

---

**FRENCH LANGUAGE DIAGNOSTIC WRITING SKILL TEST FOR JUNIOR  
SECONDARY SCHOOL STUDENTS: CONSTRUCTION AND VALIDATION USING  
ITEM RESPONSE THEORY**

**Nkechipatricia-mary Esomonu and Ogechukwu Juliet Okeke**

Department of Educational Foundations  
NnamdiAzikiwe University, Awka, Nigeria

DOI: <http://dx.doi.org/10.37500/IJESSR.2021.4227>

**ABSTRACT**

The study developed and validated a French language diagnostic writing skill-based test for junior secondary school students using Item Response Theory (IRT). The diagnostic test consisted of 50 multiple choice test items which were constructed using the French language curriculum for junior secondary school students. Five research questions were formulated to guide the study. Preliminary validation was done by two French teachers and three experienced lecturers in French language. The pilot testing was conducted to ensure grammatical checking and compatibility of the test items while trial testing was conducted on 500 junior secondary school students and this yielded the data for item analysis. Analysis was done through the BILOG-MG using Item Response Theory. The data generated from the study were analyzed using smallest likelihood estimation. The empirical reliability of the test score was 0.77. The difficulty and discrimination level of the French language diagnostic writing skill-based test items indicated that 42 items were retained while 8 items were modified. The test was found to be of good quality, valid and reliable. The French language diagnostic writing skill-based test is ready to assess examinees' writing skill in French language for JSS in Nigeria.

**KEYWORDS:** Diagnostic test, Item Response Theory, French Language test, Validity and Reliability.

**INTRODUCTION**

The multi-lingual and multi-cultural nature of Nigerian polity and the absence of national unifying indigenous language brought about foreign languages such as English and French languages as a medium of intra-national and inter-national communication (Fakeye, 2006). French language is a language of communication in twenty-six African countries. Furthermore, it is one of the core subjects for junior secondary school students and also an elective subject for senior secondary school students in Nigeria. Its curriculum aims at the development of writing skill (FME, 2014). The benefits of teaching and learning French language can be realized if the students who are taught French language know and also understand the basic needs and skills of the language.

The introduction of French language in schools is to enable Nigerians effectively interact locally and internationally with francophone countries, attend international seminars and conferences without much problem (Anneduke, 2009). Nigeria is surrounded by Francophone countries and this has made French language a necessity for Nigerians in order to have peaceful co-existence with the people in

neighbouring countries (Borode, 2014). Therefore, French was made compulsory in primary and junior secondary schools to enable students effectively use French language (FRN, 2014).

French language has been rated as the second world's most popular language which is characterized by active communication from one continent to another. Despite the importance of French language, some factors threaten the effective use of the language in Nigeria. There is the problem of decline in enrolment caused by lack of interest and poor performance of students in French language at public examinations (Examiner's Report, 2012, 2013, 2014). Another problem may be not trying to determine the root causes of students' poor language skills and poor performance in French language. The instrument that can be used for this is diagnostic test but there is lack of adequate research that focused on diagnostic test in language assessment as earlier reported by Alderson (2005). Students' poor language skills and performance in French language at both junior and senior school certificate examinations which have persisted over the years, brought about the need to train students on writing skills.

Writing skill is the ability to think and evolve ideas in written form into a grammatically error free sentence. The most common sub-skills in language writing skills, according to Okafor (2015) are organization, grammar and blank-filling. The organization sub-skills are aimed at finding the sequential presentation of ideas in writing while grammar and blank-filling sub-skills are aimed at training students to construct missing words by teaching them to read around (before and after) the gap. These sub-skills are tested by giving the students text with some key words deleted. Students are to choose correct word for each space, blank out nouns or verbs or adjectives or adverbs (Jang, 2005). These enable the test giver to test the students' understanding in grammatical relationships of words or phrases across a text. Authors like Hadfield 2008, Harmer 2007 and Okafor 2015 agreed that these sub-skills are used together; in some cases, they are interwoven and interchangeable. This means that testing for grammar skills cannot be done independently in a multiple-choice question type.

Since the introduction of French language in Nigerian schools, there have been essential ways of testing and evaluating French students in classroom and in public examinations for these various skills. The essence of using tests and other evaluation instruments during instructional process is to guide, direct and monitor students' learning and progress towards attainment of course objectives (Alonge, 2004; Kolawola, 2010). However, within the classroom interaction, teaching and assessing of student's performance are done using different tests with different functions.

