard BADA (Bénin), Pr René Daniel AKENDENGUE (Gabon), Pr Konan AMANI (Côte d'Ivoire), Pr Léonard KOUSSOUHON (Bénin), Pr Sophie TANHOSOU-AKIBODE (Togo).

Relecture/Révision

- Pr Serge GLITHO
- Pr Ataféi PEWISSI
- Pr Komla Messan NUBUKPO

Contact : Revue **Particip'Action**, Faculté des Lettres, Langues et Arts de l'Université de Lomé – Togo.

01BP 4317 Lomé – Togo

Tél. : 00228 90 25 70 00/99 47 14 14

E-mail : participaction1@gmail.com

© Janvier 2022

ISSN 2071 – 1964

Tous droits réservés

LIGNE EDITORIALE DE PARTICIP'ACTION

Particip'Action est une revue scientifique. Les textes que nous acceptons en français, anglais, allemand ou en espagnol sont sélectionnés par le comité scientifique et de lecture en raison de leur originalité, des intérêts qu'ils présentent aux plans africain et international et de leur rigueur scientifique. Les articles que notre revue publie doivent respecter les normes éditoriales suivantes :

1.1 Soumission d'un article

La Revue Particip'Action reçoit les projets de publication par voie électronique. Ceci permet de réduire les coûts d'opération et d'accélérer le processus de réception, de traitement et de mise en ligne de la revue. Les articles doivent être soumis à l'adresse suivante (ou conjointement) : Participation1@gmail.com

1.2 L'originalité des articles

La revue publie des articles qui ne sont pas encore publiés ou diffusés. Le contenu des articles ne doit pas porter atteinte à la vie privée d'une personne physique ou morale. Nous encourageons une démarche éthique et le professionnalisme chez les auteurs.

1.3 Recommandations aux auteurs

L'auteur d'un article est tenu de présenter son texte dans un seul document et en respectant les critères suivants :

Titre de l'article (obligatoire)

Un titre qui indique clairement le sujet de l'article, n'excédant pas 25 mots.

Nom de l'auteur (obligatoire)

Le prénom et le nom de ou des auteurs (es)

Présentation de l'auteur (obligatoire en notes de bas de page)

Une courte présentation en note de bas de page des auteurs (es) ne devant pas dépasser 100 mots par auteur. On doit y retrouver obligatoirement le nom de l'auteur, le nom de l'institution d'origine, le statut professionnel et l'organisation dont il relève, et enfin, les adresses de courrier électronique du ou des auteurs. L'auteur peut aussi énumérer ses principaux champs de recherche et ses principales publications. La revue ne s'engage toutefois pas à diffuser tous ces éléments.

Résumé de l'article (obligatoire)

Un résumé de l'article ne doit pas dépasser 160 mots. Le résumé doit être à la fois en français et en anglais (police Times new roman, taille 12, interligne 1,15).

Mots clés (obligatoire)

Une liste de cinq mots clés maximum décrivant l'objet de l'article.

Corpus de l'article

-La structure d'un article, doit être conforme aux règles de rédaction scientifique, selon que l'article est une contribution théorique ou résulte d'une recherche de terrain.

-La structure d'un article scientifique en lettres et sciences humaines se présente comme suit: **- Pour un article qui est une contribution théorique et fondamentale :**

Introduction (justification du sujet, problématique, hypothèses/objectifs scientifiques, approche), Développement articulé, Conclusion, Bibliographie.

- Pour un article qui résulte d'une recherche de terrain :

Titre,

Prénom et Nom de l'auteur,

Institution d'attache, adresse électronique (note de bas de page),

Résumé en français. Mots-clés, Abstract, Keywords,

Introduction, Méthodologie, Résultats et Discussion, Conclusion, Bibliographie.

Par exemple : Les articles conformes aux normes de présentation, doivent contenir les rubriques suivantes : introduction, problématique de l'étude, méthodologie adoptée, résultats de la recherche, perspectives pour recherche, conclusions, références bibliographiques.

Tout l'article ne doit dépasser 17 pages,

Police Times new roman, taille 12 et interligne 1,5 (maximum 30 000 mots). La revue Particip'Action permet l'usage de notes de bas de page pour ajouter des précisions au texte. Mais afin de ne pas alourdir la lecture et d'aller à l'essentiel, il est recommandé de **faire le moins possible usage des notes (10 notes de bas de page au maximum par article).**

- A l'exception de l'introduction, de la conclusion, de la bibliographie, les articulations d'un article doivent être titrées, et numérotées par des chiffres (**exemples : 1. ; 1.1.; 1.2; 2. ; 2.2. ; 2.2.1 ; 2.2.2. ; 3. ; etc.**).

Les passages cités sont présentés en romain et entre guillemets. Lorsque la phrase citant et la citation dépassent trois lignes, il faut aller à la ligne, pour présenter la citation (interligne 1) en romain et en retrait, en diminuant la taille de police d'un point. Insérer la pagination et ne pas insérer d'information autre que le numéro de page dans l'en-tête et éviter les pieds de page.

Les figures et les tableaux doivent être intégrés au texte et présentés avec des marges d'au moins six centimètres à droite et à gauche. Les caractères dans ces figures et tableaux doivent aussi être en Times 12. Figures et tableaux doivent avoir chacun(e) un titre.

Les citations dans le corps du texte doivent être indiquées par un retrait avec tabulation 1 cm et le texte mis en taille 11.

Les références de citations sont intégrées au texte citant, selon les cas, de la façon suivante :

- (Initiale (s) du Prénom ou des Prénoms de l'auteur. Nom de l'Auteur, année de publication, pages citées) ; - Initiale (s) du Prénom ou des Prénoms de l'auteur. Nom de l'Auteur (année de publication, pages citées). Exemples :

- En effet, le but poursuivi par **M. Ascher (1998, p. 223)**, est « d'élargir l'histoire des mathématiques de telle sorte qu'elle acquière une perspective multiculturelle et globale (...), d'accroître le domaine des mathématiques : alors qu'elle s'est pour l'essentiel occupée du groupe professionnel occidental que l'on appelle les mathématiciens (...) ».

