Exploring the Impact of Docimology on Beninese EFL Advanced learners' Writing Composition : Case Study of some Secondary Schools in Ouémé and Plateau Region

Dr. TEBA Sourou Corneille

Doctor in English Didactics, Auxiliary Master of CAMES Universities, Faculty of Arts and Human Sciences, English Department, Adjarra Campus, University of Abomey-Calavi

Abstract: *This paper exposes the impact of docimology on Beninese EFL advanced learners writing skills development. It aims at describing the importance of docimology in the development of language skills. A mixed method which consists in collecting both quantitative and qualitative data has been adopted. The instruments used, include questionnaires, interviews, classroom observations and surveys of summative examination papers. The collected data were analysed using descriptive statistic for correlation and significance test. A survey of twenty (20) examination papers and model answers designed by our sample evidences the mediocre proficiency in docimology. The impact of this docimology in writing skills development was assessed by comparing one hundred (100) learners' writing production scores using analytic and holistic assessment model. The results reveal that learners have better score for the analytic evaluation model. Some strategies are worked out to provide solutions to grading challenges.*

Key words*: Docimology, advanced, analytic, holistic, writing skills.*

1. INTRODUCTION :

Writing is the most challenging skill for language teachers because students have less experience with written expression. In fact, having a good level of spoken English does not necessarily imply good performance in writing. Most people tend to write as they speak or to translate from their first language. The outcome is unclear to the reader and/or the style may be inappropriate. There are also often grammatical mistakes caused by first language interference.

While language proficiency can be accessed through spoken language in face to face communication, academic writing is the basis upon which a candidate's achievement, learning and intelligence are assessed (Ghalib & Al-Hattami, 2015). *Good writing* skills are critical to academic, professional success and the today competitive job market. Unfortunately, writing remains neglected in EFL teaching-learning process owing to the focus on spoken language and the difficulties related to its skills development.

Effective writing skills development only takes place with an accurate assessment of learners' achievement all along the teaching-learning process. The model of assessment is determined by the principles of docimology effective in the teaching-learning approach adopted by educational stakeholders. Docimology is a subject that frameworks evaluations in a teaching-learning process in order to disclose good and bad practices and then consolidate the former and improve the latter (Gueye, 2002)

Today, for language teaching professionals, docimology is the science of evaluation and evaluation itself is at the service of decision making (De Ketele, 1993). For De Ketele, evaluation means to (1) collect a set of adequately appropriate, valid and reliable information, (2) study the adequacy rate between this set of information and another, (3) set criteria suitable to the assigned objectives from the onset or streamlined on the way, in order to make a decision (De Ketele, 1993). In writing skills development, evaluation means marking learners' writing productions. As a matter of fact, docimology for EFL teachers tends to be restricted to designing evaluation sheets and model answers and grading. The current research seeks to reflect this conception and focusses on the evaluation and the grading model part of docimology.

The model of evaluation is important as it determines the quality of the feedback provided which can drastically impact the learning process. Two models/rubrics are generally adopted in writing skill development evaluation: the analytic rubric and the holistic rubric. With the advent of the Competency-Based Approach (CBA) in Benin, the analytic scoring has been chosen over the holistic one. Still, after two decades of implementation, both models still remain. The analytic evaluation is imposed during national exams for accuracy and higher-grade purposes while the holistic model is used by EFL teachers in daily classroom practice because it is time-saving and easier, especially in large classes. The mismatch between teaching and evaluation process leads to an ineffective writing teaching-learning process and learners' demotivation. The typical consequence in Benin EFL classes is a punitive grading that leads students to skip writing tasks during evaluation and national exams.

The main objectives in carrying out this paper on the impact of docimology on EFL advanced EFL learners' academic performance are to:

- point out the impact of grading on student's academic achievement
- shed light on the requirements of a good evaluation
- to draw attention to the specificities of analytic and holistic evaluation so as to promote a consistent evaluation practice in writing skills development
- help Benin's authorities revisit English Teaching Training.

The paper turns around the following questions:

1. How important is docimology in writing skills development?
2. How do the scoring model/rubric impact students' motivation and academic achievement?
3. How can teacher use appropriate docimology principle to design evaluation paper and use the adequate scoring rubric in writing skills development?

This paper helps teachers acquire knowledge about docimology , presents strategies which can be employed in order to solve the problem of grading for an effective EFL teaching-learning-evaluation process in the Benin Republic.

2. THEORETICAL KEYSTONES:
2.1. Meaning of Docimology
The term "*docimology*" is proposed by Pieron (1964), who combines two Greek words: "*dokimê*" which means "*Test*" and "*logo*" which means "*reason*, *speech*, *study*". The docimastic "*dokimastikos*" refers to the study of exam techniques. This discipline focuses on scientific bases, the criteria of school evaluation in order to develop exam techniques and the control of students' results. This is the science of evaluation in pedagogy. One of the characteristics of the docimology attitude is the recourse to the statistics technique as a basis for granting the highest objectivity of examinations. As the systematic study of examination (modes of scoring, inter-individual and intra-individual variability of examiners; subjective factors), docimology is a science whose object is all that is related to the measurement and evaluation in the educational field.

