



Burkina Faso High School English Teaching Actors' Mastery of Assessment

Bamogo Yacouba

Université Joseph Ki-Zerbo

Dr Sanon/ Ouattara F. Emilie G.

Université Joseph Ki-Zerbo

Abstract

For any beneficial teaching, teachers must have a good mastery of assessment. This study investigates English teachers' and inspectors' mastery of assessment qualities and procedures. The data was collected in nine regions out of thirteen in Burkina Faso. It was collected from thirty-five teachers and fifteen inspectors following the principle of saturation. Using an interview guide, they were questioned on their training in testing, test validity and test writing procedure. After analysis, it came out that very few teachers got training on assessment, and inspectors were not even satisfied with that training received. Teachers and inspectors ignored the technical use of test results, washback, validity, and they followed an inappropriate procedure to design their tests. This has a detrimental impact on learning as it does not provide possibilities for remedial and does not estimate the real performance of learners. From that, suggestions were made to initiate and improve the training for teachers on assessment to attain the English teaching objective in Burkina Faso.

Key-words: Assessment; Mastery; Validity; Washback

Résumé

Pour que tout enseignement soit bénéfique, les enseignants doivent avoir une bonne maîtrise de l'évaluation. Cette étude vise à évaluer la maîtrise des qualités d'une épreuve, et la procédure d'élaboration de sujets des enseignants et inspecteurs d'anglais. Les données ont été collectées dans neuf des treize régions du Burkina Faso. Elles ont été recueillies auprès de trente-cinq enseignants et de quinze inspecteurs en suivant le principe de la saturation. À l'aide d'un guide d'entretien, ils ont été interrogés sur leur formation en évaluation, la validité d'un sujet, et de procédure de rédaction des sujets. Après analyse, il est apparu que très peu d'enseignants avaient suivi une formation en évaluation et que les inspecteurs n'étaient pas satisfaits de la leur formation reçue. Les enseignants et les inspecteurs ignorent l'utilisation technique des résultats d'une évaluation, l'effet rétroactif de l'évaluation, la validité, et ils suivent une procédure inappropriée pour concevoir leurs devoirs. Cela a un impact néfaste sur l'apprentissage car cela ne permet pas de remédier aux lacunes, et n'estime pas réellement les performances des

apprenants. À partir de là, des suggestions ont été formulées pour initier et améliorer la formation des enseignants en matière d'évaluation afin d'atteindre l'objectif de l'enseignement de l'anglais au Burkina Faso.

Mots clés : Évaluation ; Maitrise ; Validité ; Effet rétroactif

Introduction

Since 01 December 1983, it has been official through the declaration of the Minister of National Education Arts and Culture (Emmanuel Mardia Dadiouari), that in Burkina Faso, the aim of foreign language teaching is the acquisition of foreign civilisations that could enable learners to hold a good conversation with native speakers of the same level, as language and culture are inextricably linked. The only means to check the effectiveness of a curriculum is through assessment. So far, results from the assessments of the English subject tend to give satisfaction to this objective despite the obvious inability of learners to communicate effectively in English whether written or oral as posited by Some (2021). This casts doubt on the validity of the tests administered. In fact, the validity of assessment tools is questionable in the Burkinabe education system. It is even a crucial problem as Kabore the general secretary of SYNADES mentions in the newspaper Le Pays (2021) that “*j’ai suivi la réaction de certains encadreurs sur les réformes à travers les débats, mais je me rends compte qu’ils ne connaissent pas le processus d’élaboration des épreuves d’un examen*” (p.2). I witnessed the reaction of some supervisors to the reforms through the debates but I realised that they did not master the process of test design (my translation). Designing a test is more demanding than teaching on the grounds because it takes a lot of elements into account. This is true for Burkina Faso and elsewhere, and it is properly summed up by Siddiek (2010) in a research work carried out in Sudan. He said:

Although there is too much literature about tests and testing, the issue is still a highly neglected area by many language teachers and by some test designers. This inattention is due to lack of knowledge of testing techniques and lack of awareness of the importance of testing among the teachers’ community (p.133).

Consequently, students are likely to take tests of low quality from their teachers during their schooling. In such a context, it is expedient to conduct a study to find teachers' weaknesses in testing for prescriptive purposes. The aim of this study is to find out teachers’ and inspectors’ mastery of assessment.