It is vital to make a clear distinction among the tests being used within the classroom: proficiency test, formative test, achievement test and diagnostic test. The proficiency test is one that measures a candidate's overall ability in a language; it is not related to a specific course or area. The formative test is a continuous feedback on learners and learning difficulties for remediation during course or program development. The achievement test, on the other hand, tests the students' knowledge of materials that have been taught on a course. The content is derived from specific curriculum or course

of study. The assessment tool for determining causes of persistent learning problems of students for necessary interventions is called diagnostic test. Therefore, attention should be given to content area and skills while constructing a diagnostic test (Esomonu & Eleje, 2020).

A diagnostic test such as diagnostic skill-based test needs to be more specific and focused in order to determine students' level of functioning (Alderson, 2005). Test development and validation of a diagnostic test involves steps and stages of construction (Alderson, 2005; Eleje, Esomonu, Agu, Okoye, Obasi, & Onah, 2016). The revised procedural framework for the development and validation of diagnostic tests by Zhongbao (2013) will be used in the process of instrument development and validation in the work because of its detailed steps in developing language test. This involves phase 1: need analysis; phase 2: test design and operationalization; phase 3: test piloting administration and validation; phase 4: test Impact. In administering diagnostic test to students, it is expected to see the areas where there are weakness and reasons for that. Then the teacher has to play the role of a doctor at this stage, to prescribe possible solutions and intervention. Based on this statement, it is obvious that there must be feedback and remedial classes for such students who are found to be weak in some areas. The diagnostic tests have long been recognized as essential for teaching and learning because of its feedback.

However, one issue that deserves consideration in diagnostic test development is how valid the diagnostic tests are especially to give a vivid feedback. Information gotten from diagnostic test is determined through item analysis of students' individual responses to each item. The IRT analysis is a suitable tool used in measurement of examinee's ability, selection of test items and for equating tests. According to Adedoyin and Mokobi (2013) is a process which examines students' responses to individual test items in order to assess the quality of those items and of the test as a whole.

Despite the importance of using IRT in determining the quality of test items and diagnostic tests to improve the quality of learning, teaching and assessment, there seems to be limited number of tests available to the public that serve these purposes (Alderson, 2007). Esomonu and Eleje (2017) carried out a research that developed and validated a diagnostic quantitative economics skill test for secondary schools using IRT in Nigeria. Also, Ani (2014) carried out a study on IRT in the development and validation of multiple-choice test in Economics. Adedoyin and Mokobi (2013) on IRT psychometric analysis in examining the quality of junior certificate (JC) mathematics multiple-choice examination test items.

Thus, there is limited diagnostic skill test available and to the best of the researchers' knowledge there is no empirical study done on French language diagnostic writing skill using IRT. The steps and stages involved in construction of diagnostic skill test posed a huge problem to the classroom teacher. This made the researchers to be concerned and motivated to fill this gap by developing a workable multiple-choice French language diagnostic writing skill-based test for junior secondary school students using IRT.

The objective of this study therefore, is to develop and validate a French language Diagnostic Writing Skill Test (FDWST) for junior secondary school students using IRT. Based on the objective stated above, the following research questions were answered in the study:

- 1) How valid is the French language Diagnostic Writing Skill Test (FDWST) instrument?
- 2) What are the difficulty parameters of the FDWST instrument?
- 3) What are the discrimination parameters of the FDWST instrument?
- 4) What are the guessing values of the FDWST instrument?
- 5) How reliable is the FDWST?

## LITERATURE REVIEW

### Diagnostic Test

Diagnostic test is an assessment tool for determining causes of persistent learning problems of students for interventions and further remedial studies. Diagnostic test is an in-depth test used to identify students' strengths and weakness in a specific area by testing what the student knows or do not know in a language or skills they have or do not have. Diagnostic test is scored using true test score criteria. That is to say they are not norm-referenced. Sidhu (2015) confirm that diagnostic tests have much in common with criterion-referenced tests. This means that diagnostic test elicits information about an individual's performance in highly specific skills and relate this information to instructional prescriptions. Feedback is a significant feature of a diagnostic test. The diagnostic test feedback should concentrate on students' weakness and difficulties more than strengths, better if coupled with detailed descriptions about better performance (Wiggins, 1998). The feedback should be made available to administrators, teachers and students with valuable information to support their decision-making.

In reference to the usage of diagnostic tests, difference studies have found the effective use of diagnostic tests in improving teaching and learning (Esomonu & Eleje, 2017; Jang, 2009; Oyekan, 2013; Patel, 2012; Ute, 2007). However, the scarcity of French language diagnostic writing skill test necessitated this research. Therefore, making French language diagnostic writing skill test available to teachers in schools to improve teaching, learning and student's performance in French language.

