REPUBLIQUE DU CAMEROUN
Paix-Travail-Patrie
******
UNIVERSITE DE YAOUNDE I
******
FACULTE DE SCIENCE DE L'EDUCATION
******
DEPARTEMENT DE CURRICULA ET EVALUATION

S
F E

REPUBLIC OF CAMEROON
Peace-Work-Fatherland
******
UNIVERSITY OF YAOUNDE I
******
FACULTY OF EDUCATION
******
DEPARTMENT OF CURRICULUM AND EVALUATION

# TEST STRUCTURE ANALYSIS AND CREATIVE STUDENT DEVELOPMENT IN BIOLOGY IN SECONDARY SCHOOLS IN CAMEROON

*A Thesis Submitted to the Faculty of Education of the University of Yaounde 1 in partial fulfillment of the requirements for the award of a Doctorate Degree (Ph.D.) in Curriculum and Evaluation.*
*Specialty: Psychometrics*



**BY**

**BESUA ALEXANDER ATUMASA**
Matriculation No: 18W6617
M.Ed. – Educational Measurement and Evaluation – University of Bamenda
B.Ed. (Hons.) - Curriculum Studies and teaching / Biology – University of Buea

**Jury members**

| | | | |
|---|---|---|---|
| President | **Tchombe Mungah S. Therese** | **Professor** | **University of Buea** |
| Examiners | Patrick Fonyuy F. Shey | Professor | University of Bamenda |
| | Maingari Daouda | Professor | University of Yaounde1 |
| | Chaffi Cyrille Ivan | Ass. Professor | University of Yaounde1 |
| Reporters | Maureen Ebanga Tanyi | Professor | University of Yaounde1 |
| | Patrick Kongnyuy | Professor | University of Bamenda |

**November, 2024**

# CERTIFICATION

This is to certify that this research work entitled: "Test Structure Analysis and Creative Student Development in Biology in Secondary Schools in Cameroon" is the original work of BESUA Alexander ATUMASA. This work was submitted in partial fulfillment of the requirements for the award of a Doctorate Degree (Ph.D.) in Curriculum and Evaluation (Psychometrics), under our supervision.

| | |
|---|---|
| _____ | _____ |
| ***PATRICK KONGNYUY*** *Ph.D.* | ***MAUREEN EBANGA TANYI*** *Ph.D.* |
| ***Associate Professor*** | ***Professor*** |
| *DATE*_____ | *DATE* _____ |

# DEDICATION

To

My Family

# AKNOWLEDGEMENTS

Hearty thanks go to my supervisors Professor Maureen Ebanga Tanyi and Associate Professor Patrick Kongnyuy for their regular orientations and reorientation of my efforts in the process of writing this research work.

Special thanks to the University administration, the professors and doctors of the Ph.D. seminars whose lectures inspired my choice of the topic.

I appreciate Prof. Moupou Moise, Prof. Ivo Leke Tambo, Prof. Gbenga Adewale, Prof. Mbala Ze, Prof. Sakwe Peter Masumbe, Chief Dr. Dion Ngute Joseph, Dr. Egbe Martha, Mr. Belle Dimite Andre Guy Donald, Mr. Arthur Mboe Ntuba, Fuh Ernestine Lem epouse Akem among others, for their individual financial, material and moral contributions in the realization of this research project. Prof. Pierre Sakwe Nekongo particularly took time to read through the work and made some remarks in line with my supervisors and drilled me on power point presentation together with Dr. Nobel Sakwe Ikombe.

All the members of the Educational Assessment and Research Network in Africa, (EARNiA) and the online Academia for their lofty papers during our Conferences in Nigeria and Cameroon and the constructive guidance after my papers presentations.

My gratitude also goes to all my classmates, principals and staff of the schools where I did this research, for the collaboration given me while gathering the necessary data for the study. All those who in one way or the other gave a positive contribution towards the realization of this work. The mention of a few names does not cancel the fact that I appreciate the positive contributions of all those I hold in mind.

# TABLE OF CONTENTS



## CHAPTER ONE

## INTRODUCTION

## CHAPTER TWO

## REVIEW OF RELATED LITERATURE

**CHAPTER THREE**

**RESEARCH METHODOLOGY AND DESIGN**

## CHAPTER FOUR

## PRESENTATION AND ANALYSIS OF FINDINGS

## CHAPTER FIVE

## DISCUSSION, CONCLUSION, RECOMMENDATIONS AND PERSPECTIVES

# LIST OF TABLES

# LIST OF FIGURES

# ABBREVIATIONS/ACRONYMS

| | |
|---|---|
| A' Level | Advanced level |
| APA | American Psychological Association |
| B.Ed. | Bachelor Degree in Education |
| B.H.S | Baptist High School |
| CAPIEMP | Teachers Grade One Certificate |
| CAT | Classroom Assessment Technique |
| CBA | Competency Based Approach |
| COVID-19 | Corona Virus Disease of 2019 |
| CSD | Creative Student Development |
| CTT | Classical Test Theory |
| DDT | Direct Delivery Technique |
| EDSEC-2 | Second Education Sector |
| ELA | English Language Arts |
| G.B.H.S | Government Bilingual High School |
| G.C.E | General Certificate of Education |
| G.H.S | Government High School |
| GESP | Growth and Employment Strategy Paper |
| GRE | Graduate Record Examination |

| | |
|---|---|
| GRM | Graded Response Models |
| G.T.T.C | Government Teacher Training College |
| G.T.T.T.C | Government Technical Teachers Training College |
| Ha | Alternative Hypothesis |
| Ho | Null Hypothesis |
| HRD | Human Resource Development |
| H.T.T.C | Higher teacher Training College |
| H.T.T.T.C | Higher Technical Teacher Training College |
| ICC | Item Characteristic Curve |
| IDA | Item Distractor Analysis |
| IRT | Item Response Theory |
| K-R | Kuder Richardson |
| M.Ed. | Master's degree in Education |
| MKO | The More Knowledgeable Other |
| N | Number of respondents |
| NASTA | National Science Teachers Association |
| NPCE | National Paramedic Certificate Examination |
| O' Level | Ordinary level |
| OIT | Organismic integration theory |

| | |
|---|---|
| Ph.D. | Doctor of Philosophy |
| SAR/SM | Government Rural Artisanal and Home Economics centre |
| SAT | Scholastic Aptitude Test |
| SD | Standard Deviation |
| SDG | Structural Development Goals |
| SEM | Standard Error of Measurement |
| SLAT | Spatial Learning Ability Test |
| SPSS | Statistical Packages for Social Science |
| TGA | Test Grading Analysis |
| TIA | Test Item Analysis |
| TRA | Test Reliability Analysis |
| T.R.C | Teachers' Resource Centre |
| TSA | Test Structure Analysis |
| TST | True Score Theory |
| TVA | Test Validity Analysis |
| U.S/U.S.A | The United States of America |
| ZPD | Zone of Proximal Development |