- Pour dire plus amplement ce qu'est cette capacité de la société civile, qui dans son déploiement effectif, atteste qu'elle peut porter le développement et l'histoire, S. B. Diagne (1991, p. 2) écrit :

Qu'on ne s'y trompe pas : de toute manière, les populations ont toujours su opposer à la philosophie de l'encadrement et à son volontarisme leurs propres stratégies de contournements. Celles-là, par exemple, sont lisibles dans le dynamisme, ou à tout le moins, dans la créativité dont sait preuve ce que l'on désigne sous le nom de secteur informel et à qui il faudra donner l'appellation positive d'économie populaire.

- Le philosophe ivoirien a raison, dans une certaine mesure, de lire, dans ce choc déstabilisateur, le processus du sous-développement. Ainsi qu'il le dit :

le processus du sous-développement résultant de ce choc est vécu concrètement par les populations concernées comme une crise globale : crise socio-économique (exploitation brutale, chômage permanent, exode accéléré et douloureux), mais aussi crise socio-culturelle et de civilisation traduisant une impréparation sociohistorique et une inadaptation des cultures et des comportements humains aux formes de vie imposées par les technologies étrangères. (S. Diakitè, 1985, p. 105).

Pour les articles de deux ou trois auteurs, noter les initiales des prénoms, les noms et suivis de l'année (J. Batee et D. Maate, 2004 ou K. Moote, A. Pooul et E. Polim, 2000). Pour les articles ou ouvrages collectifs de plus de trois auteurs noter les initiales des prénoms, le nom du premier auteur et la mention "et al" (F. Loom et al, 2003). Lorsque plusieurs références sont utilisées pour la même information, celles-ci doivent être mises en ordre chronologique (R. Gool, 1998 et M. Goti, 2006).

Les sources historiques, les références d'informations orales et les notes explicatives sont numérotées en série continue et présentées en bas de page.

Références bibliographiques (obligatoire)

Les divers éléments d'une référence bibliographique sont présentés comme suit : NOM et Prénom (s) de l'auteur, Année de publication, Zone titre, Lieu de publication, Zone Editeur, pages (p.) occupées par l'article dans la revue ou l'ouvrage collectif.

Dans la zone titre, le titre d'un article est présenté en romain et entre guillemets, celui d'un ouvrage, d'un mémoire ou d'une thèse, d'un rapport, d'une revue ou d'un journal est présenté en italique. Dans la zone Editeur, on indique la Maison d'édition (pour un ouvrage), le Nom et le numéro/volume de la revue (pour un article). Au cas où un ouvrage est une traduction et/ou une réédition, il faut préciser après le titre le nom du traducteur et/ou l'édition (ex : 2nde éd.).

Ne sont présentées dans les références bibliographiques que les références des documents cités. Les références bibliographiques sont présentées par ordre alphabétique des noms d'auteur. Il convient de prêter une attention particulière à la qualité de l'expression. Le Comité scientifique de la revue se réserve le droit de réviser les textes, de demander des modifications (mineures ou majeures) ou de rejeter l'article de manière définitive ou provisoire (si des corrections majeures doivent préalablement y être apportées). L'auteur est consulté préalablement à la diffusion de son article lorsque le Comité scientifique apporte des modifications. Si les corrections ne sont pas prises en compte par l'auteur, la direction de la revue Particip'Action se donne le droit de ne pas publier l'article.

AMIN Samir, 1996, Les défis de la mondialisation, Paris, Le Harmattan.

AUDARD Cathérine, 2009, Qu'est-ce que le libéralisme ? Ethique, politique, société, Paris, Gallimard.

BERGER Gaston, 1967, L'homme moderne et son éducation, Paris, PUF.

DIAGNE Souleymane Bachir, 2003, « Islam et philosophie. Leçons d'une rencontre », Diogène, 202, p. 145-151.

DIAKITE Sidiki, 1985, Violence technologique et développement. La question africaine du développement, Paris, L'Harmattan.

NB1 : Chaque auteur dont l'article est retenu pour publication dans la revue **Particip'Action** participe aux frais d'édition à raison de 65.000 francs CFA (soit 100 euros ou 130 dollars US) par article et par numéro. Il reçoit, à titre gratuit, un tiré-à-part.

NB2 : La quête philosophique centrale de la revue **Particip'Action** reste: **Fluidité identitaire et construction du changement: approches pluri-et/ou transdisciplinaires.**

Les auteurs qui souhaitent se faire publier dans nos colonnes sont priés d'avoir cette philosophie comme fil directeur de leur réflexion.

La Rédaction

SOMMAIRE

LITTERATURE

1. Weaponizing the Voice in a Hostile Environment: A Reading of August Wilson's Ma Rainey's *Black Bottom*
Kodzo Kuma AHONDO.....9
2. Spaces in Nadine Gordimer's *The Pickup*: Sites of Identity Redefinition
Khadidiatou DIALLO.....27
3. Non Antagonistic Views about What It Means to be Sisters: A Reading of Danzy Senna's *From Caucasia with Love*
Alexandre NUBUKPO.....53
4. L'art et la vie dans *Flaubert's Parrot* (1984) de Julian Barnes
Astou FALL & Salif MENDY.....61
5. The African Woman in John Ruganda's *Black Mamba*
Kokouvi Mawulé d'ALMEIDA & Ayélé Fafavi d'ALMEIDA.....81
6. "The Streets Don't Go There": The Construction of the Myth of the Male and Female Subjection in Toni Morrison's *Love*
Sènakpon Adelphe Fortuné AZON.....97
7. Contextualizing Education in African and American Communities: A Marxist Reading of Joseph Coleman De Graft's *Sons and Daughters* and William Henry Smith's *The Drunkard*
James Kodjo AKAKPO-DOME.....115
8. Revisiting Racial Lexis in Peter Abrahams' *Tell Freedom*
Fougnigué Madou YEO.....139