1. Definition of key concepts

1.1. Measurement and testing

According to Brown (2018), “Measurement is the process of quantifying the observed performance of classroom learners” (p.5). It is an organized process of collecting information to make informed decisions. These decisions could be related to a student's graduation, promotion, or firing. He continues stating that a test is in simple terms a method of measuring a person’s ability, knowledge, or performance in a specific domain. The word test is sometimes used interchangeably with the word assessment or evaluation. They are so close to each other in meaning, but a test is more specific as it is about a given skill. Many theories underpin the concept of testing. The classical testing theory (CTT) according to Magno (2009) is regarded as the true score theory. The classical theory assumes that in a test, each individual has a true score which would be obtained in a test if there were no

measurement errors. For Hughes (2003), that performance would reflect the real performance of the test taker, his/her abilities, and what he/she really knows. However, due to the imperfections of measuring instruments, an individual's score may differ from his or her true score according to Awopeju and Afolabi, (2016). The items response theory (IRT) makes stronger assumptions than CTT. While CTT considers the total score of the test, IRT is focused on the score for each individual item. It is based on items analysis. A test item according to Davies (2016), is an element of the test or a question which requires a specified form of answer or response. The scores of the different items are accumulated to give the total test score.

1.2. Validity

Validity is one of the key qualities of a test. Brown (2004) argued that “By far the most complex criterion of an effective test, and arguably the most important principle is validity”. Validity is part of the assessment theory expressed by Hathcoat *et al.* (2016), who state that it is not possible to design good tests without considering the notion of validity. Once an instrument is invalid, trust in it automatically disappears. The test is then seen to be valid if results statistically highly correlate with some indices of behaviour we want them to correlate with. This approach to validity is what Weir (2005, p.18) qualifies as the “suck-it-and-see” approach. An approach to testing balanced with another called the *a priori* validity.

1.3. Reliability

It refers to the stability of test scores. That is, a test is considered to be reliable only if the test scores do not vary significantly due to the change in such external factors like day, date and place or due to the evaluation conducted by two or more sets of independent examiners, or even due to the two parallel forms of the same test. Oller (1979) considers the reliability of a test in terms of how consistently it produces similar results on different occasions under similar circumstances. Reliability is related to how consistently a test does what it is supposed to do, and is hard to be separate from validity.

1.4. Washback

In a learning context, tests are built for many purposes. Whatever the objective is, tests have an effect on the teaching and learning that specialists call washback. According to Alderson & Banerjee (2001), “The term 'washback' refers to the impact that tests have on teaching and learning” (p.214). This general term can refer to the efforts induced by tests on learners to learn their courses before and after the test. It can also be what Fahima *et al.* (2005) refer to as test consequential validity. This is the after-test scores interpretations influence on the test takers, teachers, and educational system as a whole. In fact, the nature and importance of a test incite certain behaviours of learners and teaching actors before, during and after the test. When the impact is wider and reaches the whole community, we speak of test impact. Washback is only restricted to teaching/learning and can be positive or negative.

2. Literature review

Testing is very important for society, and a lot of attention should be paid during its procedure. Test qualities (validity, reliability, practicability, and washback ...) should be considered in a test design. To ensure them, according to scholars like Hughes (1996, 2003) and Brown (2004, 2018), we usually go through three main stages during the construction process, which are the test design, drafting, and revision.

2.1. Test design level

It is about the objective or the reason for testing. At this stage, the problem is stated to clarify the kind of information we want from the test takers and why. Brown (2004) explains that test objectives involve, on the one hand, a number of simple issues about the form and function covered in a course unit, and on the other hand complex ones about constructs to be operationalized and decisions about what language abilities we want to assess. To be able to gather this type of information, Hughes (1996) proposes to look for data such as:

- The type of the test (achievement, placement, formative, summative...)
- The precise purpose, that is to clarify the objectives for testing.