# ABSTRACT

*The objective of the study was to find out if test structure analysis would correlate with creativity development in learners. The survey research design was used in the study. The researcher administered the questionnaire to 1603 students and 36 teachers from 12 schools in the four cultural zones in Cameroon. The simple random sampling technique was used in choosing the sample. The data collected was analyzed using excel spread sheet and the statistical package for social sciences (SPSS) version 23 where the bivariate correlation procedure or The Pearson Product-Moment Correlation Coefficient was used to arrive at the correlation coefficients. It was found that of the five factors of the predictive variables correlated with the criterion variable, all had a positive significant correlation. This indicated that there is a significant relationship between the test structure analysis and creative student development in learners of Biology. Therefore, test items analysis (r = 0.689 and r = 824), test item distractor analysis (r = 0.471 and r = 0.727), test validity analysis (r = 0.751 and r = 0.949), test reliability analysis (r = 0.646 and r = 0.760) and test grading analysis (r = 683 and r= 0.913) of students and teachers respectively, all proved significant at the 0.05 (2-tailed) level, thus, correlating with creativity in students. The researcher recommended that teachers and test developers should consider item analysis, distractors analysis, validity, reliability and grading of tests during test development to reflect the course content and learners' context of learning. With the strict follow-up of these practices, students would be able to use critical thinking and innovative tendencies. A sample test structure and test blueprint (or test specification) has been elaborated from the O-level 2009 syllabus in line with the study and attached to the appendix of this study. From these, a prototype Biology test has been developed. This study generally recommended that teachers in training be given lessons on test development while the old teachers be recycled in seminar workshops, not leaving out the students who should understand the purposes of tests and types of tests to enable the prepare accordingly.*

Key words: **test structure, analysis, creativity, content, context**

# SOMMAIRE

*Le but de cette étude était de vérifier si la pratique sur l'analyse de la structure de l'épreuve pouvait avoir une corrélation avec le développement de la créativité chez les apprenants. Le type de conception utilisé dans l'étude est la conception par sondage. Le questionnaire a été administre aux 1603 élèves et 36 enseignants tiré de 12 établissements rependu dans les quatre aires culturel du Cameroun. L'échantillon a été choisi par la technique d'échantillonnage aléatoire simple. L'analyse des données était faite sur Excel et SPSS version 23. La corrélation de Pearson a été faite sur les variables indépendantes et la variable dépendante et les coefficients ont été calculés. Toutes les cinq variables indépendantes ont eu une corrélation significative et positive avec la variable dépendante. Donc, l'analyse de l'item de l'épreuve (r = 0.689 et r = 824), l'analyse des détracteurs dans les réponses sur l'item (r = 0.471 et r = 0.727), l'analyse de la validité de l'épreuve (r = 0.751 et r = 0.949), l'analyse de fiabilité de l'épreuve (r = 0.646 et r = 0.760) et l'analyse des notes de l'épreuve (r = 683 and r= 0.913) pour les élèves et enseignants respectivement ont eu des corrélations significatives et positives avec le développement de la créativité chez les élèves. Le chercheur recommande que tous les enseignants et les développeurs des épreuves essaient d'analyser les épreuves pendant leurs développements pour qu'elles reflètent le contenu enseigné et le contexte où se trouvent les apprenants. Un bon suivi de ces pratiques emmènera les apprenants à développer l'esprit de créativité, de pensée critique, de l'innovation et être prêt à devenir des entrepreneurs ou employables. Une analyse de la structure de l'épreuve a été élaborée sur le programme de GCE O-level 2009 en Biologie. De ça, une bonne épreuve de Biologie est réalisée suivant les prescriptions des contenus et du contexte ensuite attaché à l'annexe du document. Il est recommandé que les enseignants en formation apprennent le développement des items, et que les anciens enseignants l'apprennent par le recyclage ou séminaires et que les élèves apprennent les types des épreuves pour mieux préparer selon les exigences.*

Mots clés : **structure de l'épreuve, la créativité, employabilité, contexte, contenu**

# CHAPTER ONE

# INTRODUCTION

## Background of the study

Creative student development could be the panacea for solving the problems of increasing unemployment and high reliance on the government for jobs as a result of increasing number of academic certificate holders at the ordinary and advanced levels of education in our country. This study considered great value in the test as an assessment instrument compulsorily used by the learners to furnish information on their progress in school. In the same line, the researcher tried to bring out the possible relationship between the analysis of this instrument in terms of its content, context and development in correlation with creative student development so as to further buttress the link for favourable learning outcomes in learners. In this light, the topic 'test structure analysis and creative student development in secondary schools in Cameroon' was coined for the study to lay a ground work for possible further studies.