LINGUISTIQUE

9. Etude lexicosémantique de termes liés au plurilinguisme-pluriculturalisme et leurs incidences sur les langues et les anthroponymes et patronymes éwé
Komla Enyuamede AGBESSIME.....157
10. Prolégomènes à l'élaboration du manuel scolaire pour l'apprentissage du français langue étrangère au Nigéria
Boniface Osikwemhe IGBENEGHU.....179
11. Some Strategies to Improve Reliability in Spoken Production Tests
Amelan Martine AKPESSI Epse YAO.....207

12. Les études littéraires et la professionnalisation à l'Université d'Abomey-Calavi
Martial FOLLY & Toussaint Yaovi Tchitchi.....231

PHILOSOPHIE ET SCIENCES SOCIALES

13. A Critical Descriptive, Interpretative and Explanatory Analysis of Vice President Kamala Harris's Victory Speech
Ferdinand KPOHOUE, Nassourou IMOROU & Edouard L. K. Koba.....255
14. Mécanismes d'accès aux terres cultivables dans la préfecture du Zio au Togo
Komivi BOKO.....277

SOME STRATEGIES TO IMPROVE RELIABILITY IN SPOKEN PRODUCTION TESTS

Amelan Martine AKPESSI Epse YAO*

Abstract

This study intends to investigate current oral proficiency tests in the Department of English of Peleforo Gon Coulibaly University in Korhogo, Côte d'Ivoire and suggest ways of improving reliability. Results revealed an inadequate preparation, a promotion of group assessments, partiality in scoring and a neglect of the analytic scoring. My suggestions for improve reliability are fourfold: allow students equal and reasonable time for both training and presentation, encourage individual assessment, multiply the number of testers and privilege the analytic scoring.

Keywords: Oral proficiency test- reliability- raters- analytic scoring- holistic scoring.

Résumé

Cette étude a pour objectif d'enquêter sur les pratiques actuelles en matière d'évaluation de la production orale des apprenants au Département d'anglais de l'Université Peleforo Gon Coulibaly à Korhogo en Côte d'Ivoire et de faire des suggestions pour plus de fiabilité. Les résultats ont révélé une insuffisance de préparation au test, une promotion de l'évaluation de groupe, la partialité dans la notation et une négligence de la notation de type analytique. Mes suggestions pour plus de fiabilité sont de quatre ordres : accorder aux étudiants un temps égal et raisonnable pour la formation et la présentation, encourager l'évaluation individuelle, accroître le nombre d'examineurs et privilégier la notation de type analytique.

Mots-clés : Test d'expression orale - fiabilité - évaluateurs- notation de type analytique- notation de type impressionniste.

* Université Peleforo Gon Coulibaly, Korhogo (Côte d'Ivoire) ; E-mail : mak72yao@gmail.com

Introduction

Regardless of its forms, testing students' speaking skills is a challenging task for, unlike a paper and pen test, oral testing is generally a face to face communication. Therefore, if we don't care, familiarity or other factors can affect our wish for fairness. So the issue is how to obtain accurate scores free from biases. This is really important for the results of oral testing could have either a positive or a negative backwash on the process of teaching and learning.

As L. Blanche (1988, p. 75) puts it, the scores obtained help learners to know "how much progress they are making and what they can (or cannot yet) do with the skills they have acquired. Without such knowledge, it would not be easy for them to learn efficiently".

In the same vein, M. Huxham, F. Campbell, and J. Westwood (2012, p. 45) maintain that "the use of oral assessment motivates students to practice and improve their English speaking skills".

Therefore, an effective measurement of learners' abilities implies taking into account some criteria among which reliability. This paper aims at improving the assessment¹⁰ of students' oral production that occurs at the end of oral practice tutorials within the Department of English of Peleforo Gon Coulibaly University in Korhogo, Côte d'Ivoire where I am currently teaching English as a foreign language. It endeavours to answer the following preoccupations: which testing strategies could help to improve reliability during spoken production tests? How could their implementation help to avoid inaccurate and subjective scores?

1. Literature review

In the field of language testing, although the terms test, testing and assessment are related, they differ from each other.

¹⁰ Although the term 'testing' and 'assessment' have different meanings, they may be used synonymously in this paper.

According to L. Bachman (1990, p. 20), a test is “a measurement instrument designed to elicit a specific sample of an individual’s behavior”.

For C. Palomba and T. Banta (1999, p. 4), assessment is “the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development”. When it comes to testing, it is the process of administering a test to measure one or more concepts usually under standardized conditions. So, a test is a means to the end of testing and assessment.

The history of language testing has been dominated by four main trends or approaches including the pre- scientific, the structuralist, the integrative and the communicative approach. The pre- scientific also known as the essay translation approach laid much emphasis on tests characterized by the use of essays, open-ended examinations or oral examining with the result determined intuitively by an authorized examiner. No special skill or expertise in testing was then required. There was little concern about reliability and objectivity of tests. Owing to the fact that this approach did not derive from any single linguistic theory, it was soon rejected. The structuralist approach is also termed the ‘discrete point’ analysis. It consisted in breaking language elements into small units and testing them separately. Although it was to provide information about learners’ ability in specific areas, it was also questioned for it’s difficult to show that any single linguistic unit is indispensable for communication. Unlike the discrete point which focused on the learner’s ability to use single units of the language, the integrative approach innovated by encouraging the use of several units of the language at a time. It involves the testing of language in context. However, it was also challenged for it did not allow to assess the learner’s communicative competence. The communicative approach to language testing relates to learners’ ability to use the language in concrete communication situations. As a matter of fact, knowing the elements of the

language counts for nothing unless the user is unable to take part in meaningful communication situation.