2.2 Docimological Principles
Stan (2001:85) states that:

> *The docimological principles are fundamental theses, general rules with descriptive and*
> *normative character, which base the evaluation project, organizing and development*
> *inorder to ensure their scientific consistency and efficiency.*

The most relevant docimological principles, which lead evaluation activity, are:

- The principle of objective evaluation character refers to the structure and organization of evaluation, so that the pupils' performances to be reflected and evaluated in a real and relevant manner, limiting as much as possible the influence of external factors;
- The principle of interactive evaluation character expresses the fact that learning evaluation is inherently connected to the evaluation and determined both by the evaluation made by the teacher and by the pupil's self-evaluation activity;
- The principle of the pupil's performances contextualization: regards the fact that during the evaluation there have to be considered the performance and there have to be used such tasks that could reflect the reality, meaning to attach the pupil's capacity to adapt the knowledge to various situations.

2.3 Docimological Test
As principal methods of objective evaluation, Nicolas, (2000:401) says that the docimological test is "*a set of questions that help one to check and evaluate the knowledge and the capacities acquiring to operate with them, by reporting the answers to a sample appreciation scale, previously elaborated*". In specific literature people also encountered other terms: pedagogical test, knowledge test, learning evaluation test, performance test or simple test to designate the instrument and method of evaluation which has a specific element the item, being characterized by a greater objectivity in evaluating the results. The quality of information offered by the testing depends on two sets of attributes the test has to possess:
1. Psycho - pedagogical: the test must be appropriate to its specific purposes and comprehensive.
2. Statistical which guarantee the perfection of the test as a measurement instrument; the most important being the accuracy and validity.

2.3.1 Test Validity
The validity represents the most important quality of the test; show whether the instrument measures what it supposes to measure and how well it makes it.
In establishing the validity of a test, two questions are asked:
1. Does the test measure what it is meant to?

2. Can it be used in taken the right decisions?

Concerning the purpose, validity can be:
• of content: the test must refer to those contents referred to during the instruction
• of criteria: involving the rapport to an external criterion

Cabac, V (1999:231) states that "*The first operation made after applying the test is to correct and marking the answers. There are two points of view in marking the answers at the items of the brief tests*"
   1. The first point of view is founded on the concept of *"errorless activity",* used in preparing the specialists in different fields of techniques. According to this point of view, in the professional activity (programming, building plans, etc.) there are no options between minor and major errors. The computer program works or not. Therefore, in evaluating the specialist preparing, it is not considered the indices of difficulty of the items, but only the fact that they are solved right or wrong, omitted or partly solved is not considered.
   2. The second point of view is founded on the thesis that the marking of the item must reflect its index of difficulty (the relative frequency of the individuals who have answered correctly of the all examinees who answered to that item).

2.3.2 Test Reliability
Reliability is defined as the extent to which a questionnaire, test, observation or any measurement tool produces the same results on repeated trials. In short, it is the stability or consistency of scores over time or across ratters. Similarly, the degree to which an individual's responses (i.e., their scores) on a survey would stay the same over time is also a sign of reliability. There are three aspects of reliability, namely: equivalence, stability and internal consistency (Miller, 2005)
a) **Test-Retest Reliability:** to gauge test-retest reliability, the test is administered twice at two different points in time. This kind of reliability is used to assess the consistency of a test across time.
b) **Inter-ratter Reliability:** this type of reliability is assessed by having two or more independent judges score the test. The scores are then compared to determine the consistency of the ratters' estimates.
c) **Parallel-Forms Reliability:** parallel-forms reliability is gauged by comparing to different tests that were created using the same content. This is accomplished by creating a large pool of test items that measure the same quality and then randomly dividing the items into two separate tests. The two tests should then be administered to the same subjects at the same time.
d) **Internal Consistency Reliability:** the objective is to judge the consistency of results across items on the same test. Essentially, you are comparing test items that measure the same construct to
determine the tests internal consistency.

The following factors will influence the reliability of a test:
-   **Administration**. It is essential that each student have the same time, equipment, instructions, assistance, and examination environment. Test directions should be strictly enforced.
-   **Scoring**. Objectivity in scoring contributes to reliability. Every effort should be made to obtain uniformity of scoring standards and practices.
-   **Standards**. The standards of performance that are established for one class should be consistent with those used in other classes. A change in grading policies not based upon facts, uniform standards, and experience factors gained from other classes will affect the reliability of test results.
-   **Instruction**. The reliability of tests results will be affected if the instruction presented to a class tends to overemphasize the teaching points included in the examination. This is often known as "*teaching the test*" and is undesirable. When the instructor gives students obvious clues as to the test requirements, he not only affects the reliability of the test, but he insults the intelligence of his class.
-   **Length.** The more responses required of students, the more reliable will be the test or measuring device.

2.3.3 The Competency-Based Approach Concept related to Docimology
Lasnier (2000:9) defines competency as "*a complex practical knowledge resulting from the integration and mobilization of a set of abilities and pieces of knowledge efficiently used in a correlated situation to solve more or less complex situations*". With such a problem-solving objective, the Competency-Based Approach is therefore a curriculum aiming at developing in students, the competencies they need to successfully face life-problems in the future. These competencies are classified into three main categories as follows: disciplinary competencies; transversal competencies and transdisciplinary competencies. According to the same author, the disciplinary competencies are related to the school subjects such as French, English, Mathematics, Physics and Biology etc.

Disciplinary competencies are the abilities each subject aims at developing the needed skills in every student, in the different fields of human knowledge. In English, they are: (1) communicating orally in English; (2) reacting to texts after reading or listening to them; and (3) producing texts of various types and functions.