A test has been the main instrument used in the assessment of both learners and teachers in our schools. It is from the test results that inferences are drawn for the following; promotion of learners to the next class, scholarship awards, certification etc. The role of the test in the Cameroonian school system needs not to be overemphasized. Tests are used for all forms of assessments and evaluations including diagnostic, formative and summative evaluations. They help the teacher to understand the previous knowledge of the learners in order to help them advance in learning and for remediation (diagnostic evaluation). Tests also serve for ranking of learners, for placement, for aptitude and attitude. Tambo,(2003). A learner may abstain from attending classes but would never expect to stay away

from a test as it is the main form of assessment and evaluation. A test is also that instrument used in measuring some psychological traits of a learner during the process of assessment. A test is made up of items (or questions) to which the learners must respond in order to earn a score. Tanyi, (2019) course notes. The content and context of each test item build up to the general content and context of the test. These aspects of content and context come from the prescription of the syllabus gotten from the subject based on the curriculum used. The curriculum in itself is gotten from the government policy laid down by experts in the domain based on the kind of person to be developed according to the need of a particular society or the country as a whole. Republic of Cameroon, (1995)

The taxonomy of Bloom in its revised model, Anderson and Krathwohl, (2001) has replaced the last level of the six-level taxonomy of educational objectives; (evaluation) with creativity. In another interpretation, the 'synthesis' level has been replaced by 'creativity' and taken above the 'evaluation' level putting the 'ability to create' to be the highest cognitive level in the taxonomy. Thus, learners are expected to evaluate after applying the skills then create in their own knowhow. From knowledge, understanding, application, analysis, synthesis and evaluation to remember, understand, apply, analyse, evaluate and create. Krathwohl et al, (2001) This has not left me indifferent in the quest to find out how test development could meet up with this new trend, so we thought that in analysing the structure of a test to meeting the psychometric properties, we would have taken a step forward in enhancing quality education as enshrined in the United Nations Development Goals number four (4) on quality education adopted by Cameroon. To amplify the need to use the taxonomy of Bloom in ensuring a good testing, Tanyi and Wirngo (2015) highlighted that 'evaluators, assessors, test developers and teachers lack skills in test construction as they usually construct

poor test items without taking good implication of the different levels of the bloom's taxonomy of educational objectives'.

The information gathered during an assessment using a test, is quantified according to the individual scores per item and the sum of all the scores on the items in the test give a total score for the test; say 15/20, 18/20, 9/20. These scores given in numbers or quantitatively (depicting measurement) can be used to give the results of the learners in a classroom; where learners may be ranked as first second, third, fourth or fifth in that order falling in the ordinal level of measurement, or the scores are later transformed qualitatively to provide scores in the form of letters or grades such as A, B, C, D, E grades depicting a pass in the Advanced level or A, B and C grades depicting a pass in the Ordinary level, Tanyi, (2019). The grading of the scores from the figures to the letters, move the measurement from the ordinal scale of measurement to the nominal level. These grading forms which are carried out in our school system fall under the final stage of the evaluation process.

In formal education, official curricula and syllabuses are drawn to be followed by the teacher and students during the teaching learning process. These official documents take into consideration the contents expected to be learned by the students of the first and second cycles and the contexts there of. It is these official programmes from the government that have to be used in the teaching/learning processes in our schools. Contrarily, in the quest for good results from students' performances and recognitions thereof, many teachers tend to use the examination syllabuses provided by the GCE board to teach and to set classroom items and finally the end of course tests or mock examinations. Also, we sought to find out whether the test items set in the classroom tests are valid and reliable before they are put into the test. Are there items and/or test banks kept by individual teachers or in the biology department or in the school system? Do the teachers follow any

test development model; Tanyi et al in developing an item for a test or in developing the test? Pondering on these pertinent questions drew my attention as to the need for a test structure analysis in the process of developing a test item and the test as a whole.

On the other hand, with the coming in of the need of learners to be competent in whatever they learn in order to be creative, develop problem solving skills and innovative, (the Competency Based Approach), MINESEC, (2012), we sought to find out if the test structure analysis could enhance the creativity of the learners in Biology such that they could meet the sustainable development goal number four (4) of the United Nations in order to enhance quality education ( innovative, self-reliant, entrepreneurs or self-employed) of the students so as to become useful to their immediate societies in solving problems Shey & Manju (2019), even while still studying thus, fulfilling the need for quality education.

It is on these premises that we decided to carry out a survey in some secondary schools in Cameroon to find out if test structure analysis had any relationship with the development of creativity in students studying biology then, propose to the educational stakeholders the test structure analysis processes that could better lead to develop student's creativity, by providing a model test structure table and a test specification table (or test blueprint) followed by the classroom test items..

Tests are of different forms, including; the written test commonly known as the pen-and-paper test, the practical test, the oral test, the computer-based test, the portfolio test and the study of files, Atumasa et al (2021). These tests could be presented in various structures such as the multiple-choice questions (MCQ), the structural test questions (fill-in-the blanks, matching items, labelling, sorting and rearranging) and the essay type questions. While this study took interest in all the

test forms, more emphasis was placed on the classroom written, practical and oral tests and on the MCQ and structural tests because the origin and psychometric properties of the test item needed to be analysed before the item integrates the test. Tanyi et al. To actualise these, we used five variables to operationalise the independent (predictive) variable; the test item analysis (TIA), the test item distractor analysis (IDA), the test validity analysis (TVA), the test reliability analysis (TRA) and the test grading analysis (TGA). Based on these, five items per variable were validated to form the questionnaire. In the same light the dependent (criterion) variable; creative student development, was researched using five items on the questionnaire. The data from the five items of each independent variable were used to correlate with the data from the responses of the dependent variable.

All the scales of measurement: Nominal, ordinal, interval and rational (NOIR) scales were implicated at the various levels of the analysis to form the mixed method of reporting in the thesis. Nevertheless, more emphasis was on the ordinal scale as the results were given in numbers which did not consider internal interval or ratio in them.

The field of Psychometrics in which this research is found, falls under the programme of Curriculum and Evaluation and in the domain of measurement and evaluation in education.

According to Tchombe (2019) in her preface, “the test-driven nature of curriculum promotes mechanistic pedagogical practices which do not prepare the learners adequately for personal and skills development.” It is in the light of this assertion that we found out that if tests are given more attention in their development, it would go a long way to address many worries that plague the creativity of the learners since they are most of the time taking one test or the other for various educational purposes.

Tests constructed by the teacher who taught the class or a group of students are of prime interest in this study. The test is mostly based on the content taught; sometimes the cultural context in which the learner is found or in which learning took place, Shey & Fangwi (2020). Some aspects of the activities and methods used in teaching and the teaching/learning materials are seldom taken into consideration in the test. Nenty, (2014). This very powerful and indispensable instrument used for the measurement and assessment of the level of attainment of the learners must not be developed at anyone's discretion without proper follow-up. Tanyi, (2020). The structure and form must be given special attention as to address both the test takers and their various needs. Tanyi, (2016).