According to M. Canale and M. Swain (1980, p. 30), communicative competence encompasses grammatical competence, socio-linguistic competence, discourse competence and strategic competence. Therefore, testing learners' communicative competence implies taking all those four aspects into account.

Depending on their goals, we have different types of tests including the placement test, the proficiency test, the diagnostic test, the aptitude test and the achievement test (A. Hughes 1989, p. 11). Although they differ from each other, they share common features regarding qualities involving authenticity, interactiveness, practicality, washback, validity and reliability. Unless those criteria are met, a test might not be considered good or efficient. Authenticity is defined as the relationship between test task characteristics, and the characteristics of tasks in the real world. Interactiveness is the degree to which the individual test taker's characteristics (i.e language ability, background knowledge and motivation) are engaged when taking a test. Practicality has to do with the extent to which the demands of a particular test specification can be met within the limits of the existing resources. Washback refers to the effect of testing on teaching and learning. Validity relates to the appropriateness of a given test or any of its component parts as a measure of what it is purposed to measure. A. Hughes (2000, p. 22) believes that "a test is said to be valid if it measures accurately what it is intended to measure". Thus, in the academic setting, a valid test is the one that matches with course objectives and the syllabus. As far as reliability is concerned, it has to do with consistency, generalizability and stability of a test.

R. Brennan (2001, pp. 295-6) defines it as “a measure of the degree of consistency in examinee scores over replications of a measurement procedure”.

A. Hughes (2003) agrees with this definition and even provides some guidelines in order to reach reliability. He says:

What we have to do is construct, administer and score tests in such a way that the scores actually obtained on a test on a particular occasion are likely to be very similar to those which would have been obtained if it had been administered to the same students with the same ability, but at a different time. The more similar the scores would have been, the more reliable the test is said to be (36).

A. Hughes (2003) was so aware of the importance of reliability as a prerequisite in language testing that he declares “if a test is not reliable, it cannot be valid” (34) to prove that validity cannot be achieved unless reliability is already established. He even added that “to be valid a test must provide consistently accurate measurements” (50). There are different types of reliability including rater reliability, inter-rater reliability, intra-rater reliability, test-retest reliability, alternate-forms reliability and score reliability. Rater reliability entails human subjective nature which could affect scores. Inter-rater reliability is established by comparing the scores of different raters. It relates to the extent to which different raters agree in their assessment decisions. Intra-rater reliability expresses how constant is the rating of a rater on different times. Test-retest reliability refers to the ability of a test to achieve the same result time after time with the same group of individuals. Parallel or alternate-forms reliability is the consistency of test takers’ scores across different editions of the test, containing different questions or problems testing the same types of knowledge or skills at the same difficulty level. Score reliability in oral proficiency test relates to scoring and criterial levels.

For a test to be labelled as reliable, A. Hughes (2003) lists the following criteria that should be met regardless of the skill being assessed:

1. Take enough samples of behavior for it has been demonstrated empirically that the addition of further items will make a test more reliable;
2. Exclude items which do not discriminate well between weaker and stronger students;
3. Do not grant candidates too much freedom because the more freedom you give them, the greater the differences between the performances;
4. Write unambiguous items; candidates should not be presented with items whose meanings are not clear;
5. Provide clear and explicit instructions for it's, by no means, always the weakest candidates who are misled by ambiguous instructions; indeed it is often the better candidates who are able to provide the alternative interpretation;
6. Ensure that tests are well laid out and perfectly legible;
7. Make candidates familiar with exam format and testing techniques; for this reason every effort must be made to make sure that all candidates have the opportunity to learn what will be required from them;
8. Provide uniform and non-distracting conditions of administration; every precaution should be taken to maintain a quiet setting with no distracting sounds or movements;
9. Use items that permit scoring which is as objective as possible;
10. Provide detailed scoring keys;
11. Train scorers;
12. Agree on the score to be given for acceptable responses at the outset of scoring;
13. Identify candidates by numbers, not names;
14. Employ multiple, independent scoring keys (pp. 44-48).

Moreover, for Oral Proficiency Interviews (OPI), he itemizes these additional criteria in order to improve reliability:

1. Make the oral test as long as feasible; it is unlikely that much reliable information can be obtained in less than about 15 minutes while 30 minutes can probably provide all the information needed for most purposes;
2. Plan the test carefully;
3. Give the candidate as many 'fresh starts' as possible;
4. Use a second tester for interviews;
5. Set only tasks and topics that would be expected to cause candidate no difficulty in their own language;
6. Carry out the interview in a quiet room with good acoustics;
7. Make candidates comfortable so that they can show what they are capable of ;
8. Collect enough relevant information;
9. Do not talk too much;
10. Select interviewers carefully and train them (A. Hughes 2003, pp. 124-126).

Although the above points are suggested for oral interviews, they might be valid for the other kinds of oral assessments including paired tests, role-plays and group discussions.

When it comes to grading, two ways of scoring are generally used: either the holistic or the analytic scoring. The holistic scoring also referred to as impressionistic or global scale aims to look at the overall performance of the testees.

X. Xi and P. Mollaun (2006) explain that

In holistic scoring, raters consider the combined impact of delivery, language use, and topic development, and make a judgment about a person's performance on a particular task. During this process, raters attempt to weigh the impact of different dimensions on the overall effectiveness of communication to come up with a holistic score (32).