The transversal competencies are the ones students need to develop throughout all the school subjects while carrying out the teaching/learning activities. They are the competencies that will help students face real life situations better in their future life. They are eight: (1) exploiting available information; (2) solving problem-based situations; (3) using one's critical sense; (4) displaying one's critical sense; (5) managing one's acquisitions to perform tasks; (6) working in cooperation; (7) displaying one's sense of ethics; and (8) communicating in a precise and appropriate way. Dealing with docimology, the concept of disciplinary competencies takes into account the marks giving to students after their reactions from the text given; it would be the result of the emotional intelligence of those EFL learners. In the case of texts production of various types and functions, it leads to a variety of marks given to students. In the concept of the transversal one, students exploit the available information, solve problem-based situations, use and display their critical sense and manage their acquisition to perform tasks. So, it can be said that disciplinary and transversal competencies are part of docimology because, without students' productions, docimology cannot be applied.

Concerning the transdisciplinary competencies, they are referred to as competency of life, as they are the one directly connected to the socio-professional life (Bankole-Minaflinou, 2013). They aim at preparing the students and equipping them with the qualities and abilities of a responsible citizen fully aware of his/her role in his/her community. There are six trans disciplinary competencies: (1) asserting one's personal and cultural identity in a world in constant change; (2) acting individually and collectively in mutual respect and open-mindedness; (3) getting ready to integrate the professional life in a prospect of self-achievement and social integration; (4) PDF Pro Evaluation having riskless behaviors in matter of health, sexuality and safety; (5) acting in harmony with the environment and in prospect of sustainable development; (6) acting as a wise consumer by a responsible use of goods and services.
In every sequence of class, most of these competencies are developed directly or indirectly.

2.4 Importance of Grades/Scores

Wormeli (1961:76) explained the role of grades by stipulating that "*Grades are first and foremost communication; they are information, nothing more. The moment we make them something more, we corrupt their constructive use*". This restrict grade to data to be analysed so that decision be taken which is interesting. Grade provide a valuable feedback on the whole teaching learning process.

Grades are an effective means of motivating students to do their best in school. If there is no grade, students will not work hard. By receiving grades, students recognize their positions in class, their weaknesses. If their positions came to be low, they then try to achieve higher positions by working hard.

2.4.1 The Grading/Scoring Models

The grading model also known as scoring rubric is a tool for evaluating the quality of student work on a continuum of performances from excellent to poor (Schafer, 2004). It contains a set of well-established criteria corresponding to a scale of possible points to be assigned in scoring a piece of work, spoken or written (Campbell, Melenyzer, Nettles, & Wyman, 2000). The best performance is assigned the highest point and the worst the lowest point on the scale. A scoring rubric provides descriptors for the different levels of proficiency on the scale. These descriptors are detailed enough to enable sufficiently fine judgments, and rich enough to enable reliable, unbiased and valid discrimination.

Herman, Aschbacher, and Winters (1992) posited four characteristic features of a rubric—criteria, standards, scale, and examples. An effective rubric encompasses with (1) a well-defined list of criteria clearly specified for both teachers and learners to know what is expected for them, (2) standards of excellence for the different levels of performance, (3) gradations of quality, or a scale, based on the degree to which the standard has been met. The gradations are constituted by detailed descriptions that represent what should earn which point on the scale, (4) modal exemplars of expected performance at the different levels on the scale.

Weigle (2002) argued that there are three types of rubrics used in the evaluation of written proficiency. These are primary trait, holistic and analytic scoring rubrics. These three types differ in their impact, discriminatory power, inter-rater reliability, the degree of bias, and the cost-effectiveness—in terms of time, effort and money (Kuo, 2007). The choice of one scoring rubric or the other is significant because if "*represents, implicitly or explicitly, the theoretical basis upon which [a] test is founded*" (Weigle, 2002:109). This study focusses on holistic scoring rubrics and analytic scoring rubrics,

2.4.2 Holistic Grading/Scoring

Holistic scoring is "a global approach" to scoring that is underscored by the idea that *"writing is a single entity which is best captured by a single scale that integrates the inherent qualities of the writing"* (Wiseman, 2012:59). As

such, holistic model considers the entire written composition and assigns an overall score to the performance (White, 1985; Weigle, 2002; Hyland, 2002). This cost-effectiveness of holistic scoring makes it a suitable approach for large-scale assessment of written performance, especially for decisions concerning placement (Cumming, 1990; Hamp-Lyons, 1990; Reid, 1993).

Holistic scoring criteria consist of general guidelines that define good performance at each score point. Number of researchers (e.g., White, 1985; Cohen, 1994) argue that holistic scoring focuses on the strengths of the writing rather than on the deficiencies.

The holistic scoring generates a composite score that *"does not provide specific evidence of where and how much additional writing instruction is needed"* (Becker, 2011:116). Although may be seen this shortcoming, Weigle (2002) contends that holistic scorings are very practical. They are short, do not include detailed criteria of evaluation, and make possible the evaluation of an essay by assigning one score to it after only one reading. Holistic scorings are therefore appropriate for evaluating written performance in large-scale assessment contexts. This reflects the typical teaching conditions in Benin with an average of 60 students per classes and 5 to 6 classes per EFL teacher for a total number of students ranging from 250 to 350 students. Also, holistic scoring is also ideal assessing written performance in the computer-based Test of English as a Foreign Language (TOEFL), Graduate Record Examination (GRE), and Graduate Management Admission Test (GMAT).

2.4.3 Analytic Grading/Scoring

As an alternative, analytic grading, which involves "*the separation of the various features of a composition into components for scoring purposes*", has received significant scholarly attention (Wiseman, 2012:60). An analytic scoring typically includes writing components relating to the test taker's lexical, syntactic, discourse, and rhetorical competence. As such, an analytic scoring rubric offers more detailed information about a test taker's writing performance than does the single score of a holistic scoring rubric. An analytic rubric provides organized and comprehensive feedback to teachers and enable them to discriminate the weak and strong aspect in students writing performance (Hamp-Lyons, 1995; Crehan, 1997). In other words, an analytic rubric has higher discriminating power (Mendelsohn & Cumming, 1987).