2.2. Drafting your Test

A first draft of the test gives a good idea of the look of its content and how students will judge its face validity. This includes writing the test items specifications, considering the level of the answer satisfaction, and specification of the number of grades to grant depending on the level of the response satisfaction for opened questions. Specifications are information about the test like the content, the format, the duration, the critical level of performance, and the scoring. It is a set of guidelines which help to write test items. A test item is a kind of stimulus that calls for a given response. Test specifications help to achieve balance in the test content. The easiest way to have specifications that reflect the content of a syllabus is to turn the course objectives into specifications (Brown 2004). Then, we sample to get items representative of the course content, the time used to teach those contents, and the educational goals.

2.3. Revising the Drafted Test

After drafting your first test, you will work through the items you have devised to improve them. This revision is necessary to iron out some threats to validity (the layout in the face, the items map to the construct measured, how representative it is for the syllabus, and the prediction of the candidates' future performance). According to Brown (2004), there are some questions that could help to improve the quality of the test items.

- Are the directions to each section absolutely clear to avoid ambiguity?
- Are there some example items for each section?
- Is each item related to a specified objective?
- Is the language used for items simple and clear?

- Are multiple-choice item distracters not ridiculously easy?
- Are items language difficulty adequate to students' level
- Do the sum of the items and the test as a whole adequately reflect the learning objectives? This is only a sample of the criteria that could be used to improve the quality of a test item. One can refer to each type of validity criteria to go through the drafted test.

2.4. Editing the Test

In an ideal situation, you would try your test on some students before administering it. Colleagues can also help to proofread the test beforehand. Anyway, one must do his/ her best to bring to students an instrument that is, to the best of one's ability, practical, reliable, and valid. Before you edit your test, save it in a PDF format to ensure there will not be some inadequate modification at the printing stage.

2.5. The use of test results

One of the objectives of a test is to induce improvement in learning. So, after administering and scoring a test, the results can be analysed to find out the information about the candidates' performance (the overall appreciation we can give to the class performance, the students who failed and those who succeeded, in which aspect of the course there are still some difficulties or not well assimilated). A test's results can provide all this information through statistical analysis of the test performance. This is done through the analysis of the mean score, the standard deviation, the paired and independent t-tests, the share of students in the test... A deep analysis can help use test results for political reasons, like pedagogical orientation and national education planning. Each individual item will inform the teacher on the aspect of the course that was well or badly assimilated. The different information related to assessment procedures and qualities are references from which the researchers inspired to design the study tools.

3. Methodology

This study combines a qualitative and quantitative method. The mixed method is the palliative one, which helps solve the pitfalls of quantitative and qualitative studies Creswell (2009). The data was collected in nine regions out of 13 of Burkina Faso. The information on classroom testing practices was obtained from English teachers (24 males and 11 females). They are mainly second-cycle teachers and the first concerned in this research as they are test builders, administrators and graders. Among them, 29 are full-time teachers, 4 are part-time and two are trainees. They have one to 26 years of teaching experience. Regarding their study level, 15 have a bachelor, 20 of them have a Master's degree with 9 who majored in English linguistics. The second targeted population is English inspectors. They were interviewed for their opinion on teachers' knowledge of testing and teachers' formative classroom testing. The inspectors interviewed were 15 (13 males and two females). They have three to 21 years of job experience. The sampling technique to get English teachers and inspectors is snowball

sampling. The opinions were collected following the principle of saturation. However, for quantitative studies, the sample size is important as we sample for representativeness. This led to taking up to 35 teachers for statistical considerations.

To gather the data, an interview guide was addressed to teachers and inspectors of English. The interviews were conducted face-to-face and by telephone. The reason for using the phone and media is related to the distance with some interviewees and the impossibility, both financially and practically, of making more than 50 research trips while other accepted means exist. The other reason for using interviews is linked to their advantages. Kumar (2011) highlighted some, like the appropriacy for complex situations, the ability to explain the question for more understanding, and the ability to get in-depth information. The interview was mainly based on the informants' knowledge of testing (the training received on assessment and test items analysis). In addition, there were the key elements of a test quality like test validity, reliability and washback. The last section of the interview pertained to test construction procedures and their statistical knowledge of scores interpretation. For ethical considerations, research permission and consent letters were sent to informants before data collection. Then, a schedule was set up to meet the informants at their convenience. To analyse the data collected, qualitative information was grouped according to topics discussed, and then information was quantified by computation to estimate them in terms of number and the proportion they represent.