To address test takers requires the knowledge about their traditions and cultural beliefs in relation to certain natural phenomena and organisms. To understand their needs on the other hand requires that the teacher probes into the needs of the society where the learners evolve.

Some students in many regions of Cameroon avoid the study of biology especially at the high school level. The reasons of this avoidance are based on the fact that the immediate environmental conditions and the organisms are not totally respected in the study. The use of foreign names on things they are familiar with makes them feel like strangers learning in their own environment. They expect familiar names and appellations to the organisms. Students are expected to dissect a fish in the Far North region at the same time with those in Kumba or Bamenda; by the time a fish leaves the water it gets dried and deteriorates within minutes, losing blood and the rotting of the bowels whereas everything remains intact in Kumba. Examples of plants like pawpaw and coconut would not serve good examples in the northern regions while millet, sorghum; cactus plants are bad examples to the students of the southern regions who might have never seen them. The use of unfamiliar

scientific names of organisms and the study of these foreign organisms scare most from getting interest in Biology. They could apply what they study to their local environment but most students feel that it would not be of their interests to concentrate on studying local names and organisms since they would have to pass the final certification examinations based on the curriculum and/or the examination syllabus.

Similarly, to find out the level of mastery or retention of what the students study in relation to the tests taken in Biology, many students feel that they just write what they can as given on the test paper and the rest is left for the markers to determine. Some could not give a summary of what the tests required of them at each instance, they would shy away and could not even remember an item or two and their possible answers they gave. While some put these lack of creative tendencies as due to the lack of trained teachers in the subject, and inappropriate infrastructures like the laboratories, gardens, culture sites or ecological niches others still put it that they learn more of abstract facts than the contextual and cultural realities. Shey & Fangwi (2020). Even in areas like the centre region where students do biology, many students confer that the organisms treated are mostly of foreign origin since the names are mostly of foreign coinage.

Asked to recognise or identify the organisms treated in their lessons in their direct contexts or locally, many students failed or do wrong identification. Putting all these facts together and taking introspection in the study of Biology since secondary school I realised that many organisms like “Rose Periwinkle” we learned from the primary school but have not been able to find and identify till date.

Biology closely linked to geography in terms of the study of fauna and flora could serve as a backbone to our society today in agriculture; animal husbandry,

animal breeding, crop cultivation, aviculture among others. The contextualisation of it as the students study would therefore motivate the learners to pursue the course up to professional level Kongnyuy, (2021).

The many students who study the biology do their best to have a pass in the subject by consulting huge and a variety of textbooks as they are guided by the teachers. Many students aim at having a pass in the examination and end up having very good grades but consulted to manipulate either an earthworm or millipede, they shy away. Tanyi et al following the true score theory: T = O + E, observed that such evaluation totally based on the certificate attest to their knowledge (Observed score, O) paying little interest on the learners' real life experiences and applications like in the students' creativity, know-how or competencies (True score, T), as a result of the cut-off-point deliberations to enable more candidates to pass, reducing pass scores to as low as 6/20 Baccalaureate, (2013), constitute an error (E) in the measurement and evaluation Process. Most of the students end up in many domains of study other than biology.

As a scholar in the quest of finding out how the natural environment could be used in the study of biology Atumasa, (2014) and the need to empower students to sustainable development and creativity development Atumasa, (2018) (Masters dissertation), the researcher decided to find out if the analysis of classroom tests in biology could have any significant relationship with this highly rated level of creativity development in students of biology as prescribed by the taxonomy of Bloom. This state of affairs has placed many school graduates unemployed and dependent on the state for more competitive entrance examinations such that they could be trained before being absorbed by the public service. Fozing, (2009)

As a teacher having taught at all levels of the educational system I asked myself the question; if the learners are aware of tests being the main instrument to

measure their learning achievements, and are obliged by this to seat for examinations in which tests are given, then can the test not be made in such a way that after answering it, the learners may be sure of their abilities, or the test made in such a way that in the course of answering the test question the learners are exhibiting their acquired kills or talents or trying to solve a common problem to the society? The test could be analysed and given some aspects of 'flow' in the process of answering it. Csikszentmihalyi, (1996).

Though an imperative instrument in educational assessment in Cameroon schools, the development of a test lies in the hands of the individual teacher who would like to do it following the conditions in which he/she is faced. Tambo, (2010) many teachers take past questions and administer to the learners for assessment. The teachers put very little effort to develop valid and reliable items and the consequences are seen on the learners who also become very dependent on other successful persons to employ them as a favour to them or remain unemployed and do menial jobs after school and upon obtaining a certificate. Fozing, (2014). Many teachers do not consider many psychological aspects of their learners in the development of the test. If tests are analyzed to address the needs of the learners, many learners in the course of learning as well as teachers who set the tests would meet the requirements of the society and may become creative; self-reliant, self-employed, problem solvers, creating jobs for others and contribute positively to the development of their communities Fonkoua & Marmoz (2010).

A combination of the above worries and others led the researcher to choose the topic 'Test Structure Analysis and Creative Student Development in Biology in Secondary Schools in Cameroon.'

**Historical Background**

A pioneer in education; Horace Mann postulated an idea in1845 which provided that it was by the wring of examinations that the children of the school should have been able to prove that they had acquired knowledge and not by the oral examination which were written annually.

According to Gallagher, (2003) in her review on educational Psychology article, based on the standardized testing history; 'Mann's interest aimed at finding the best method that could be used in teaching and try to repeat or reproduce them to enable children of all backgrounds to have similar or same opportunities in their learning'.

Quite a bit later after 2003, the proponents of the standardization of tests did not have mastery in handling that aspect in a way that could be appreciated, but in the past decades, they have been able to slowly make it up, and taking a central stage in the working out of educational systems. Today, modern secondary and high school education are centralized around standardized tests.

Diverse from Mann's examination, the first standardized tests that where adopted and used in schools were designed to measure ability and not the achievement. The intelligence test and other assessment tests in that light that became renowned early in the twentieth century were centered on the objectivity that was scientifically proven. ' the alpha and beta tests of the army, that were developed in the period of first world war to bring out able soldiers measured by the ability to do mental activities properly, was taken as a sample of reference in the schools of the area, Prometric, (2016).