The present study adopts, indeed adapts, Bachman and Pamer's (1996) model of communicative language ability and the rubric based on the model. According to the model, the ability to write an essay requires knowledge schemata (knowledge of the topic), strategic competence (strategies for content development), rhetorical knowledge (strategies for producing cohesive supporting arguments), grammatical competence, and knowledge of vocabulary and register.

2.4.4 Analytic versus Holistic Grading

There are a number of studies that have compared the behavior of holistic and analytic rubrics with interesting findings. Chi (2001) compares holistic and analytic scoring rubrics, using many-faceted Rasch measurement, in terms of the appropriateness of the scoring rubrics, the agreement of the student scores, and the consistency of rater severity. The study reports significant differences between raters using holistic scoring rubrics, but not analytic scoring rubrics. Other studies confirm this advantage of analytic scoring in terms of inter-rater and intra-rater reliability (Al-Fallay, 2000; Easy & Young, 2007; Knoch, 2009; Nakamura, 2004). Analytic scoring also provides an individualized profile of the test taker's written performance (Weigle, 2002) and direct, useful feedback to students and teachers (Brown & Hudson, 2002). For this reason, analytic scoring rubrics are often chosen for placement and diagnostic purposes (Jacobs et al., 1981; Perkins, 1983; Hamp-Lyons, 1991).

By contrast, holistic scoring rubrics offer the advantage of reduced cost in time and money (Wiseman, 2012). Bauer (1981) compared the cost-effectiveness of analytic and holistic scoring rubrics in scoring secondary school students' essays. The study reports that the time needed to train the raters to use the analytic rubrics was two times the time needed to train raters to use the holistic rubrics, and the time needed to grade the essays using the analytic rubrics was four times the time needed to grade the essays using the holistic rubrics. Other studies in different other contexts have reported similar findings (Klein et al., 1998; Arter, 1993; Bainer & Porter, 1992). For this reason, holistic scoring is the preferred method of scoring in large-scale testing contexts that involve a large concentration of test takers taking the test at the same time (Becker, 2011).

The choice of one type or rubric or the other, therefore, depends mainly on the purpose of using the rubric and is driven by context-specific considerations. The present study is an extension of this tradition of examining the performance of holistic and analytic scoring rubrics.

3. METHOD:
3.1 Target Population

This research paper is carried out in some Beninese secondary schools of Ouémé and Plateau regions with EFL teachers and students. To observe the impact of evaluations and the grading system on EFL beginners' academic performance, teachers and learners are the centre in the field of this investigation because they are the main actors.

3.2 Sampling

A sample of one hundred (100) students and twenty-five (25) teachers is selected for the account of this study.

Table 1: Selected Teachers and Learners

Secondary schools	Status	Teachers'sample	Learners' sample
CEG Honvié	Public	05	19
CEG Avrankou	Public	04	23
CEG Malanhoui	public	03	8
CEG Tchaada	Public	05	15
CEG1 Ifangni	Public	04	12
CEG1Sakété	Public	04	23
	Total	**25**	**100**

3.2.1 Teachers

In the process of docimology, teachers are the main actors. It is up to them to use it during classroom activities in order to make students get a good academic achievement. They are taken into account for this study because the experiences and difficulties they encounter while teaching are helpful.

3.2.2 Learners

They are important because no teaching activities can be effective if learners are not taken into consideration. They are the target of any teaching and are directly affected by the teacher's way of marking in the class. So, a few questions to them would surely add to the data, the analysis of which is necessary for a better understanding of the impact of docimology on their writing composition achievement.

3.3 Research Instruments

The main instruments used for this paper are:
- Questionnaires to both EFL teachers and learners,
- Interviews with EFL teachers and other teachers,
- Classroom observations.
- Survey of summative examination papers and model answers
- Assessment of students' performance with good and bad evaluation

Each of them is worth describing to have a clear understanding.

3.3.1 Questionnaires to Learners and Teachers

➤ **To learners**

A set of questionnaires is addressed to the students in advanced classes (2nd and 1ere) to have their opinions about docimology. This questionnaire is made of four questions described as follows:

Question n°1 and n°2 deal with student's perception of teachers grading models. Question n°3 exposes learners difficulties related to writing skills development. Question n° 4 deals with learners' suggestions.

➤ **To Teachers**

The teachers' questionnaire contains six (06) questions and is described as follows:

Question n°1 assesses the importance of evaluations. Questions n°2 discloses teachers' challenges as far as marking papers is concerned. Question n°3 check teachers' priorities during the grading process. Question n°4 and n°5 exposes the different biases affecting writing compositions assessment. Questions n°6 and n°7 investigates hindrances to writing skills developments and the grading rubrics used in the evaluation process.

3.3.2 Interviews

Interview is a research design which consists in asking questions to experienced people in the field of one's research.

To collect information five (5) teachers' advisers purposely selected, have been interviewed. The interview aims at deepening the answers provided by the teachers through the questions such as:
- ✓ What is docimology?
- ✓ Which grading/scoring model do they recommended and why?
- ✓ To what extent do the scoring/grading affect writing skills teaching learning experience?
- ✓ What are their perception of biases that could affect the teaching learning process and to what extent can docimology help?

Their various answers have enabled me to know what they think on using this science in EFL classes. For further information about the problems, classroom observation has been undertaken.