4. Results and Discussions

4.1. Training on language testing

The researchers identified training to be a critical factor that can inform teachers' knowledge on language testing. Two people confirmed they got training with the American Centre and under the French Cooperation. The others resort to what they have gained from Ecole Normale Supérieure (ENS) or during their Master's classes at university.

T1¹: "I got a training at ENS but this lesson has nothing to do with what we really do in class. It was on testing globally but not specifically in English".

T2: "There is a course on evaluation but what I don't agree with this evaluation is that, because of my experiences on language testing from my Master's course. I saw the big difference, the big gap between the course of evaluation there (at ENS) and the course of language evaluation. It was not the same. The one from the training school looks like evaluation of a project".

T3: "Yes, I got some of them. Testing in language may mean to know how people express themselves, to know how they can teach also, about language, to express themselves just for understanding of other". This is a response from a part-time teacher who pretended he got a training. It is clear from his definition that he has no clue. This response followed with the definition of testing led me to go deeper with next questions to know deeper their understanding of testing. Nine people said they did not have a training on testing.

¹ T1 : Teacher one

T4: “No, I try to evaluate my students as I can”.

T5: “No, no, no, I have never got”.

T6: “Did we get it during the Licence”? To mean “no” as he/she is a part time teacher without any training and without such a course on testing during his/her Bachelor studies.

From the answers above it is clear that teachers who said they got a training at ENS did not because the information from in depth questions showed that the class was on *Mesure et Evaluation* which is a common class given to all trainees. It is not on language testing specifically. This interpretation was obtained thanks to the quality of the research instrument. The face-to-face interview helped to inquire and understand more on the content of that training. If the tool were just a questionnaire, then all these “yes” answers would be accepted to be true. Weir (2005) posits that validity is not in the claim we are making, but the evidence brought to support that claim. Definitely those who can be considered to have had a language testing training are those who did it in the Master’s programme in linguistics and the two other teachers who got it from the American centre and the French cooperation. The confrontation of the teachers’ information with inspectors’ apprehension of the training confirms that the training is inadequate. fourteen inspectors, which corresponds to 93.33% of them were uncomfortable with the teachers’ training at ENS on testing. Siddeik (2010) already raised this concern in Sudan, saying that testing is neglected in the education system. From this light, we can foresee that teachers will not have a good mastery of testing skills.

4.2. Mastery of test quality

Having training does not ensure that one knows how to devise a valid test in as much as there are doubts about the training provided to teachers at ENS. Some more sub-questions were devised to ensure the teachers know what testing is, the test items analysis, washback, validity, and reliability. Mastering these elements will attest that the respondent can write a good quality test. The following answers were obtained:

- Only two teachers had an idea of the statistical interpretation of a test’s score. Both of them did a Master’s in linguistics.
- T1 said: “We talked about this in class of testing and assessment with Dr X”. All the remaining thirty-three teachers (94.28%) just said they had no idea on the statistical interpretation of tests results.
- One inspector had an understanding of items analysis but not of items discrimination. Insp1²: “Items analysis is referred to the answers students give to different individual items so as to have an idea of the items themselves and of the whole test itself”.
- Washback

Three inspectors out of seven proved to have an understanding of the term washback

² Insp1 : Inspector one

Insp1: “test washback can be referred to the impacts or the inference of tests on teaching and learning. Four respondents simply said they ignore what it is about. As for teachers, they do not master washback. This means that in the formative assessments and exam tests, teachers will not look at the effect that their test can have on fostering or harming the teaching and learning. Washback is the most important thing we need to consider in test design in relation to learning. When it is negative, then the test is negative towards learning.

- Validity

Six inspectors have a good understanding of validity. However, when asked to give examples of validity, all of them do not know. This is a sample of responses:

Insp1: “Validity for me it is this aspect of test which should make the test evaluate what it is supposed to evaluate”.

Insp2: “Your test is valid if it measures exactly what you have taught to your students. »

Insp3: “Validity, let's say, I don't know what to tell you, but I don't know if you want to test or not when for example you did a test you certainly want to know if, let's say, what the students want to be tested on correspond to what you expected them to do..... This is what I understand by validity”. This is the answer of teachers’ supervisor taken word for word.