"With testing, fast children may work hard to create things or new ways of doing things and the same time enabling us to avoid the spoiling of valuable

resources on children who are slow in learning." In this light learners were followed up academically to put them on the path linked with their career abilities that were suitable for them. The "college entrance examination board" was the most relevant ability test. The name was later changed to 'scholastic aptitude test' or (SAT) which started around the 1920s by the federal government of the United States of America and designed to assess the methods of teaching and in schools in a particular area or the country as a whole. For more than a decade the global economy and the cold war periods emphasized on schools to produce work force that was skillful as they placed a lot of weight on the standardized tests. A major worry remained about the standardized tests as to whether the tests aided in the success of the students from poor home backgrounds or they disturbed them? The complexity of the relationship between educational testing and social inequality was reviewed in a paper for the annual review of sociology'. Grodsky, Warren and Felts (2008).

Most promoters of testing have propagated it to serve to its lowest, partly as an antidote to classroom structures that are stereotyped and difficult to change. Some tests like the scholastic aptitude test, as an example, was said to have been designed to make top colleges into places for brilliant students from all backgrounds on one hand and not only the offspring of the elite class. On the other hand, critics of nowadays note that standardized test scores reflect to a large extent some socio-economic privilege. This could partly be explained that children from rich homes with mediocre marks could improve on their results with expensive private test preparation courses. Nevertheless, differences in test results within learners from different backgrounds could be related to a diversity of issues starting from early childhood malnutrition to differences in available resources at the schools in context.

According to Grodsky et al (2008), ‘tests do not necessarily create more social stratification. They mostly seem to instead reflect the academic advantages that go with socio-economic privilege within the students in the United States of America. This is off-course the evidence that despite Horace Mann’s hopes for standardized tests, equal opportunity for all children still has not become reality’.

According to prometric (2016), “your tests serve like something promised, providing a guarantee from your organization that those granted license or accreditation have met the required level of standards”.

The above citation portrays an important role of the test. It shows that with the administration of a test and the success of a candidate in passing the test gives the guarantee that the successful candidate has gained approval of the attainment of certain aptitudes or competences related to the field of what was tested, indicating that tests give rights for accreditation, attestation and certification. This important tool if well analyzed in its conception and development will therefore give great assurance as to the type of individuals to send out to the society. The standards for educational and psychological testing, (2004), spells out that ‘the development of tests should be scientifically grounded and the need for the documentation of that evidence of the scientific approach’. By this, the standards shows that valuable aspects of test development like test item analysis, the distractor analysis, validity of the test and the reliability, fairness, grading and administrative systems should be well outlined and documented to accompany the test such that the conditions surrounding the test administration, grading and publishing of the results may be considered in every context. ‘Knowledge about the history of learning revealed that all forms of training and education; academic and work-related, are to a large extent the outcomes of social and economic realities. As time passes, the meanings of training and education seem to be similar, overlapping, or totally discrete with

the bases on the prevailing social, political, and economic conditions'. UNESCO – EOLSS (2010).

In education and the world of work, we find roots of human resource development as a field of study. It also has broad foundations in relations to social and technical domains. Primarily, human resource development is a domain of education that stands for work and is about work. The human resource development domain, despite its significant development since the World War II, originated historically and cannot be sidelined from the histories of training and education. Both the history of human resource development and that of education and training for work are the same.

The roots of the history of human resource development can be traced in the origins of training and education. Finding out or tracing where the human resource development (HRD) begins, the progress of man is drawn from learning driven by the need to survive, to the education that has been programmed by Christian tendencies and classics, to war influences and strategies of the military on technical and scientific education, such that the training based on skills and scientific management boosted by the era of industrialization, and finally to the training based on contemporary workers, managers and technicians; who were the immediate predecessors of contemporary Human resource development.

**Contextual Background**

"Anybody who teaches must be able to assess the learners on what they have been retained". This assertion is a common one used by most administrators to remind the teachers of their pedagogic duties. Whether trained or not, these teachers tend to apply this statement to the latter by developing a test which would help them get information for the evaluation of their learners before forwarding the

final reports to the hierarchy. The knowledge on test development would only be determined by the ability to provide scores and the final report. Only a few administrators worry about what type of test was given, whether they followed the test norms or not: The verification of the right item with respect to the content, context and purpose of the test, validity and reliability of the test and the items, the pertinence of the distractors per item response, the grading procedure among others. All that matters are the test scores to fill in the report cards.

In this light the test is valid or reliable only to the one who administered it. Since the hierarchy is interested in the fact that the learners have assessment marks on which they are evaluated, little is done to fine out the limitations of the test or the worries of the test takers. This applies to both trained and untrained teachers. Those trained to some extent follow the Bloom's taxonomy of educational objectives which highlighted on knowledge, understanding, application, analysis, synthesis and evaluation. They tend to use the appropriate verbs for the test, but they often do not develop a test structure guide or even build up a test blueprint or specification table. They therefore do not take care of the needs of the learners in terms of their cognitive domain, affective domain or psychomotor domain as most teachers and school administrators are contented with scores and reports without caring about the validity and reliability of the scores or reports. The tests are done superficially, using the appropriate verbs for those who master them. For those who do not master the verbs and the untrained teachers, they copy the test structure of an old or more experienced teacher without questioning or improving on them, and keep doing so till they are given some knowledge in that line either individually through reading or in seminars and workshops.

In this light, the act of analyzing the structure of a test is not a common phenomenon. Tests are produced on daily basis as the teachers tend to assess their

learners. Teachers ignore the known structures and procedures for setting standards of classroom tests. Most of the tests set by individual teachers to some extents are given to colleagues or the head of department to verify some aspects like the content or level of the language to the targeted test takers. Little or nothing is calculated and documented as to the item quality, reliability of the test or the validity of the test.

From the responses got in the process of carrying out the study, the pilot testing of the instruments, experience and literature present that teachers write test items even at the last minute to the test time, without prior pre-test or verification for validity or reliability of the item or even making the marking guide. It is after the administration of the test that they turn to make a marking guide; for those who even think to do so or mark the script of the one they consider as the best student, and use his/her copy as the reference marking guide. It is also after testing that they discover that some items were not at the level of the learner, were wrongly set or wrongly phrased. The results thereof are mostly superficial and not representative of the learners. The learner with the worst mark may not have been addressed properly by the test items or might have gotten the wrong understanding of the items due to the ambiguity in the test items or the test as a whole. It is at the backdrop of these contextual worries that the researcher has decided to engage in this study.