### Item Response Theory

Item Response Theory is also called latent trait theory or true score theory. The Item Response Theory is the study of test and item scores based on assumptions concerning the mathematical relationship between abilities and item responses. IRT is used in developing and refining tests and examinations, maintaining banks of items for examinations and comparisons between results over time. The IRT has the possibility of obtaining item characteristics which are not group dependent; ability scores, which are not test dependent; and a measure of precision for each ability level. Under IRT, item difficulty describes where an item functions along the ability scale (Baker, 2001). It allows item difficulty to be estimated in an unbiased way. Item discrimination in IRT is the correlation between the item and test

performance (-1.00 to 1.00). In distracter, quality can alter the performance on a test item. It allows test writer to identify poor performing items and revise or omit them. They are instructional intent, specifying the domain, item development, item review and test development.

IRT attempts to model the ability of an examinee and the probability of answering a test item correctly based on the pattern of responses to the items that constitute a test. Using the appropriate IRT model, the ability level of an examinee is accurately estimated with any set of items that measures this ability. The IRT is mostly used for modeling responses to items and scoring of educational tests. IRT is based on the idea that the probability of a correct response to an item is called latent trait or ability.

The number of item parameters to be estimated determines which IRT statistical model will be used, and the test item analysis of any examination is based on item discrimination, item difficulty and the guessing parameters. There are three basic models in item response theory. They are:

**(1) b parameter:** it is also known as the difficulty parameter or the threshold parameter. This parameter tells us how easy or difficult an item is. It is used in the one parameter (1P) IRT model. The more difficult an item is, the higher an examinee's ability must be in order to answer the item correctly. Items indicating high b values are difficult items, that is to say, values of b greater than 2 are very difficult items and low ability examinees are not likely to answer them correctly. When the values of b are between -2 to +2, then the test items are considered to have a good difficulty parameter (Baker, 2001).

**(2) a parameter:** It is also called the discrimination parameter. This value tells us how effectively an item can discriminate between highly proficient students and less proficient students. The two-parameter (2P) IRT model uses both a and b parameters. According to Adedoyin and Mokobi (2013), one characteristic of a good test item is that high ability candidates will answer it correctly more frequently than lower ability candidates. The parameter expresses how well an item can differentiate between examinees with different ability levels. A test item has positive discrimination when higher ability students have a high probability of answering an item correctly and lower ability students have a low probability of answering the item correctly. The discrimination values (a-values) of good items range from 0.35 to 1.69 (Esomonu & Eleje, 2017).

**(3) c parameter:** it is also known as the G parameter or the guessing value/ parameter. This value tells us how likely the examinees are to obtain the correct answer by guessing. According to Adedoyin and Mokobi (2013), the c parameter expresses the likelihood that an examinee with very low ability will be able to guess the correct response to an item and therefore has a greater-than-zero probability of answering correctly. This parameter is the lowest value. For example, an examinee who randomly selects one response from others can answer these items correctly about, 1 out of 4 times. That is to say, the probability of guessing correctly is about 0.25.

Another purpose of IRT is to provide a framework for evaluating how well assessments work, and how well each item on assessment works. According to Abanobi (2013), the most common application of IRT is in education, where psychometricians use it for developing and refining examinations, maintaining banks of items for exams, and equating for difficulties of successive versions of exams (for example, to allow comparisons between results over time).

According to Adedoyin and Mokobi (2013), the IRT model assumes that the performance of an examinee can be completely predicted or explained from one or more abilities. IRT models the probability of a correct answer using three logistic functions. The one-parameter logistic (1PL) model attempts to address the probability of a correct answer by allowing each question to have an independent difficulty variable. For instance, one-parameter model allows each question on a test to have an independent difficulty variable. The two-parameter logistic (2PL) model attempts to model each item's level of discrimination between high and low ability students while 3PL model adds a third item parameter which is called guessing parameter that reflects the probability that an examinee with a very low trait level will correctly answer an item solely by guessing. This implies that students can correctly answer an item in a test by guessing. The Item Response Theory is related to the study because the study is developing a French language diagnostic skill-based test that intends to find the parameters of the items to be constructed and validated in the study.

## **METHOD**

The study used instrumentation research design. The population of the study comprised 2,426 Junior Secondary School students. The sample for the tryout consisted of 500 students drawn through simple random sampling from 20 schools from government-owned junior secondary schools in Awka Education Zone.

The phases of diagnostic tests development used are that of Zhongbao (2013) that involved (1) needs analysis (2) test design operationization (developing test specifications, selecting test tasks, and designing feedback) (3) pilot testing and administration and validation (trial testing, test administration and evaluation).