Insp4: “Validity is related to the fact that if we take a test, we should design it in such a way that when we test it in one situation, and whatever the others situations, the test would give the same results. This is what makes it valid”.

From the information above, it can be concluded that teachers and inspectors have a limited knowledge of the technical aspects of testing. Statistical analysis of tests, like items analysis, discrimination is not mastered. Inspectors have a good understanding of testing in general, but not on statistical analysis of test results. During their comments on the importance of testing, none of the inspectors nor teachers tried to associate testing with the teaching policy. This means that they ignore that the attainment of English teaching goals in Burkina Faso is measured by tests. Siddeik (2010) explained it, saying that “test results are benchmarks and tools of quality control with which decision-makers can secure the national educational goal” (p.135). Unless we keep in mind that the tests we administer show our success or failure to contribute to achieving the national education goal, we will not induce improvement in our practices and students’ performance. Weir (2005) explained that feedback from tests is essential for the teachers, who are responsible for meeting the instructional objectives, while for the students, it illustrates their performance. This is, in fact, a betrayal of our oath to contribute to quality education and the country's development if we test students using a flawed framework. The inability of inspectors to have a sound item analysis, discrimination, and even to analyse statistically test results will harm teaching practices.

4.3. Mastery of test writing procedure

To confirm the respondent's knowledge of validity, the question was openly asked to some, and inferences were made from the procedure they go through to design their tests. The procedure followed by teachers to design their tests is shallow. It is, however, fair for teachers who did a master's in linguistics (nine people). Their procedure considers the content and some aspects of face validity like mistakes and text suitability with students' level. None had criteria for determining a text difficulty level when selecting texts according to students' level. Let us consider some of the responses on the procedure they follow to write their tests:

T1: "Before I design any test, I have to take into consideration certain criteria among which for example we have the content and the form"

T4: "Ok, the different steps? I take the text, and I ask different questions... comprehension questions. I start with comprehension questions, after that I give them a writing assignment and sometimes, translation". These answers respect some procedures given in the literature by Hughes (2003), Brown (2004), and Weir (2005). It corresponds to nine teachers (26.47%), and the number of those who did a Master's degree in linguistics. None of the others follow an acceptable procedure to get a valid test. This contradicts Hughes's (2003) recommendations, which posits that teachers should follow a suitable test design procedure to lessen threats to validity.

Nine inspectors were questioned on the same issue. Two responses were considered fair, given that they considered some aspects of face and content validity. The procedure given by inspectors does not consider construct and predictive validity. This backs Backman (1995), who assumes that the non-mastery of the testing procedure is a threat to test validity. It naturally leads to grading validity issues as items construct quality is critical to systematic grading. Inspectors' limited knowledge about test construction comforts teachers in the status of ignorant in testing procedure.

Another way to improve the validity of a test is the revision done after writing to iron out the possible weaknesses. The following are answers related to tests revision by teachers:

T1: "I go through for mistakes I can correct and sometimes I may try to answer the different questions before I submit it to students"

T2: "I can ask colleagues to discuss the answers and try to see if the test is not too hard for the pupils". This corroborates with Hughes (2003) who suggests to moderate test items by asking at least two colleagues to revise it.

T3: "Have I ever looked for that? I'm not sure"

Some teachers try to proofread their tests before implementation even though they do not have some validity fixed criteria to follow during the proofreading to ascertain they are respected. Others also ignore proofreading their tests. Not revising tests is dangerous and harmful to validity as some teachers do not care about the tests they produce. This situation entails that one key validation process is skipped. That should not be the case because a test designer's first concern is ensuring the quality of the test items, as recommended by Awopeju and Afolabi

(2016). Not revising tests will result in weaknesses in tests' face validity, which could be easily avoided and other forms of validity are also exposed. It is a logical result of the non-mastery of test design processes and a source of assessment weaknesses. To solve that, Hughes (2003) proposes to design a checklist for each test rubric for items moderation.