Furthermore, with the revised bloom taxonomy from Knowledge understanding, applying, Analysis, synthesis and evaluation, to remembering, understanding, applying, analysis, evaluation and creation replacing the position of "evaluation" at the last stage with "creation", while replacing the position of "synthesis" with "evaluation", learners are called now to create new things, new ways and new manners of doing things. The outcome of such innovations would

ensure quality education as prescribed by the structural development goals (SDG) of the United Nations: To promote innovation, critical thinking skills, participation in societal development and craving to become self-reliant in taking care of oneself and the needs. Thus, the quest for creativity in students is our main problem in this study.

## Conceptual Background

Here the terms and concepts used are defined as implied in the context of the study as compared with the dictionary meanings and other definitions by authors.

### Psychometrics

According to pmetricsoc (2021), 'the society of psychometrics explains that the devoted society interested in the quantitative measurement advancement takes lessons and training in some psychological studies and the social sciences. This description could be a very general one with regards to psychometrics; however the word quantitative is emphasized as in the previous sentence'. It is in line with this definition that emphasis is placed on the ten point Likert scale in the questionnaire and reporting more of the data gathered in the questionnaire than the semi-structured interview and checklist. Individually, other researchers on psychometrics like Borsboom defined it as follows;

Borsboom (2006), defines psychometrics as follows; Psychometrics can also be said to be 'a scientific discipline with interest in the construction of tools used in assessment, measurement instruments and models that have been formalized and which may be used to link to or relate with phenomena that can be observed; like answering of questions in a test of intelligent quotient (IQ), to attributes which are theoretical like intelligence'.

In line with both the general definition and the personal definition presented above, this study fall in education and is based on the quality of a test as a measuring instrument in trying to analyse the test structure in bringing out the test item analysis, the item distractor analysis, the test validity analysis, the test reliability analysis and the test grading analysis to light and presenting a formal way of presenting the test structure.

In this study, psychometrics is considered as the process of ensuring that the items in a test instrument measure observable phenomena typical of the learning environment of the students (Biology) in connection to the theoretical attribute (creativity). The study is focused on test structure analysis and the main actors are the teachers.

**Test Structure**

Test structure is a concept commonly used to describe the presentation or form of a test. Most of the common usage refers to the parts of the test on one hand like part one; objective type questions, part two; structural type questions, and part three; essay type questions. On the other hand, test structure may refer to a test which has the oral part, written part and/or the practical part.

Here, in this study the "test structure" refers to the framework of a test as an instrument for assessment with the building blocks being the test items. It is considered to be the sum total of all the traits that make the test fit to address the needs of the learners so as to enable them to develop creativity in order to take initiative to carry out projects that will enable them innovate while learning and be imbibed with entrepreneurial skills such that upon graduation they could become employers or innovative employee by virtue of their endowed skills, instead of seeking to be employed by virtue of the certificates obtained. Test structure could

be explained to mean the image or representation of the items of a test following particular aspects like the content, context and objectives peculiar to the geographical or cultural setup in which learning took place.

**Test**

According to the Cambridge advanced learner's dictionary (third edition); a test is way of discovering, by questions or practical activities, what someone knows or what someone or something can do or is like. Miriam Webster Dictionary define test as something (as a series of questions or exercises) for measuring the skill, knowledge, intelligence, capacities, or aptitudes of an individual or group.

**Structure**

Structure is the form or framework that serves as the backbone to the general presentation of the items as a whole giving an impression of a sum total of a collection.

**Analysis**

This is the overview of a test from the general viewpoint to specific aspects that make up the whole. These specific aspects are treated as to understand their individual peculiarity and function in relation to the whole. A test is made up of items with various characteristics like the difficulty index, cognitive level etc. Here the analysis involves test items, validity, reliability, distractors in proposed responses and the grading of tests. This is what we call the Operationalisation of the dependent variable into working variables.

## Creativity

Amabile (1996 p. 33), says that 'the responses or products quality accepted to be creative as per the observers considered to be appropriate, and ….the procedure through which anything considered as such is produced'.

Csikszentmihalyi, (1996, p.28), says that creativity 'is any idea, product or act that causes a particular existing field or domain to change or that makes a new domain from an older one in existence', with domain considered as 'procedures and a set of rules symbolic to it'. Ford,(1995 p.17), considers creativity to be a judgment which is subjective, specific in context of the new aspect and value of the particular outcome of the behavior of an individual or a collectivity'. Woodman (1995 p. 293) added that 'the creation of useful, valuable new service, idea, product, process or procedure by some persons who work in an organization considered to be a complex social one'.

## Creative student

This is a student who uses the knowledge acquired in the classroom to solve problems in context and out of context. They use old procedures to achieve new results and new procedures to improve on old products. A creative student thinks beyond his classroom and immediate environment to impact other domains. Critical thinking is part of the life of a creative student; they match theory and practice by implementing what they have studied and keep track of the results obtained. **A** creative student takes initiative in their domain to innovate, modernize or ameliorate the current learning environment with respect to products or services of the school and the immediate community. These are persons who in no way can be replaced by robots because they create new ways of doing things in the learning

process, design new procedures and innovate in terms of technological knowledge and other competences.

**Student development**

Independent application of skills and competencies in real life situations and matching theory with practice following fixed objectives and diverse contents and contexts such that even after graduation the student is able to carry out certain innovative projects to the benefit of themselves and their communities.

**Creative Student Development**

The acculturation/enculturation of learner into the culture of innovation by allowing them and giving them the opportunities to apply the knowledge acquired in the classroom to their environment for the immediate benefit of the community as a whole.

**Biology**

A college subject in the domain of sciences in which students learn about living things, their structures, functions, habitat and their overall impact to the economic development of the human and the environment as a whole.

**Secondary schools**

The level of formal education we have after the completion of the primary school. The secondary school in Cameroon is divided into two the first and second cycle. The first cycle lasts for five years and the second cycle last for two years. At the end of the first cycle students sit for the General Certificate of Education (G.C.E.) Ordinary Level (O Level) and at the end of the second cycle students sit for the G.C.E. Advanced Level.