3.3.3 Classroom Observation

In order to assess real life teaching practices of the selected population, classroom observations are carried out. The purpose of the classroom observations is to check if verbal scales are used during activities. This instrument provides relevant data that helped to crosscheck the data obtained from the other research instruments.

3.3.4 A Survey of Summative Examination Papers and Model Answers

For evaluations to reach their full potential, their quality should be as high as possible. Four (criteria) requirements are important in this respect: reliability, validity and accuracy. The examination papers and their model answers (correction key) are evaluated using the following grid. Each item scores from 1 to 5. The total of the (3) items gives the overall score of the evaluation papers.

Table 2: examination paper assessment grid

Item	1	2	3	4	5
Reliability					
Validity					
Accuracy					

1=Very poor; 2=Poor; 3= Fair; 4= good; 5= excellent

3.3.5 A Survey of Student Writing Composition

This instrument illustrates the impact of docimology on students' performance in essay writing. Ninety-five (95) Students writing composition papers are graded using both holistic and analytic (criteria-based) scoring as recommended in the CBA. The scores are compared and analysed using the descriptive statistic for correlation and significance test. The dependent variable here is learners score and the independent variable is type of grading. The Analysis of variance ANOVA of the result is provided.

3.4 Data Collection and Analysis Procedures

Questionnaires have been distributed to teachers as well as learners of the targeted schools. Many of the students succeeded in filling the boxes and returned the questionnaires since the French version of their questionnaire was made. To teachers, twenty-five (25) questionnaires have been addressed and all the answers from them have been got back. One hundred (100) questionnaires have been addressed to the students and all have been returned. Classroom observation was undertaken during this research work in the chosen schools (CEG Honvié, CEG Avrankou, CEG Malanhoui, CEG Tchaada, CEG1 Ifangni, and CEG1 Sakété) and some EFL teachers have been met for the interview.

Formative and summative evaluation have been analysed. All the data collected have been analysed and the results have been presented in tables, charts and pie-charts according to each instrument.

4. RESULTS :
4.1 Results from Questionnaires
4.1.1 Teachers' Responses
4.1.1.1 Importance of evaluation

Table 3: Importance of evaluation

Importance of evaluation	Frequency	Percentage
Monitoring the Teaching learning process	24	96
Motivating learner and teachers	16	64
Measuring improvement over time	25	100

Table 3 shows that most teachers (96%) acknowledge that evaluations monitor the Teaching process. In addition, they unanimously claimed that evaluations help to measure improvement over time. However, the role of motivation tool of evaluations is recognised by only 64% of the respondents.

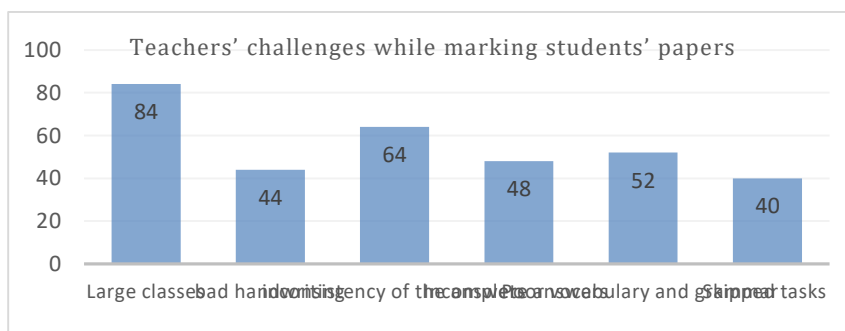### 4.1.1.2 Teachers' Problems while marking Students' Productions

Figure 1: Teachers' challenges while marking students' papers

According to the results of figure1, the main challenges teachers face while marking papers, are related to large size EFL classes (21%), inconsistence of learners' answers 16%, poor vocabulary and grammar 13% and bad handwriting (12%.) Other challenges such as skipped tasks during evaluations and incomplete answers respectively rank 11% and 10%.

4.1.1.3 Teachers' Priority while Marking Students' papers

Table 4: Teachers' Priority while Marking Students' papers

Teachers' Priority while Marking Students' papers	Frequency	Percentage
Accuracy of the provided answers	10	40
Skills/subskills development	7	28
Overall achievement	*8*	*32*

Predictably, table 4 shows that for 40% of our sample the priority while marking papers is the accuracy of the provided answers. Regrettably, 28% only, focus on learners' skills development. The remaining respondents, that is 32 %, opt for overall achievement.

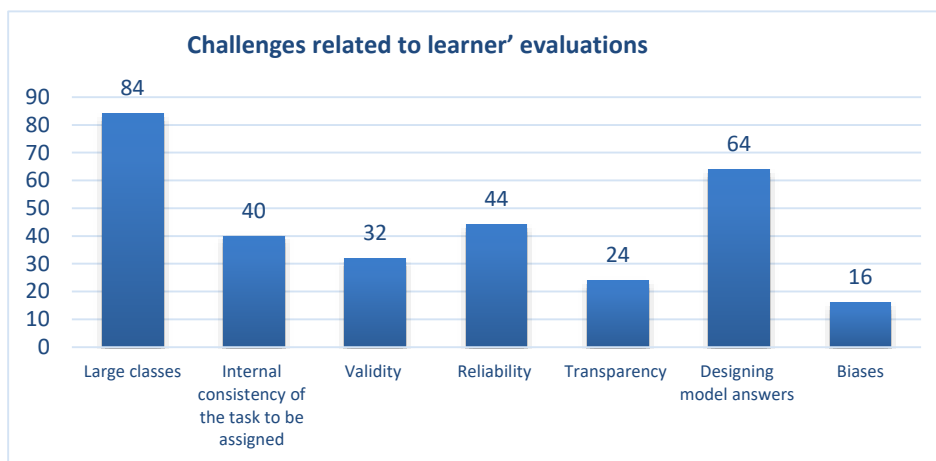**4.1.1.4 Challenges related to Learner' Evaluations**

Figure 2: Challenges related to learner' Evaluations

Figure 2 exposes that Large size EFL classes 84%, and model answers designing 64% are the core hindrances experienced by our sample. Still, essential quality features of evaluation papers such as validity 32%, reliability 44%, transparency 24% and internal consistency 40% do not get much attention. The current table also discloses that biases that can affect teacher perception are taken into account by only 16% of the respondents.