4.4. The language skills tested

To ensure content validity, the skills and constructs included in the classroom assessment should be taught in class. Thirty-four teachers do not teach the reading comprehension techniques for the reading comprehension skill. Only one respondent mentioned that he teaches his students scanning and skimming. Inspectors said they did not notice during their class visits teachers teaching reading comprehension strategies and techniques. They have reading comprehension classes with students, but no lesson is designed to teach the strategies and techniques. As for translation, teachers admitted they do not teach translation techniques to students. No inspectors admitted having met a teacher teaching translation techniques. They are all (100%) unanimous on that. They gave the following comments. The last tested skill is essay writing. The writing skills were not also taught. They learn how to write through practice. No one mentioned to be teaching the writing skills like organisation, clarity, word choice, cohesion, coherence and mechanics. They did not also say that they teach persuasive, argumentative, descriptive writing or writing correspondence, which raised the issue of construct underrepresentation. Teachers' answers were confirmed by inspectors as only one of them said they were teaching writing. However, he added that it is « not very often ». The other eleven inspectors said they overlooked that. On the whole, teachers do not design a lesson to teach the techniques of different skills tested. They do exercises instead of teaching the construct of each tested skill. Normally, the students should be taught the skills tested as suggested by Hughes (2003), so that they understand what it entails. We cannot test something we do not teach. These kinds of tests violate the principles of content validity. When you analyse the information on testing practices, teachers' understanding and test design, it implies that the classroom formative tests are of bad quality.

4.5. Teachers' use of old tests

Knowing that each test is context-dependent and a valid test for one situation is not compulsorily valid for another, teachers were asked to know their test creativity. All teachers interviewed said they use old tests. Some said they tried to make some modifications before, but it still remains an old test. Other teachers even admit they have test bunches where they go and pick some to administer. The following is a sample of responses.

T1: "I use tests I have been given by my former teacher in 1997"

T2: "I sometimes use old text but I try to change the questions"

Old test use from the light of these responses is a common practice for teachers, whether young or old, from villages or towns. A teacher witnessed that "Many teachers, when they have a test, they go and pick up old tests. After they have difficulties to correct them". This situation is acceptable in our context as, so far, all data have

indicated that teachers were not trained in language testing. Some will call it laziness, but it can be justified by the lack of training and limited knowledge on testing as one respondent said, “I try to evaluate my students as I can”. Having a collection of old tests becomes a solution, and this does not encourage learning how to design a test and tests creativity. If we consider the teacher who said he is still using tests given to him by his teacher since 1997, it shows that we are lagging in the domain of testing. The use of old tests is a source of threats to validity and creates troubles in the validity of assessments. Washback on learning in such testing practices is high.

To sum up, teachers test their learners as they can, and negative washback on tests of these kinds is high. Those tests cannot inform much about teaching, learning, students’ level, the achievement of English teaching goals. However, Siddeik (2010) posits that each test is a learning situation. In this study, the consideration of testing helps to understand why people learn English for many years and are still weak at English. In fact, tests are used to inform teaching, like changing the instruments, the method, or returning to the aspects of the course students did not understand for remediation. Teachers ignore this. Then, tests are used to give the students' actual level so that they can work on the aspect of the learning that is not processed yet. The test designers’ mastery of assessment in Burkina Faso cannot help to assess the real performance of the learners. Then, this gives a wrong appreciation of students’ marks by students, the teachers, the school, and the parents. Considering the importance of a test as a quality control tool, that is, to see if the syllabus's role to push learners up to a given level is achieved, it can be said that the test design procedure could not certify tests’ quality. So, students' marks in class will not reflect their actual level. Teachers and inspectors’ ignorance of the use of tests results makes that automatically, one cannot tell if the syllabus attained its objectives or not. With such assessments, the conclusion is simply that students cannot communicate and keep changing teaching methods. This fact illustrates that English teaching actors' understanding of testing is weak, and it clarifies that testing continues to be neglected in Burkina Faso as mentioned by Zabramba (2000). Evaluation is everywhere in life.

Conclusion

The study aimed to assess English teaching actors’ mastery of assessment. After interviewing them, it came out that their understanding of testing, its elements and procedures is weak. All the respondents, whether inspectors or teachers, could not respond satisfactorily to all the questions. They do not master the technical issues related to assessment, such as item analysis and statistical interpretation of test results. The key elements of test quality, like validity, reliability and washback, are not mastered. This led to the use of unsatisfactory test design procedures, which do not capture complete content, construct, face and predictive validity. They prefer using old tests for assessing their learners. In a word, the English teaching actors have a weak mastery of assessment. The detrimental washback for the lack of mastery is high as it leads to little gain in learning. From the weakness revealed, the following are suggested for testing policy and classroom practices.