The Junior Secondary School curriculum for French Language, Examiners' Reports of previous years as well as prescribed textbooks and some teaching materials, examination scripts and documents were reviewed and analyzed to ascertain the inclusion of the targeted topics, objectives, skills, students' learning needs and difficulties on which the diagnostic skill-based test were based. The skill-based content areas cut across French verbs and conjugation, genders, use of articles, negation, structure and grammatical aspect of the language. The stems and options were provided and only one of the five options was the key (the answer). The tables of specification of the test consisted of 50 items.

The preliminary validation of the test was done by presenting initial draft of 50 skill-based items of the instrument to two French teachers and three experienced French lecturers. The test-retest reliability

method was used to ensure the initial stability of the instrument. The reliability index of 0.82 was obtained.

The researchers administered the test to 500 junior secondary school French students of the sampled schools through the principals and French language teachers of the schools involved. The scripts gotten from the administration of the test were scored by allocating a single mark for a correct response and no mark for a wrong response or a choice of more than one response for an item. The total correct score was determined, and the percentage of the score, out of the possible scores, was calculated.

The final validation version of the instrument was done, using the smallest -2Log likelihood value which indicated the model best fit (Thorpe and Favia, 2012). Therefore, the 2PL model fits best for the French language Diagnostic Writing Skill Test (FDWST). The final version of the test was administered to 500 French junior secondary school students and the final reliability was established using the Item Response Theory.

The statistical analysis package used for the diagnostic skill-based test item is the Item Response Theory analysis software package. The IRT software package was used to determine the item discrimination parameter, item difficulty parameter and the guessing parameter of the test items. Analyses of the difficulty parameter of these items were calculated and items that had difficulty parameter between 0.35 to 1.69 were selected and considered appropriate for inclusion in the instrument (Baker, 2001; Esomonu & Eleje, 2017).

**Table 1: Interpretation of Values for Difficulty Parameter; from Baker (2001, p.34)**

Less than -2	Very Easy
-0.50 to -2.00	Easy
-0.49 to 0.49	Average
0.50 to 2.00	Difficult
Greater than 2.00	Very Difficult

According to table 2 below, an ideal discrimination parameter should range from -2 to +2. This implies that high discrimination level indicates that the item discriminates well between low and high skilled individuals. The decision rule is that discrimination parameter that range from -2 to +2 should be selected.

**Table 2: Interpretation of Values for Discrimination Parameter; from Baker (2001, p.34)**

0.01 – 0.34	Low
0.35 – 1.34	Moderate
1.35 - 2.00	High
2.01 and above	Very High

**Table 3: Decision table for difficulty and discrimination indices**

Difficulty Level	Discrimination Level	Decision
Easy	Low	Revise
	Moderate	Retain
	High	Retain
Average	Low	Revise
	Moderate	Retain
	High	Retain
Difficult	Low	Revise
	Moderate	Retain
	High	Retain

The final selection of each item was done using the interpretation in table 3 showing the difficulty level, discrimination level and decision taken for an item to be selected or modified or revised. In order to preserve the content, skill or area of the test and retain the total number of items so that no learning point was omitted, all items that were rejected were revised (Ceniza & Cereno 2012). The distracters that no one selected and items that were out of range were modified.

**RESULT**

**Research Question 1**

How valid is the French language diagnostic skill test?

**Table 4: Model fit information for French Language Diagnostic Writing Skill Test (DFWST)**

DFWST	
-2Log likelihood <sub>1-PL</sub>	24064.058
-2Log likelihood <sub>2-PL</sub>	22525.689
-2 Log likelihood <sub>3-PL</sub>	23577.385

The model with the lowest -2log likelihood is 2 parameter logistic model (2PL), which is the best fit.

**Research Questions 2, 3 and 4**

What is the item threshold, slope and guessing values/levels of FDWST items based on two parameters logistic (2PL) model?