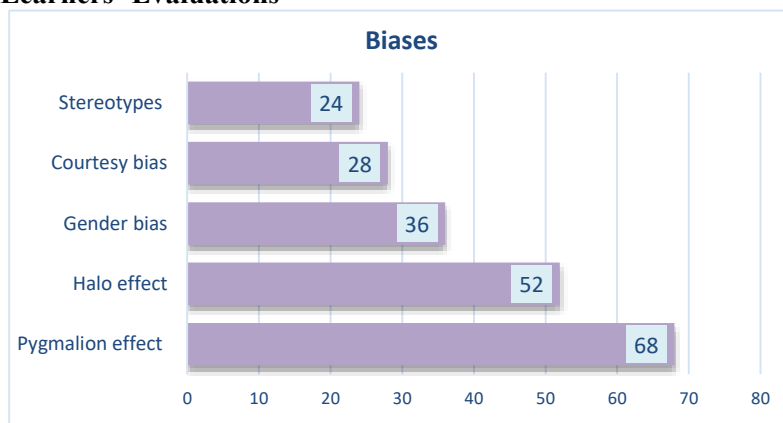**4.1.1.5 Biases related to Learners' Evaluations**

Figure 3: Biases related to Learners' Evaluations

The data of figure 3 evidence that there are some biases that affect teachers' perceptions in the process of evaluations such as the Pygmalion effect (68%), the halo effect (52%), the gender biases (36%), the courtesy biases (28%), and other stereotypes (24%).

4.1.1. 6 Teachers' Biases blind Spot

Table 5: Teachers' Biases blind Spot

Opinion	Frequency	Percentage
Strongly agree,	*1*	*4*
agree,	*5*	*20*
neither agree nor	*10*	*40*
disagree,	*9*	*36*
strongly disagree	*0*	*0*

Predictably, from table 5, most of the teachers (40%) are neutral as far as the mentioned biases are concerned while 36% contend their evaluations are not biased. Happily, 24 % of them that is 20% (agree) and 4% (strongly agree) acknowledge the impact of biases in the evaluation process.

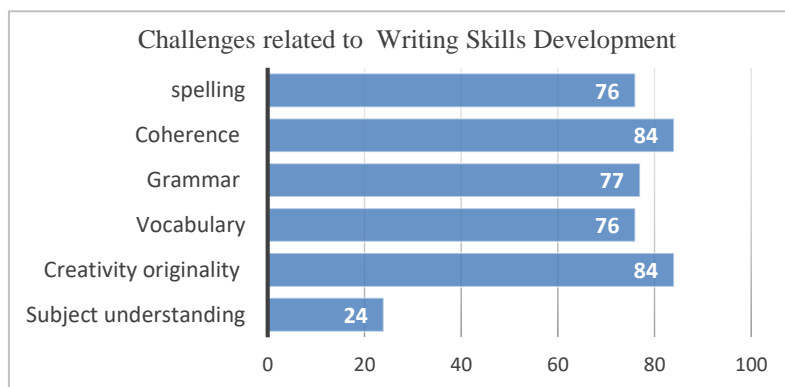#### 4.1.1. Challenges related to Writing Skills Development

Figure 4: Challenges related to Writing Skills Development

Figure 4 discloses the challenges related to writing skills development. Understandably, EFL teachers share the same difficulties with learners' creativity and coherence which score 84% each. Also, hindrances related to Vocabulary 76%, grammar 77% and spelling 76% skills are almost equally experienced within our sample. Comfortingly, only 24 % of the respondent teacher reference subject understanding as a hurdle.

4.1.1 Appropriate Writing Skills Assessment Strategy

Table 6: Writing Skills Assessment Strategy

Writing skills assessment strategy	Frequency	Percentage
Holistic	18	72
Analytic	4	16
Both	3	12
Total	25	100

The data from table 6 reveal that the holistic assessment is used by 72% of EFL teachers against only16% for analytic assessment. The remaining 12% of our sample state that they use both types of assessment.

4.1.2 Students' Responses
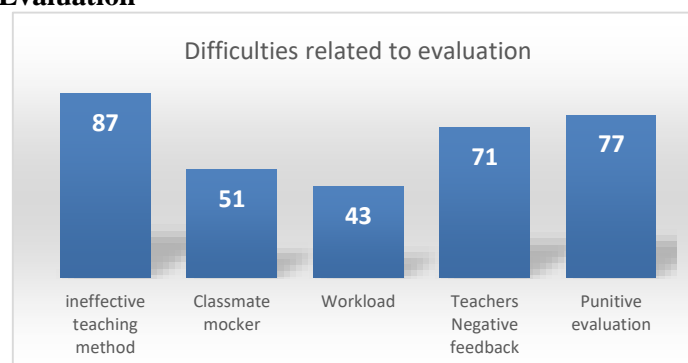#### 4.1.2 Difficulties related to Evaluation

Figure 5 : Difficulties related to Evaluation

Figure 5 evidences that the challenges faced by learners are mostly the ineffective teaching method 87%, teachers' negative feedback 71% and the fear of evaluation 67%. Other noteworthy impediments that must be tackled are related to classmate mockers 51%, and the workload 43%.