## Test theory

'In essence, test theory is the grouping of mathematically related concepts that clarify and formalize some questions or items about using and constructing tests, which are in turn provided with methods that can be used to give responses to them' McDonald, (1999 p.9). Cited in Ji and Adam (2009 p. 3) 'test theories put in place a global framework that create a liaison of observable variables like the test scores and item scores, with unobservable variables like ability scores and true scores. Due to the generality of the specifications, test scores could not be verified and published as to be of use or not to be of use. Therefore, a test theory engaged in the introduction of such concepts as test scores, true scores and error scores cannot be judged as being useless or useful unless it was specified fully in forms of models particular to it. Hambleton and Jones, (1993).

## Test model

Particular models of test were formulated in the midst of the test theory framework and they specify in detail which were considerable, the relationships within a set of theoretical test concepts linked with a collection of assumption sets and the relationships with the concepts. The worthiness of the said models could be evaluated in consideration of a particular set of test data. Such an evaluation might be done by conducting a well-designed set of suitable investigations which are empirical or studies which are model fit. In addition, a logical analysis of the assumption model in relation to the test data could be carried out to verify the validity of the model.

## Best fit line, trend line, and gradient line

These words mean the line used to indicate the direction of the scatter plots

(positive or negative) with regards to its inclination on the X and Y axes of the plot.

**Theoretical Background**

Knowledge in our secondary schools is being transmitted vertically from top to bottom or from the macro to the micro level in a cognitive manner where only written tests determined the success of a learner from one class to the other since independence. These cognitive structures tend to interpret knowledge and reality in a predictive dimension and trying to transform the cognition into practical life skills. These cognitive structures convey the politics of the system in order to make the sense of existence, reality and knowledge sustainable. The transformation of this political drive in the provision of knowledge to a scientific (universal) drive in terms of content, context and pedagogy will provide a more universally legitimate system especially in the study of Biology such that learners become proactive in the search for solutions in biological problems within and around them. The dependence on the cognition in the realization of life realities needs to be accompanied with many other philosophies to realise the solutions in education and ensure sustainability. The inclusion of the other domains of the Bloom's taxonomy (psychomotor and socio-affective) would complete the requirements demanded by the classical test theory and the item response theory in the test structure analysis and buttress the demands of the socio-constructivist view of studies and the self-determination theory in the quest for creativity in students. The motivational theory and the autotelic flow theory that link both the independent and the dependent variables will have a grounded place in the study.

**Justification of the Study**

This study is justified in its content and context as it was based on the possibility of enhancing good classroom test (effectiveness and efficiency in test development) in relation with the development of students with creative tendencies (quality education). Following the government's Growth and Employment Strategy Paper (GESP);(2010/2020), on education, training and research in Cameroon, instead of laying emphasis on the learners and the learning process, in providing conducive environment for the acquisition of knowledge and the growth and development of the learners, the long-term education sector (EDSEC-2) component on effectiveness and quality has elaborated as long-term goal to enhance effectiveness and quality of secondary education in other sectors other than the learning itself. The results expected by 2020 were to ensure that Pedagogic supervision in secondary education was competitive in success rates, teachers and infrastructural availability and supervision. Instead of centering on the quality of the learners, the indicators were as follows:

- The rate of internal effectiveness of the system

- The success rate of males and females in classroom examinations and competitive entrance examinations,

- A minimum of 80% of both male and female teachers and supervisors of pedagogy obtained a pedagogic assessment score which was satisfactory.

- The rate of completion of the school cycle according to sex (male or female)

- The rate of supervision

- The distribution of teachers according to sex and the hazards involved in the secondary education

- A minimum of 80% of technical/commercial High Schools were involved in alternative vocational and technical supervision

- A mini-incubator per region in a technical secondary school.

From the above facts it could be seen that the goal of education was centered on the effectiveness and quality of the supervisors, infrastructure and the quantity of learners according to sex. The emphasis here was on those who carried out the checks and balances of the education and not the learning process and the learners. This study therefore comes to fill in the gap in support of the quality education but laying emphasis on the students and their learning process which will lead to quality assessment using quality assessment tools (test) and the development of creative students.

From the outside of the school, supervisors tend to provide information about the functioning of the school. This mostly took place through correspondences or communication while the reality on the field remained what it was. The main actors of the educational system; teachers and learners were seen only in the distance. The teachers gave information about their teaching to the school administrators who sent the reports to the hierarchy. Everything centered around teaching and not on learning. The number of hours taught, the percentage of syllabus coverage, the success rate and all other aspects surrounding teaching were reported by the teacher. The only opportunity given to the learner to report their learning was only after a test has been written. From the test scores it was possible for a student to say I had passed or failed. Yet they did not have the choice to decide whether to pass or fail as it was still the teacher who decided on the criteria of pass or fail, the learner had no contribution to it, thus, the place of this study to bridge the gap and give some value to the learning process and make the learner a worthwhile actor in their learning.

This research work therefore intended to valorize the place of the learner in the pursuing of their personal studies project. The new curriculum of the secondary school placed the learners at the center of pedagogy following the Competency Based Approach to pedagogy, yet all efforts still tend to relegate the learners to the peripheries of the process. With the completion of this study based on test structure analysis for creative student development, learners' personality traits, basic competencies and their contextual background are expected to be considered while developing a test such that the society in which they find themselves would be a major force to project them to success than limiting them from achieving their goals.

Test would have to address the needs of the learners and the specific society in which they are learning such that in a day-to-day manner the learners may apply what they learnt in the resolution of the immediate societal problems. This would enable them gain self-reliant capacities and be valued in the society. The researcher intended to develop a sample test structure that will serve as a blue-print to the development of Biology test following given contents and varied contexts in line with the Bloom's taxonomy and the various test theories and learning theories used in this research work.