**Table 5: Item threshold (difficulty estimates) and slope (discrimination) values of FDWST items**

Item	Threshold (Difficulty)	Difficulty Level	Slope (Discrimination)	Discrimination Level	Remark
1	0.53	Difficult	0.31	Moderate	Retain
2	0.43	Average	0.38	Moderate	Retain
3	0.49	Average	0.32	Moderate	Retain
4	1.21	Difficult	1.16	Moderate	Retain
5	0.32	Average	1.44	High	Retain
6	1.27	Difficult	1.18	Moderate	Retain
7	0.93	Difficult	0.45	Moderate	Retain
8	1.70	Difficult	0.26	Low	Revise
9	0.56	Difficult	0.45	Moderate	Retain
10	0.38	Average	0.34	Moderate	Retain
11	0.38	Average	0.36	Moderate	Retain
12	0.56	Difficult	0.69	Moderate	Retain
13	0.68	Difficult	1.32	Moderate	Retain
14	1.93	Difficult	1.43	High	Retain
15	0.68	Difficult	0.74	Moderate	Retain
16	0.31	Average	0.48	Moderate	Retain
17	0.27	Average	0.98	Moderate	Retain
18	0.20	Average	0.61	Moderate	Retain
19	2.18	Very Difficult	0.67	Moderate	Revise
20	0.63	Difficult	0.41	Moderate	Retain
21	0.48	Average	0.77	Moderate	Retain
22	0.36	Average	1.34	Moderate	Retain
23	1.59	Difficult	0.99	Moderate	Retain
24	1.80	Difficult	0.77	Moderate	Retain
25	2.40	Very Difficult	0.55	Moderate	Revise
26	0.86	Difficult	1.13	Moderate	Retain
27	0.43	Average	0.54	Moderate	Retain
28	0.41	Average	0.50	Moderate	Retain
29	0.93	Difficult	0.48	Moderate	Retain
30	2.49	Very Difficult	0.58	Moderate	Retain
31	1.15	Difficult	1.43	High	Retain

32	1.85	Difficult	0.32	Moderate	Retain
33	0.71	Difficult	1.24	Moderate	Retain
34	2.31	Very Difficult	0.35	Moderate	Revise
35	1.31	Difficult	0.23	Low	Revise
36	3.36	Very Difficult	0.40	Moderate	Revise
37	1.73	Difficult	1.37	High	Retain
38	0.95	Difficult	0.67	Moderate	Retain
39	1.11	Difficult	0.15	Low	Revise
40	0.91	Difficult	0.51	Moderate	Retain
41	0.92	Difficult	1.42	High	Retain
42	0.49	Difficult	0.41	Moderate	Retain
43	0.63	Difficult	0.34	Moderate	Retain
44	0.89	Difficult	0.41	Moderate	Retain
45	0.27	Average	1.37	High	Retain
46	0.42	Average	0.58	Moderate	Retain
47	0.34	Average	1.47	High	Retain
48	0.82	Difficult	1.32	Moderate	Retain
49	1.31	Difficult	0.54	Moderate	Retain
50	1.88	Difficult	0.56	Moderate	Retain

In Table 5 the 2-parameter logistic model (2PL) has the smallest -2Log likelihood. Therefore the 2-parameter logistic model that consist of discrimination parameter and difficulty parameter were used in analysis. The item threshold (difficulty) and item slope (discrimination) values indicated that forty (42) items be retained while ten (8) items (8, 19, 25, 29, 34, 35, 36, 39) to be revised.

**Research Question 5**

How reliable is the entire FDWST according to IRT model?

**Table 7: Summary of Statistics from Item Analysis Results**

DFWST	
N	500
Mean	-0.0341
Standard Deviation	0.8857
Variance	0.7845
Empirical Reliability	0.7710

The empirical reliability index as given by IRT model (Bilog MG) is 0.77. This is the Kuder-Richardson formula 20 (KR-20) reliability index.

## Discussion of Findings

### Analysis of French Language Diagnostic Writing Skill-Based Test

The model fit information indicated that -2Log likelihood value (2PL) indicated the model best fit.

### Item Threshold (Difficulty) Values of FDWST

It was revealed in table 5 that items 1,2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 26, 27, 28, 30, 31, 32, 33, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47,48,49 and 50, that is forty-two (42) items or eighty percent (80%) of the DFWST were within the b-value range of -2 to +2, while items 1, 2, 8, 19, 25, 29, 34, 35, 36 and 39 were outside the -2 to +2 range and thus were revised. The findings of the study are in line with those of Ani (2014) who pointed out that difficulty parameter or the threshold parameter value tells us how easy or how difficult an item is. The selection of items ranging from -2 to +2 corresponds with the criteria stated by Baker 2001.

### Item Slope (Discrimination) Values of FDWST

In table 5 of FDWST, forty (40) items had Moderate discriminating values, three (3) items indicated Low discriminating values, and seven (7) items indicated High discriminating value. The discriminating parameter revealed how well an item discriminate between respondents below and above the item threshold parameter as indicated by the slope of the characteristic curves (Reeve & Fayers, 2005). The selection of items was based on interpretation of discrimination values in table 3 which is in agreement with the criteria description of Baker (2001).It was also revealed that the threshold (difficulty level) and item slope (discrimination level) values indicated that for FDWST, forty-two (42) items were retained.

### The Reliability of FDWST

The summary of results in table 7 shows that the empirical reliability of FDWST was 0.77. This can be interpreted that 77% of the consistency of the FDWST in yielding approximately same result repeatedly (Cherry, 2005). This shows that the test were reliable. According to Ceniza and Cereno (2012), the reliability coefficient within the range of 0.81 to 1.0 indicated high reliability; 0.61 to 0.80 signified a moderate reliability; also 0.41 to 0.60 signified fair reliability; 0.10 to 0.40 signified slight reliability; and less than 0.10 signified no reliability.