4.1.3 Data collected from the Experimentations
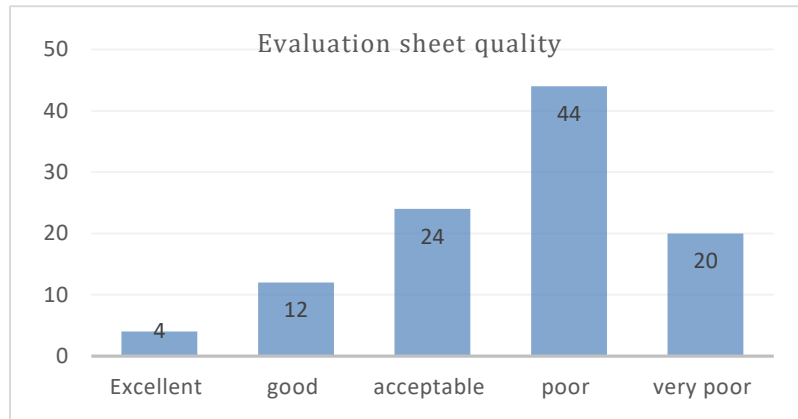## 4.1.3.1 Teachers Proficiency in Docimology

Figure 6: Teachers Proficiency in Docimology

Regrettably, the survey of writing evaluation sheet, displayed in figure 6 reveals that the assessed papers score from acceptable (24%) to very poor 20%), with the majority that is 44%, unpredictably, scoring very poor.

4.1.3.2 Students' Writing Composition Survey
The impact of docimology (grading/scoring rubric) in language development can be measured through analysis of variance (ANOVA) of our sample summarized in the following Tables.

Table 7 ANOVA table: Overall score in writing * type of evaluation

			Sum of Squares	df	Mean Square	F	Sig.
overall score in writing * grading rubric	Between Groups	(Combined)	450.667	1	450.667	48.927	.000
	Within Groups		865.833	94	9.211		
	Total		1316.500	95			

Table 8: Measures of Association of the overall Score in Writing * Status

Measures of Association		
	Eta	Eta Squared
Overall score in writing * grading rubric	.585	.342

The measure of association Table 2 indicates that in the current case the Eta squared value expressing the Effect size is ES= .342 as $0.04 < ES \leq 0.36$, the correlation between the dependent (overall score in writing) and the independent variable (the type of evaluation) is significant but moderate.

These data are interesting as they confirm the predicted negative impact of poorly implemented docimology on literacy skills development.

4.1.3 Results from Interview
The interviews show that teachers have limited knowledge about docimology. However, they acknowledge its importance for students' achievement. They suggest appropriate training so that they master evaluation implications for the more efficient teaching learning evaluation process.

4.1.4 Classroom Observation Report

The Classroom observations reveal that the teachers start their course in a formal way according to the curriculum and attempt to manage their classes. Interestingly, most of the observed teachers use verbal scales to encourage students. This create a participative atmosphere necessary for effective teaching learning process.

4.2 Discussion of Results

Most of Beninese EFL teachers are unfamiliar with docimology. Predictably, the survey of writing examination paper reveals their mediocre performance. There are manifest problems of reliability, validity and accuracy. This is a situation justified by their flawed initial training and/or in-service training. Hence, effective evaluation becomes challenging as teachers do not know more than the traditional task of designing evaluation papers and grading according to empirically inherited models. This results into biased evaluations, harmful for the teaching-learning-evaluation process where only grades and average are prioritised in detriment to learner's performance. Understandably the respondents alarmingly underestimate the biases they are subject to. Poor docimology engenders poor teaching experience.

Docimology is vital for effective evaluation and appropriate grading. In essence, grading is an exercise in professional judgment. It involves the collection and evaluation of evidence on students' achievement or performance over a specified period of time, such as nine weeks, an academic semester, or entire school (Pollio, 2000). The ANOVA of analytic and holistic grading shows a significant correlation between learners' performance and grading model. It turns out that learners get better grade with the analytic grading. This does not mean that learners perform better. It means that they have rather positive feedback and a comprehensive view of their performances in each aspect of writing composition skill development.

Guskey & Pollio (2001) advocate that Grades have some value as rewards, but no value as punishments. In EFL environment where students lack intrinsic motivation, the existence of grades are the best motivator/demotivator and important factors in determining students' commitment to learn. Most students view high grades as positive recognition of their success, and some work hard to avoid the consequences of low grades. Instead of prompting greater effort, low grades usually cause students to withdraw from learning. This corroborate the outcomes of the interviews and classroom observations where students disinvest creating a boring class experience. Many students regard the low grade as irrelevant and meaningless. Other students may blame themselves for the low mark, but feel helpless to improve. There is strong evidence that this is detrimental to relationships among students and between teachers and students. Modern research has also shown that the seemingly direct relationship between aptitude or intelligence and school achievement depends on instructional conditions (Pollio, 2000).

Nordquist (2019) states that *Holistic grading* is a method of evaluating a composition based on its overall quality. Developed by the Educational Testing Service, holistic grading is often used in large-scale assessments, such as college placement tests. Graders are expected to make judgments based on criteria that have been agreed upon before the start of an evaluation session. Holistic grading is only useful as a time-saving approach, as it does not provide students with detailed feedback (Nordquist, 2019). But, how effective can a teaching learning process be without detailed, accurate and objective feedbacks? Those feedbacks help to monitor learners' progress over time and strategies. They are vital for adjustments, remediations for the improvement of the learning experience. Teachers who practice holistic grading refuse to break down a student's essay into separate problems like punctuation and paragraphing, but base their grade on their immediate 'sense of the whole' derived from a deliberately 'non analytical' reading (Rosenthal, 1984). Obviously, it leaves little chance for effective writing skills development.