Here, the different components of speaking (fluency, vocabulary, grammar, pronunciation, cohesion, etc.) are considered as a whole. So the rater focuses on the overall oral performance of his testee instead of relying on specific abilities. The holistic rubric appears to be easier and faster than the analytic one since it takes less time for grading. It's usually appreciated by teachers who have to assess large numbers of students in a short time. Conversely, the analytic or discrete rubric

seeks to separate out salient features of performance and to evaluate each one individually and independently on its own subscale; the analytic approach thus focuses attention on discrete qualities of performance, typically combining scores on the separate subscales to produce an overall score for speaking, and sometimes reporting the subscores as well to provide a richer level of source information, which can be useful for diagnostic purposes to guide future teaching/learning objectives (L. Taylor and E. Galaczi, 2011, p. 177).

Yet, M. Llach (2011, p. 57) scorns the analytic scoring for he argues that “[o]ne of the main disadvantages of analytic scoring is the difficulty in providing clear-cut and unambiguous definitions for each descriptor”.

Nevertheless, K. Finson, C. Ormsbee, and M. Jensen (2011, p. 181) maintain that “[a]nalytic rubrics (...) support a more objective and consistent assessment of student work”. Discrete scoring is more detailed and allows learners to enhance their oral proficiency for it provides them with a clear idea of their individual performances; their weaknesses and their strengths as well as the areas where improvements are expected.

L. Tuan (2012, p. 673) favours the analytic scoring too and even suggests “[d]epending on the purpose of the assessment, speaking performance might be rated on such criteria as content, organisation, cohesion, register, vocabulary, grammar, or mechanics”. Undoubtedly, the analytic scoring could help to achieve a higher score reliability.

Interestingly, the analytic scale is the one that is used during IELTS¹¹ or TOFEL¹² exams.

2. Theoretical Framework

Communicative Competence lays the background of this study. M. Canale and M. Swain (1980) define it as:

A synthesis of knowledge of basic grammatical principles, knowledge of how language is used in social contexts to perform communicative functions, and knowledge of how utterances and communicative functions can be combined according to the principles of discourse (20).

As seen above, communicative competence includes four main dimensions: grammatical competence, discourse competence, sociolinguistic competence and strategic competence (M. Canale and M. Swain 1980, p. 30). Grammatical competence for M. Canale and M. Swain (1983, p. 7) involves “features and rules of the language such as vocabulary, word formation, sentence formation, pronunciation, spelling and linguistic semantics”. Discourse competence relates to the ability to combine language structures into different types of coherent and cohesive texts so as to produce meaningful utterances.

Sociolinguistic competence according to L. Bachman 1990, p. 94) is the “sensitivity to, or control of, the conventions of language use that are determined by the features of the specific language use context”. Simply puts, it has to do with the mastery of the sociocultural code of language use. It is the appropriate application of vocabulary, register, politeness and style in a given situation.

Strategic competence is explained by L. Diaz-Rico and K. Weed (1995, p. 15) as “the manipulation of language in order to meet

¹¹ IELTS stands for International English Language Testing System

¹² TOEFL stands for Test of English as a Foreign Language

communication goal”. It focuses on the ease with which speakers flexibly manipulate the language in order to meet communication demands. Communicative competence is at the core of the oral proficiency test.

D. Brown (2007, p. 352) does not contest the idea since according to him pronunciation - fluency -vocabulary - grammar and discourse feature are the skills that are generally assessed during oral proficiency tests. And these skills remind us the components of communicative competence.

“The role of the language teacher is to help learners get along in real-life situations”

2. Context of the study and methodological framework

2.1. Context of the study

The study was carried out at the Department of English of Peleforo Gon Coulibaly University in Korhogo, Côte d’Ivoire. The population consists of first year English as Foreign Language (EFL) students. The language classes include Reading, Writing, Listening and Speaking (Oral Practice). Twelve hours are initially devoted to the teaching of oral practice which comprises both teaching and testing learners’ speaking abilities.

2.2 Research paradigm

The paradigm adopted in this work is action research which, according to A. Burns (2010, p. 2), involves “taking a self-reflective, critical, and systematic approach to exploring your own teaching contexts”. She believes that “the central idea of action research is to intervene in a deliberate way in the problematic situation in order to bring about changes and, even better, improve-ments in practice” (id). Consequently, A. Burns (2019) provided a framework for a successful action research when she declares:

The action component requires some kind of planned intervention, which deliberately puts into place particular strategies, processes, or activities in the research context. Interventions are introduced in response to a perceived issue, puzzle, dilemma, or question that people in the immediate social context wish to understand, improve, change, or mediate in some way in order to create a more positive educational outcome (992).

From this, it's clear that action research entails important steps including the identification of a problem followed by an intervention. With such objectives, action research could rightly serve as a means for improvement of curriculum and teaching. In the present study, I first investigated the assessment of students' oral skills in the department of English; I then identified some weaknesses regarding reliability. Finally, I made some suggestions for increased reliability during speaking tests.

2.3. Sampling and data collection

2.3.1 Sampling

The technique of convenience sampling which involves the selection of individuals on the ground of their availability was adopted. In order to have reliable data, the research covers two academic years: 2018-2019 and 2019-2020. Thus, the first sample consisted of eighty (80) students while the second one included one hundred and twenty two (122) students, making it a total number of two hundred and two (202).

2.3.2 Data collection

Two different perspectives were used to collect the data. Firstly, a questionnaire was designed to collect data from the students (testees). Secondly, two semi-structured interview guides were used for collecting data from some students and their teachers (testers).

2.3.3 Data analysis

The data were analyzed according to two different perspectives. The one collected from the questionnaire went through quantitative analysis (frequency analysis) whereas those collected from the interview were submitted to qualitative analysis (content analysis).

3. Presentation of the results

3.1 The results of the quantitative analysis of the data

The results of the quantitative analysis of the data are presented in four (4) multidimensional tables.