## CONCLUSIONS

From the findings it was revealed that the French language Diagnostic Skill-Based Test that was developed is valid. The 2 parameter logistic model is the best fit. The empirical reliability was 77%. This implies that the DFWST is reliable.

## RECOMMENDATIONS

Based on the findings of the study, it was recommended that teachers should be encouraged to use French language skill-based test for identification of students' weak areas for remediation and follow-

up studies. It was recommended that the examination bodies and teachers should be encouraged to use IRT in developing test items. Therefore, examination bodies should come up with regular training, workshops and seminars for item writers, and supervisors and examiners to update them on test construction.

## REFERENCES

- Abanobi, C.C. (2013). Psychometric properties of NABTEB Economics multiple choice test items from 2005 - 2009. Unpublished Master's Thesis. Nnamdi Azikiwe University, Awka, Nigeria.
- Adedoyin, O.O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asian Social Science*, 3(4), 992-1011. Retrieved from [http://www.aessweb.com/journal etail.php?id=5007](http://www.aessweb.com/journal%20etail.php?id=5007).
- Alderson, J.C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J.C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring. In J. Fox, M. Weshe, D. Bayliss, L.Cheng. C. Turner, & C. Doe (Eds.). *Language Testing Reconsidered*. Ottawa: University of Ottawa Press, 21-39.
- Alonge, M.F. (2004). *Measurement and evaluation in education and psychology*. Ado-Ekiti: Adebayo Printing Nig. Ltd.
- Ani, E.N. (2014). Application of item response theory in the development and validation of multiple-choice test in economics. Master's thesis. University of Nigeria, Nsukka.
- Anneduke, I.C. (2009). *The significance of French language in Nigeria education system*. Asaba: Maks Publishers.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
- Borode. B.R. (2014). Effects of lecture and activity based methods on the attitudes of junior secondary school students to essay writing in French. *European Journal of Educational Studies*, 6 (1), 43-49.
- Ceniza, J.C., & Cereno, D.C. (2012). Development of mathematics diagnostic test for DORSHS. [http://www.docscst.edu.ph.index.ph\[academics/graduates school/publication/category/5-volum -1-issue -1-2012?](http://www.docscst.edu.ph.index.ph[academics/graduates%20school/publication/category/5-volum%20-1-issue-1-2012?)

Cherry, K. (2005). Reliability – what is reliability? Retrieved from <http://www.psychology.about.com/od/researchmethods/>

Eleje, L.I., Esomonu, N.P.M., Agu, N.N., Okoye, R.O., Obasi, E., & Onah, F.E. (2016). Development and validation of diagnostic economics test for secondary schools. *World Journal of Education*, 6(3), 90. doi:10.5430/wje.v6n3p90.

Esomonu, N.P.M, & Eleje, L.I. (2020). Effect of diagnostic testing on students' achievement in secondary school quantitative economics. *World Journal of Education*,10(3), 178. DOI: 10.5430/wje.v10n3p178

Esomonu, N. P. M. & Eleje, L. I. (2017). Diagnostic quantitative economics skill test for secondary schools: Development and validation using IRT. *Journal of Education and Practice*, 8(22),110-125. Examination Development Center. (n.d.). Statistical reports. Awambia: Examination Development Center.

Fakeye, D.O. (2006). ICT-Assisted instruction and students' vocabulary achievement in selected senior secondary schools in Ibadan. *Journal of Humanities in Education*, 1 (1) 14-21.

Federal Ministry of Education. (2014). National curriculum for secondary schools: French language. Lagos: Federal Government Press.

Federal Republic of Nigeria. (2014). National policy on education. Lagos: NERDC Press.

Hadfield, J. (2008). Introduction to teaching English. Eigel S.A Portugal: Oxford University Press.

Harmer, J. (2007). The practice of English language teaching. England: Pearson Education Limited.

Iwuji, V.B.C. (2013). Measurement and evaluation for effective teaching and learning. Awka: Summer Educational Publishers (NIG) Limited.

Jang, E.E. (2005). A validity narrative: The effects of cognitive reading skills diagnosis on ESL adult learners' reading comprehension ability in the context of next generation TOEEL (Doctoral Thesis). University of Illinois: Urbana Champaign.

Kolawale, E.B. (2010). Principles of test construction and administration. Lagos: Bolabay Publishers.

Ministry of Education. (2013). Diagnostic assessment in support of school learning. Ontario: Policy/Program memorandum.