In EFL context, Holistic grading is probably not the best tactic, even if it seems the easiest and quickest. Assigning a single score, grade, or judgment leaves the student unsure about both quality and content. (Robert, 2013). When the main objective of evaluation is discounted, no real writing skill development can take place. As the results evidence, Beninese EFL learners writing achievement is merely mediocre. EFL teacher justifiably argued that the choice of holistic grading is evident if the poor teaching conditions, the curriculum and the implementation of the CBA are not improved and adequate support in terms of training and in-service training provided. But, should EFL teacher wait for hypothetic changes from the education stakeholders before working for efficiency? Clearly, they underestimate the impact of the grading model on their professional achievement. Doing that, they are jeopardizing their own performances.

The requirements of the CBA induct a rational analytic assessment model that enable EFL teachers to monitor learners progress in skills and subskills development rather than an economic approach perfect for getting marks to set average and restraining English learning to a '*demotivating school subject*'. The development stage of language requires an analytic approach that benchmarks critical aspects of language skills improvement in a continuing action and feedback process where teachers-learners partnering lead to a cohesive teaching learning experience. Indeed, such an approach is resource demanding, but the needs of consistent writing skill development imposed it.

5. CONCLUSION :

The basic aim in carrying out this research work is to find strategies and techniques to solve the problem of grading.

Through the answers provided by both respondent students and teachers, in addition to the data collected from classroom observations and conversations, results show for Beninese EFL students, English learning and writing kills development are affected by a great deal of obstacles due to the grading rubric. Moreover, from my findings many teachers due to the time and resources constraint rely on holistic grading in detriment to an effective and analytic approach required by the CBA. Such conditions do not allow learners to get a good academic achievement.

To crown it all, it is suggested to train all EFL teachers for the acquisition of instructions, strategies and techniques in docimology so as to cope with the grading challenges in Beninese educational system.

REFERENCES :

1. Bachman, L. F., & Palmer, A. S. (1996) Language Testing in Practice. Oxford: Oxford University Press.
2. Bankole-Minaflinou (2013) Using Grouping to Foster Effective Classroom Management in the Competency-Based Approach Context in the Republic of Benin. *Particip'action.* Volume 5, Nº2.
3. Becker, A. (2011). Examining Rubrics Used to Measure Writing Performance in US intensive English Programs. The CATESOL Journal, 22(1), 113-130.
4. Cabac, V. (1999). *Evaluareapr intesteînÎnvaţământ*, Ed. Universităţii Al. Russo Bălţi 2000.
5. De Ketele, J.M., Roegiers, X. (1993). Méthodologie du Recueil d'Information. Bruxelles, De Boeck, p. 74.
6. De Landsheere, G. (1980). *Evaluation Continue et Examens - Précis de Docimologie, Bruxelles,* Ed. Labor, Education 2000.
7. Disha, M. (2017). Evaluation In Teaching and Learning Process/Education. Retrieved from www.yourarticlelibrary.com/static-2/evaluation-in-teachingandlearning-process-education/92476
8. Ghalib, T. Al-Hattami, A. (2015). Holistic versus Analytic Evaluation of EFL Writing: A Case Study. English Language Teaching. 8. 225 - 236. 10.5539/elt.v8n7p225.
9. Gueye, B. (2002) The Challenges for Science. *Education for the Twenty-First Century Pontifical Academy of Sciences,* Scripta Varia 104, Vatican City 2002 www.pas.va/content/dam/accademia/pdf/sv104/sv104-gueye.pdf
10. Guskey, T., R. & Bailey, J., M. (2001). *Developing Grading And Reporting Systems For Student Learning*. Thousand Oaks, Ca: Corwin.
11. Guskey, T., R. (2001). Helping Standards Make the Grade. *Educational Leadership 59 (1):20–27.*
12. Mahmoodi-Shahrebabaki, M. (2018). Assessment, Evaluation, and Testing: What are the Differences?. https://www.researchgate.net/publication/323218570
13. Miller, M.J.(2005) Reliability and validity. *Graduate Research Methods, Western International University*
14. Nicola, I., (2000) *Tratat de pedagogieşcolară*, Ed. Aramis, Bucureşti.
15. Nordquist, R.(2019) Holistic Grading (Composition).Glossary of Grammatical and Rhetorical Terms https://www.thoughtco.com/holistic-grading-composition-1690838?print 1/3
16. Peggy Rosenthal, P. (1984), *Words and Values: Some Leading Words and Where They Lead Us.* Oxford University Press.
17. Pierron H., Henri L.,H. (1964). *Études Docimologiques sur Le Perfectionnement des Examens et Concours*, Paris, Conserv.nat. des arts et métiers.
18. Pollio, H., R.; & Beck, H,. P. (2000). When the Tail Wags the Dog: Perceptions of Learning and Grade Orientation in and by Contemporary College Students and Faculty. *The Journal of Higher Education.* 71:84–102.
19. Robert, C. & Miller, R. (2013). *Best Practices in Writing Instruction*, 2nd ed., edited by Steve Graham et al. Guilford Press.
20. Stan, C. (2001) *Auto-evaluareaşievaluarea Didactică*, Ed.Presa Universitară Clujeană, Cluj-Napoca.