Table 1: Types of classroom activities

Activities Frequency	Grammar		Vocabulary		Discussion		Role play		Interviews	
	Nber	%	Nber	%	Nber	%	Nber	%	Nber	%
1	202	100	202	100	0	0	50	24.75	155	76.73
2	0	0	0	0	0	0	82	40.59	47	23.27
3	0	0	0	0	94	46.53	31	15.35	0	0
4	0	0	0	0	108	53.47	39	19.31	0	0
Total	202	100	202	100	202	100	202	100	202	100

As revealed in table 1, much of classroom activities (100%) are devoted to discussion. Role play accounts for 34.66% of classroom activities while less time is devoted to interview which represents 23.27% of the activities performed during tutorials. Besides, grammar and vocabulary are completely ignored since all the 202 respondents (100%) confess that these kinds of activities were never done.

Table 2: Types of activities during oral testing

Activities Frequency	Grammar		Vocabulary		Discussion		Role play		Interviews	
	Nber	%	Nber	%	Nber	%	Nber	%	Nber	%
1	202	100	202	100	0	0	121	59.90	162	80.20
2	0	0	0	0	0	0	48	23.76	40	19.80
3	0	0	0	0	107	52.97	27	13.37	0	0
4	0	0	0	0	95	47.03	06	2.97	0	0
Total	202	100	202	100	202	100	202	100	202	100

All the 202 (100%) students acknowledge that oral proficiency tests are usually based on discussion at the expense of interview (19.80%) and role play (16.34%). In addition, according to the students, they are never asked to do grammar and vocabulary exercises and this idea is confirmed by all the 202 respondents.

Table 3: Types of assessment

Types Frequency	Individually		In pairs		In groups	
	Nber	%	Nber	%	Nber	%
1	162	80.20	202	100	0	0
2	40	19.80	0	0	0	0
3	0	0	0	0	94	46.53
4	0	0	0	0	108	53.47
Total	202	100	202	100	202	100

All the 202 students (100%) agree that they are generally assessed in groups while individual assessment is rarely done (19.80%). When it comes to pair assessment, all the 202 respondents (100%) confirm that this way of testing learners' spoken production is never used.

Table 4: Grading criteria

Criteria Frequency	Grammar		Vocabulary		Fluency		Pronunciation		All of the above		Others	
	Nber	%	Nber	%	Nber	%	Nber	%	Nber	%	Nber	%
1	123	60.89	49	24.26	0	0	0	0	72	35.64	86	42.57
2	79	39.11	84	41.58	0	0	0	0	0	0	61	30.20
3	0	0	69	34.16	94	46.53	102	50.50	130	64.36	55	27.23
4	0	0	0	0	108	53.47	100	49.50	0	0	0	0
Total	202	100	202	100	202	100	202	100	202	100	202	100

Among the grading criteria, all the students (100%) mention fluency and pronunciation. For 79 students (39.11%), correct use of grammar is also required whereas 123 (60.89%) others don't think so. 84 students (41.58%) think that vocabulary accuracy is also checked while 49 others (24.26%) don't see things this way. More than half of the students (64.36%) are aware of the grading criteria even if they are ignored by 35.64% of the respondents. Still, apart from the above criteria, for 116 students (57.43%), other criteria are also considered.

It appears that the approach used during oral practice courses coincides with the one used for testing learners' spoken production. In both cases, an emphasis is laid on discussion (or communicative activities) at the expense of role play and interview. In addition, grammar and vocabulary activities are completely overlooked in both cases. Moreover, group assessments are favoured over individual and pair assessments. Furthermore, many students are aware of the grading criteria.

3.2 The results of qualitative analysis of data

The lines below present the results of qualitative analysis of the data collected from interviews.

3.2.1. Students' responses

The students interviewed acknowledged that the activities performed during oral practice tutorials prepare them for oral proficiency tests.

Olive¹³ thinks: "the tutorials prepare us for the test to the extent that we are given the kinds of exercises we are likely to meet during the test".

Angeline will not say the contrary; she even adds that "the training received during the course helps us to overcome fear and anxiety. Without that training, I'm sure none of us would be able to speak in front of an audience. It's a way to increase confidence".

Even if classroom activities aimed to prepare students for the test, the duration of that training poses a real problem. According to Daniel

We cannot consider these four hours of tutorials as training. Do you feel what we did is a training? I'm not sure, for only few students are given the opportunity to express themselves. What about the others? Four hours cannot turn us into good speakers, ready to do the test scheduled in the afternoon. It's not enough. Unfortunately this way of doing things tends to belittle the speaking skill. We have the impression that oral skill is something we have to rid ourselves of and move to more important things. It was so quick that we did not have enough time for preparation.

¹³ All participants' names are pseudonyms

Kpanigo goes along with Daniel when he suggests: “it would be better to have two or three days tutorials and a whole day for preparation before the test.”

When it comes to the chance of success offered to students, Sialou frankly admits:

We are not given the same chance. The first groups that were assessed had less time (15minutes) for preparation and 10 minutes for presentation while the last groups had one (1) or two (2) hours preparation. Consequently the last groups did well than the first ones. I wish we were given the same amount of time for preparation. If we could be rated simultaneously by three or four teachers, nobody’s interest would be harmed.

Olive will not oppose the idea. She even goes further by suggesting that:

Giving equal chance presupposes that individual assessment should be favoured. That could help learners to know their individual performance. Even if group assessment is privileged by the teacher, we should be tested according to individual performances.

The students believed that the activities carried out during tutorials could prepare them to oral proficiency tests. Yet, they suggest enough time for training as well as fairness in scoring.

3.2.2 Teachers’ responses

Teachers’ responses corroborate those provided above by the students for they confess that the communicative activities carried out during the tutorials were designed to prepare their students for the test.

As Doctor Agnimel puts it: “testing provides a feedback to the teaching. So, how could we ask testees to accomplish unfamiliar tasks during a test? Honestly, it’s our duty to help them be prepared for the oral proficiency test”. During tutorials, the students are given the kinds of exercises that are likely to be encountered during the test.