National Examinations Council. (2014). Statistics of candidates in French 2005-2014 BECE. Minna: NECO.

Nkpono, H. L. (2001). Application of latent trait models in the development and standardization of physics achievement test for senior secondary students. (Unpublished doctoral dissertation), University of Nigeria, Nsukka.

Okafor, J.N. (2015). Diagnosing test in foreign language. Onitsha: Creative Publisher.

Popham, W.J. (2009). Test better, teach better: The instructional role of assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Reeve, B.B. & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire items and scale properties. In P. Fayers and R.D. Hays (Eds.), *Assessing quality of life in clinical trials: method of practice*. (2nd ed.). USA: Oxford University Press. Retrieved from [http://cancer.Unic.edu/research/faculty/display\\_member-plone.asp?ID-694](http://cancer.Unic.edu/research/faculty/display_member-plone.asp?ID-694).

Thorpe, G. L. & Favia, A. (2012). Data analysis using item response theory methodology: An introduction to selected programs and applications. Retrieved from [http://digitalcommons.lidrary.umaine.edu/psy\\_facpub/20](http://digitalcommons.lidrary.umaine.edu/psy_facpub/20)

West African Examination Council. (2012). Chief examiners' report (May/June). Lagos: WAEC.

West African Examination Council. (2013). Chief examiners' report (May/June). Lagos: WAEC.

West African Examination Council. (2014). Chief examiners' report (May/June). Lagos: WAEC.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass Publishers.

Woolfolk, A., Hughes, M., & Walkup, V. (2008). *Psychology in education*. England: Pearson Longman.

Zhongbao, Z. (2013). An overview of studies on diagnostic testing and its implications for the development of diagnostic speaking test. *International Journal of English Linguistics*, 3(1), 41-45. Canada: Canadian Center of Science and Education.

Appendix A

The French Language Development Skill-Based Test Blue Print  
Writing Skill Based Test

Content and Sub-skills	Know.	Comp.	Appli.	Total
French verbs and conjugation	4	2	4	10
Use of articles and prepositions	4	3	3	10
Negation	2	1	2	5
Language structure and grammar	9	10	6	25
<b>Total</b>	<b>19</b>	<b>16</b>	<b>15</b>	<b>50</b>

Appendix B

FRENCH LANGUAGE DIAGNOSTIC WRITING SKILL-BASED TEST

**Instruction: Attempt all questions and circle the letter that corresponds with the correct answer.**

**Change the underlined words in number 1-10 to opposite**

1. Antoinette est grosse. (a) court (b) mince (c) large (d) grand (e) long
2. Cet enfant est paresseux. (a) paresseuse (b) intelligent (c) malade (d) travailleur (e) presente
3. Les élèves habitent loins de l'école. (a) sur (b) près de (c) derrière (d) sous (e) en face de
4. Madame Ike est en bonne santé. (a) belle (b) jolie (c) malade (d) riche (e) bon malade
5. La classe est brillante. (a) brillante (b) calme (c) vide (d) large (e) jeune
6. Mes sœurs sont pauvres. (a) humble (b) responsables (c) pantalons (d) riches (e) grandes
7. Monsieur Eke est très sévère. (a) sérieux (b) jeune (c) calme (d) gentil (e) bavarde
8. Ogechukwu est mon ami. (a) frère (b) ennueie (c) ennemi (d) fils (e) tante
9. Le filme m'intéresse beaucoup. (a) m'embarrasse (b) m'occupe (c) me plait (d) me fait plaisir (e) m'ennueie
10. Madame Martha est jeune. (a) jolie (b) petite (c) vieille (d) belle (e) laide

**Choose the correct preposition to fill in the gap**

11. La police travaille ..... frontier. (a) de la (b) de (c) du (d) à la (e) au
12. Je vais .... Togo. (a) on (b) en (c) de (d) a (e) au
13. Nous allons ..... église. (a) à la (b) de l' (c) à l' (d) de la (e) en
14. Oluoma a beaucoup ..... fruits. (a) des (b) d' (c) du (d) de (e) de l'
15. Elles vont à Onitsha .... pied. (a) à (b) à l' (c) au (d) aux (e) à la
16. Ifeoma prend ..... pain. (a) de (b) de l' (c) de la (d) des (e) du
17. Je vais .... Lagos. (a) au (b) chez (c) en (d) avec (e) à
18. Nneka va à Londres ..... avion. (a) par (b) sur (c) pour (d) avec (d) à
19. Votre professeur boit ..... thé. (a) du (b) de l' (c) des (d) de la (e) de
20. Il y a beaucoup ..... personnes dans la rue. (a) des (b) de (c) du (d) de la (e) de l'