First, we can set up a testing institution with a clear testing policy. This will call for training specialists, who will train teachers and design documents. For that, at the beginning of the school year or during holidays, inspectors and specialists in testing can organise seminars and conferences so that the English teachers can meet

to level up their knowledge on language testing. Documents can serve as guidelines on testing for teachers. These documents should be embedded in the philosophy of the objectives of language teaching in Burkina. Second, to empower testing policy, checking formative tests must be part of supervising objectives. Third, the teachers' training programme at ENS on testing should be revised. The dissatisfaction of teachers and inspectors' criticism of the training given at ENS on testing justifies the need for change in the training provided. Finally, cooperation with colleagues is to be promoted. What can also be done for teachers is to set a mechanism to check tests quality before their use. Instead of letting each teacher test his/her students, they can try to make a common test for 2, 3 or 4 classes so they would pay more attention to the testing. They should have the same progression so that the designed tests carry the language item already studied by every school member. In addition, the school vice-principal has to be trained to check from time to time the record book to see if the assigned progression and the contents are taught before the testing takes place. Teachers can try the tests themselves before.

References

- Alderson, J. C. & Banerjee, J. (2001). *Language testing and assessment* (Part I). Cambridge University Press
- Awopeju, O. A. & Afolabi, E. R. I. (2016). Comparative analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of senior school certificate Mathematics examination. *European Scientific Journal*, vol.12, No.28, pp. 263-284. ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431
- Backman L. F. (1995). *Fundamental considerations in language testing*. Oxford University Press
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger Postgrad Med J* 2015; 22:195-201.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson
- Brown, H. D. & Priyanvada A. (2018). *Language Assessment Principles and Classroom Practices*; third edition, Pearson.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Inc.
- Davies, A. (2016). *A glossary of applied linguistics*. Routledge
- Educational Testing Service (2009). *Guidelines for the assessment of English language learners*. Educational Testing Service (ETS).
- Fahima, M. B. Saidatul A. Z. A. & Asiah J. (2005). A Validation process of ESP testing Using Weir's socio-cognitive framework. *Procedia - Social and Behavioral Sciences* 202 pp. 199 – 208
- Hughes, A., (1996). *Testing for language teachers*. Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press; 2nd ed.
- Hathcoat, D. J, Sanders B. Courtney & Gregg N. (2016). *Selecting and designing instruments: Item development, development, reliability, and validity*. James Madison University.
- Kumar R. (2011). *Research methodology: a step-by-step guide for beginners*. 3rd edition, Sage.
- Le Pays n° 7318 du jeudi 29 avril 2021

- Magno C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment* April 2009, Vol. 1, Issue 1, pp. 1-11
- Oller, J. W. (1979). *Language tests at school*. Longman Group Ltd.
- Richard J. M. & Sheila C. M. (1999). *Item Analysis for Criterion Referenced Tests*. Research Foundation of SUNY/Center for Development of Human Services.
- Siddiek, G. A. (2010). The Impact of test content validity on language teaching and learning; *Canadian Center of Science and Education* pp.133-143
- Some, G. E. (2021). Implementing communicative approach in English language teaching: the case of Burkina Faso. *International journal of English studies*, pp.52-56.
- Sugianto, A. (2011). *Analysis of validity and reliability of English formative tests*. *Journal on English as a Foreign Language*, Volume 1, Number 2, pp.87-94.
- Sugianto, A. (2016). Technique and procedure of statistical analysis of test item validity of objective English language test. Proceedings of the 2nd National Conference on English Language Teaching (NACELT) IAIN Palangka Raya. *ResearchGate* (p.p 5-18) ISSN: 2502-3225
- Weir, J. C. (2005). *Language testing and validation an evidence-based approach*. Centre for Research in Testing, Evaluation and Curriculum (CRTEC) Roehampton University
- Zabramba, J. (2000). *Formative and summative assessment/ evaluation*. The Ohio State University.