Doctor Tiacoh agrees with him but he complains about the time devoted to oral practice when he says:

Our duty is to train our students. But unfortunately, the time devoted to the training is nothing in comparison with the aim of the introduction of the subject discipline within the syllabus. Initially twelve hours are devoted to oral practice but actually due to room problems both the tutorials and the assessment take only 8 hours. Both are scheduled for a single day. Honestly, that is not easy at all for students who have just entered university.

For Doctor Agnimel,

Preparation is necessary. But everything cannot be done during the class. Students should be involved in their own training by listening to radio programs or joining English clubs for improving their language proficiency.

Also, the teachers investigated acknowledge that the principle of fairness was not always respected for several reasons.

Doctor Rodriguez thinks that the main reason at this level is time shortage. He explains:

Could you picture the scene? You have to assess a large number of students alone. Surely, the first groups will not have enough chance as the last ones. In addition, some factors such as tiredness or time pressure might influence negatively students' performances.

Doctor Tiacoh is not of a different view. He even suggests we seek the assistance of workmates (other teachers) by declaring:

To my mind, raters need to be assisted by their colleagues so that instead of rushing, each testee will have reasonable time for both preparation and presentation. So doing will considerably facilitate the rater's work and reduce stress or tiredness. It will help to achieve fairness in scoring. Normally, individual assessment is the type of assessment that provides a clear picture of a student's speaking abilities.

Like their students, the teachers interviewed agree that the communicative activities done during the tutorials helped to prepare the students for the test. Nevertheless, they complain about time shortage for both preparation and testing. Besides, they suggested the assistance of other colleagues in order to meet the principle of fairness. The data collected from both the questionnaire and the interviews show that students' were never asked to

perform unfamiliar tasks during oral proficiency tests. Moreover the students in general are aware of the existence of grading criteria. However, time shortage, the type of assessment (group assessment) and the number of raters remain an impediment to reliability.

3.3 Interpretations and discussion

Both qualitative and quantitative data reveal that the students were never asked to perform unfamiliar tasks during oral testing. The format and the content of the test generally echoed classroom activities (See table 1&2). This way of doing things might help to achieve reliability for it is one of the criteria of test reliability according to A. Hughes (2003, pp. 44-48) mentioned earlier. However, apart from this positive aspect, one can conclude that the assessment of students' speaking abilities failed to meet all the criteria of reliability. The explanation for this situation is fourfold. First, although the activities carried out aimed at preparing the students for the exam, what is actually being done at Peleforo Gon Coulibaly University might not be called preparation as such. For, the time allotted to the tutorials (4hours) could not permit an effective training. Unless the students resort to other opportunities, they would not be able to pass the test equipped with such an inadequate preparation. Unfortunately, this seems quite impossible since the test is scheduled to take place the same day, only three hours after the lecture.

Two, the results from the data also show that due to time pressure, group assessments were privileged to the detriment of pair and individual assessment (See table 3). Sadly enough this situation did not permit the students to know their individual language proficiency so as to make necessary improvement. Group assessment encourages holistic scoring which considers the overall performance rather than specific abilities. Thus, it fails to provide a reliable picture of students' language skills. Therefore,

to enhance reliability, it is advisable for teachers in charge of speaking classes to promote individual assessment. Yet, in case of larger groups of testees, instead of giving the impression that the syllabus is something they want to get rid of, it would be better to seek the assistance of other teachers. Such initiative might help to cope with time pressure, reduce tiredness and achieve inter-rater reliability.

Three, the results also tell us that fairness in scoring was not attainable owing to the fact that the testees were not given equal time for preparation. With a single rater it's quite impossible for all the candidates to have equal time for both preparation and presentation. Definitely, in such a situation, some testees might be privileged than others. However, having more than one rater might help to achieve fairness in scoring (A. Hughes 2003, pp. 124-126). Four, another information deriving from the analysis of the data instructs us that the students are aware of some grading criteria among others pronunciation - fluency -vocabulary and grammar (See table 4). Making those criteria available to testees help them to know what is expected from them and is necessary for an effective preparation. According to A. Hughes (2003), these criteria confer a degree of reliability to a test. However, knowing the grading criteria might be useless if the students would not be submitted to individual assessment or analytic scoring.

Consequently, teachers in charge of oral practice tutorials should acknowledge that discrete scoring might help to avoid subjective scores and thus achieve score-reliability (K. Finson, C. Ormsbee and M. Jensen 2011, p. 181).

3.4 A didactic framework for an improvement of reliability in spoken production tests

Table 5: A didactic framework

Reactions Stages	Principles	Teacher's role	Student's role	Outcome
During tutorials	Make students	Designs tasks that	Participates actively	Enhances test

	familiar with testing format and technique	students are likely to meet during the test	in classroom activities	reliability
	Present the grading criteria	Explains every criterion to the students by means of illustrations	Takes note of the criteria	Helps to achieve score reliability
During the test	Allow candidates reasonable time for preparation	Gives at least 15-30 minutes to every candidate for preparation	Identifies the main ideas involved in the topic	Increases fairness and test reliability
	Do not be content with a single rater	Asks the assistance of other raters	Accepts being assessed by any rater	Fosters inter-rater reliability
	Promote analytic scoring	Uses a marking scale including the grading criteria	Keeps the grading scale in mind in order to perform accordingly	Promotes score reliability