**Choose the correct option of the sentence from the verbs in brackets**

21. J'(avoir) douze ans. (a) a (b) ai (c) aurais (d) avons (e) es
22. Il (s'appeller) Okechukwu. (a) c'appelle (b) m'appelle (c) s'appelle (d) s'appellent (e) t'appelle
23. Juliette (être) petite. (a) es (b) est (c) étais (d) était (e) sont
24. Vous (aller) au marché. (a) allons (b) allez (c) aller (d) allions (e) allions
25. Vous (manger) du pain ce matin. (a) mangerons (b) mangez (c) mangeons (d) mangent (e) mangerez
26. Le matin les enfants (prendre) du thé. (a) prend (b) prenais (c) prennent (d) prendrez (e) prendons
27. Elles (parler) français. (a) parle (b) parlent (c) parlez (d) parlerons (e) parlons
28. Okechukwu (préférer) la danse que le football. (a) préfère (b) préfères (c) préfères (d) préfèrent
29. Tu (aimer) des ananas. (a) aiment (b) aime (c) aimes (d) aime (e) aimez
30. Il (faire) du bruit. (a) faites (b) fait (c) faites (d) font (e) fait
- Change the sentences in number 31-35 into negative
31. Je suis étudiant. (a) Je suis ne pas étudiant. (b) Je ne pas suis étudiant (c) Je ne suis étudiant pas (d) Je ne suis pas étudiant (e) étudiant je ne pas.
32. Ogonna a 12 ans. (a) Ogonna ne a pas 12 ans (b) Ogonna n'a pas 12 ans (c) Ogonna pas ne a 12 ans (d) Ogonna ne pas 12 ans (e) Ogonna a ne pas 12 ans
33. Tu as un oncle? Non,..... (a) je n'ai pas de l'oncle (b) je n'ai pas d'oncle (c) je ne ai pas d'oncle (d) je ne pas ai d'oncle (e) je n'ai d'oncle pas.
34. Elle parle anglais. (a) elle parle ne pas anglais (b) elle ne Parle pas anglais (c) elle parle ne pas anglais (d) elle ne parle anglais (e) elle ne parle anglais.
35. Nkechietmoi parlons français. (a) Nkechietmoi ne parlons français (b) Nkechi et moi parlons ne pas français (c) Nkechi et moi ne parlons pas français (d) Nkechietmoi pas parlons français (e) Nkechi et moi parlons rien.
- Choose the correct option
36. Josephine .... de lit (a) tombe (b) tombes (c) tombons (d) tombez (e) tombent
37. Vous allez ..... le dentiste. (a) a (b) au (c) chez (d) de (e) rien
38. .... homme aimable me plaît beaucoup. (a) cet (b) ce (c) cette (d) ces (e) c'est
39. .... se trouve la porte? (a) Ou (b) Que (c) Qui (d) Où (e) On
40. Comment ..... vous mes enfants? (a) allons (b) allez (c) vas (d) vais (e) va
41. .... garçon est beau. (a) La (b) Le (c) L' (d) Des (e) Les
42. Elle est allée ..... marche. (a) en (b) au (c) à le (d) aux (e) à la
43. Ces filles sont..... (a) beaux (b) belles (c) beau (d) belle (e) bons
44. Emeka a ..... ses examens. (a) fini (b) écouté (c) finir (d) font (e) écouter
45. Marie ..... comme infirmière. (a) porte (b) joue (c) travaille (d) regarde (e) reste
46. Nous venons..... notre père. (a) avec (b) sont (c) en (d) le (e) sur
47. .... les bienvenus, mes amis. (a) Ah! (b) Allo (c) D'accord (d) Soient (e) Soyez
48. A..... heure vous allez à la plage? (a) quel (b) quelle (c) quels (d) quelles (e) que
49. ...., je parle français, et..... tu parles Yoruba. (a) Toi, lui (b) Lui, moi (c) Toi, nous (d) Moi, toi (e) nous, toi
50. Je n'ai pas..... riz. (a) de (b) de la (c) du (d) de l' (e) des

Appendix C

SCORE GUIDE FOR FRENCH LANGUAGE DIAGNOSTIC WRITING SKILL-BASED TEST

**1.B 2.D 3.B 4.C 5.B 6.D 7.D 8.C 9.E 10.C 11.D 12.E 13.C 14.D 15.A 16.E 17.E 18.A 19.A 20.B  
21.B 22.C 23.B 24.B 25.B 26.C 27.B 28.A 29. C 30.A 31. D 32.B 33.B 34.B 35.C 36.A 37.C 38.A  
39.D 40.B 41.B 42.B 43.B 44.A 45.C 46.A 47.E 48.B 49.D 50.C**