This didactic framework presents five (5) main principles whose implementation might improve reliability in oral proficiency tests. During the first stage (tutorials), the implementation of two principles is necessary. Principle (1) relates to making students familiar with testing format and technique. This implies that during tutorials, the teacher in charge of oral practice class designs tasks that students are likely to encounter during the test. On their side, the students should actively participate in classroom activities to benefit from the training. This first principle helps to increase test reliability for it is frustrating for students to be submitted to unfamiliar tasks. Otherwise, the backwash might be negative. Principle (2) has to do with the presentation of the grading criteria to students by means of illustrations. Knowing those criteria is necessary for students to identify which aspects should be given much attention. Definitely, this second principle also permits to increase test reliability. The second stage comprises three other principles which should be considered during the test. Principle (3) insists on the fact that candidates should be granted reasonable time for preparation. At this level, teachers should give at least 15-30 minutes preparation to every candidate. Doing so is necessary to identify the main ideas involved in a topic. Moreover, it increases fairness and test

reliability. Therefore, a deep involvement of all the stakeholders (especially the administration, syllabus designers, language teachers and learners) is urgently requested for an increase in the time allotted to oral practice tutorials. Otherwise, our suggestion will remain at a theoretical level. Principle (4) shows that teachers should not be content with a single rater but might ask for the assistance of other raters to foster interrater reliability. Therefore, candidates should willingly accept to be assessed by any rater. Principle (5) encourages analytic scoring. This last principle could be implemented if a marking scale including grading criteria is used by raters during the assessment in order to achieve score reliability.

Conclusion

The aim of this study is to investigate current oral proficiency tests and suggest ways of improving reliability. The article intended to answer the following questions: which testing strategies might help to improve reliability during spoken production tests? How might their implementation help to avoid inaccurate and subjective scores? In order to find an answer to the above questions, I felt it necessary to investigate the exam rooms.

The data collected by means of questionnaires and interview guides prove that although the students were familiar with testing techniques and formats, much is left to be done in terms of reliability for four main reasons. One, an inadequate preparation due to time shortage. Two, group assessments were privileged at the expense of pair and individual assessments. Three, we noticed partiality in scoring due to the fact that there was a single teacher (rater). Four, the analytic scoring which aims to achieve score reliability seemed to be neglected for the holistic one.

In order to improve reliability of oral assessment tests, my suggestions are fourfold. First, students should be given equal and reasonable time for both preparation and presentation. Secondly, test

administrators should encourage individual assessment which might provide accurate picture of students' speaking abilities. The implementation of both suggestions enhances test reliability. Thirdly, apart from test reliability, inter-rater reliability should also be achieved. This could be by increasing the number of testers which is a prerequisite to achieve fairness in testing. Fourthly, score - reliability is also another criterion which really matters. It can be reached through the analytic scoring which provides a more objective and consistent assessment of students' work. However, for the success of these suggestions, students should be informed about the grading criteria as well as the testing techniques and formats from the outset.

Despite the relevance of the issues addressed in this paper, I could not pretend that the implementation of the above suggestions is enough to attain 100% reliability in oral proficiency tests. Rather, a lot remains to be done. Other criteria of reliability might be investigated for making oral tests much more reliable.

References

- BACHMAN Lyle, 1990, *Fundamental considerations in language testing*, Oxford, Oxford University Press.
- BLANCHE Patrick, 1988, "Self-assessment of foreign language skills: implications for teachers and researchers" in *RELJ Journal* Vol. 19, No. 1, pp.75-93.
- Brennan Robert, 2001, "An essay on the history and future of reliability from the perspective of replications" in *Journal of Educational Measurement*, 38(4), pp. 295-317.
- BROWN Douglas, 2007, *Language Assessment: Principles and Classroom Practices*, San Francisco State University, Longman.

- BURNS Anne, 2010, *Doing Action Research in English Language Teaching: A guide for practitioners*, New York, Burns Taylor & Francis.
- , 2019, *Action Research in English Language Teaching: Contributions and Recent Developments*, Switzerland, Springer Nature.
- CANAL Michael & Swain Merrill, 1980, "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing" in *Applied Linguistics*, volume 1, Issue1, Spring 1980 pp1-47, London, Oxford University Press.
- , 1983, "From communicative competence to communicative language pedagogy" in J. C. Richards and R. W. Schmidt (eds.) *Language and Communication* (pp. 2-27), New York, Longman.
- DIAZ-RICO Lynne & Weed Kathryn, 1995, *The Cross-cultural, Language, and Academic Development Handbook*, Boston, Allyn and Bacon.
- FINSON Kevin, Ormsbee Christine, & Jensen Mary, 2011, *Differentiating Science Instruction and Assessment for Learners with Special Needs, K-8*, Thousand Oaks, Corwin Press.
- HUGHES Arthur, 1989, *Testing for Language Teacher*, First published, Cambridge, Cambridge University Press.
- , 2000, *Testing for Language Teachers*, Beijing, Foreign Language Teaching and Researching Press.
- , 2003, *Testing for Language Teachers*, Second edition, Cambridge, Cambridge University Press.
- HUXHAM Mark, Campbell Fiona, & Westwood Jenny, 2012, "Oral versus Written Assessments: A test of Student Performance and

Attitudes” in *Assessment & Evaluation in Higher Education*, 37(1), pp. 125-136.

LLACH Maria, 2011, *Lexical errors and accuracy in foreign language writing*, Bristol, Multilingual Matters.

PALOMBA Catherine and Banta Trudy, 1999, *Assessment essentials: Planning, implementing, and improving assessment in higher education*, San Francisco, Jossey-Bass.

TAYLOR Linda & Galaczi Evelina, 2011, “Scoring validity” in Linda Taylor & Cyril Weir (Eds.) *Examining speaking*, pp. 171-233, Cambridge, Cambridge University Press.

TUAN Luu, 2012, “Teaching and Assessing Speaking Performance through Analytic Scoring Approach in Theory and Practice” in *Language Studies*, 2 (4), 673-679, Finland, Academy Publisher.

XI Xiaoming & Pam Mollaun, 2006, *Investigating the Utility of Analytic Scoring for the TOEFL Academic Speaking Test (TAST)*, Princeton, ETS.