**Statement of the problem**

Many aspects come into play in order for students to be creative in the study of biology in the secondary schools in Cameroon. However these aspects are either grossly neglected by the students and teachers or are considered just partially. The students have the duty to carry out hands on activities or practical as they study biology in school and after school hours. On the contrary, students tend to wait only for the teacher to expose content by topics before the students can make little efforts to study orally in lectures or practically in the laboratory. Learners solely depend on teachers to the extent of even demanding speculated questions from teachers for some high stake examinations or certification examinations. Thus it is common to find a student scoring an 'A' grade in Advanced level in Biology with no accompanying creative skills to show for. They get the grades but when asked to contextualize their knowledge or solve common daily life problems in the light of biology, they shy away and avoid the subject as they further their studies.

These tendencies which make the students solely dependent on the teachers relegate initiative and all that goes with creativity in biology. Students confidently deny biology with pretexts that it involves more of reading and cramming aspects that they do not have a mastery on; like scientific names of organisms and their classifications whereas those organisms are not commonly found in their immediate environments.

This laxity on the part of the students and the stakes in the study of biology made the researcher to find out possible ways of engaging the students in the study of biology such that they could develop creative minds while in the learning process.

It was then observed that every student; serious or not, would want to have a score in biology as part of their end-of-term examinations or promotion examination since the subject is compulsory in many secondary schools. In this light, the researcher sought to find out if the test they take could in any way have a relationship with the development of creativity in students since they all take biology tests? It was as a backdrop of these facts that the researcher coined the topic; 'test structure analysis and creative student development in Biology in secondary schools in Cameroon,' as a thesis topic for research for a Ph.D. in curriculum and evaluation specialized in Psychometrics.

**Objectives of the Study**

Generally, this study intended to find out if test structure analysis has a significant relationship with creative student development in Biology in secondary schools in Cameroon. The researcher intends to consequently provide a test structure analysis procedure from the study.

The following specific objectives guided the study:

1- To find out if test item analysis significantly relate with creative student development in Biology in secondary schools in Cameroon.

2- To check out on the possible level of relationship between test item distractor analysis with creative student development in Biology in secondary schools in Cameroon.

3- To determine the relative significance in the relationship between test validity analysis and creative student development in Biology in secondary schools in Cameroon.

4- To find out at what level of significance the test reliability analysis correlates with creative student development in Biology in secondary schools in Cameroon.

5- To find out if there is a significant relationship between test grading analysis and creative student development in Biology in secondary schools in Cameroon.

**Research questions**

This study generally aimed at investigating the possibility of the analysis of test structure having a significant relationship with the development of creative students in Biology in secondary schools in Cameroon.

In other words: Can there be any relationship between the analysis of test structure and creative student development in Biology in secondary schools in Cameroon?

The following specific research questions enabled the researcher to further explore the study.

1- To what significant level does test item analysis correlate with creative student development in Biology in secondary schools in Cameroon?
2- What relationship does the test item distractor analysis have with creative student development in Biology in secondary schools in Cameroon?
3- At what level of significance does test validity analysis tend to relate with creative student development in Biology in secondary schools in Cameroon?
4- To what extent does the reliability of test relate with creative student development in Biology in secondary schools in Cameroon?
5- What level of significance can test grading provide in relationship with creative student development in Biology in secondary schools in Cameroon?

## Research Hypotheses

As a general research hypothesis; the analysis of test structure has a significant relationship with creative student development in Biology in secondary schools in Cameroon.

The alternative hypotheses that follow, guided the study:

Ha1-Test item analysis to a significant extent relates with creative student development in Biology in secondary schools in Cameroon.

Ha2- There is a significant relationship between test item distractor analysis and Creative student development in Biology in secondary schools in Cameroon.

Ha3- To a significant level, test validity analysis relate with creative student development in Biology in secondary schools in Cameroon.

Ha4-Significantly, test reliability analysis relates with creative student development in Biology in secondary schools in Cameroon.

Ha5- There is a significant level of relationship between test grading analysis and creative student development in Biology in secondary schools in Cameroon.

## Scope of the Study

The scope of this study was done in three main domains following the sociological domain, psychological domain and the Philosophical domain involved in the study. The content delimitation and time have been considered under the psychological domain and present the domain in which the study falls and the time of the study. Both the content delimitation and time delimitation are given detailed implications as used in the study respectively. This study was carried out at a period when the COVID – 19 was very active and ravaging the world and the restrictions were enormous (2019 to 2021). The population per class was 50

students maximum and many streams were created which attended school in shifts of morning and afternoon alternatively.

**Sociological domain**

This domain takes into consideration the geographical and cultural scopes as used in the study. Geographically, this study covered the whole of Cameroon. It took into consideration the cultural diversity of the country. From the four main cultural spheres, schools were drawn to give the approximate representation of the area. These included the grass field cultural sphere, the Sudano-Sahelian cultural sphere, the Fang-Beti cultural sphere and the Coastal or Sawa cultural sphere. The Grass field comprised the North West (NW) and the West (W) regions, the Sudano-Sahelian zone comprises the Adamawa (A), the North (N), and the Far North (FN) regions, the Coastal (Sawa) Zone comprised the South West (SW) and the Littoral (L) region, while the Fang-Beti or Centre south zone comprised the Centre (C), the East (E) and the South (S) regions. A representative population was taken mainly from one or two of the regional headquarters representing each of the zones. Bafoussam and Bamenda were used from the grass field zone, Ngaoundere for the Sudano-Sahelian zone, Douala and Buea for the Coastal (Sawa) zone and Yaounde for the Centre South or Fang-Beti zone. Schools were drawn from these towns following a stratified sampling technique as to represent the government schools, the lay private schools and the denominational or mission schools.

Culturally Cameroon is made up of more than 250 ethnic groups, many of which are made up of diverse tribes and to clans. Despite this diversity, Cameroon uses two official languages: English and French, with only two of the ten regions typically dominated by English while the other eight are dominated by French. This study took into consideration, these diversities in carrying out the study in the

four main cultural spheres (Grass field, coastal, central plateau and the sudano-sahelian) and in both the English speaking (North West and South West) and the French speaking regions (Adamawa, Centre, Littoral and West); altogether six regions involved in the study. The study however considered only those students and teachers of the English sub-system of education in government, lay-private and mission secondary schools and bilingual secondary schools in the selected towns of the study. Even though students of both English and French backgrounds attend the same school, the study was carried out typically in English as the instruments were only in English and the respondents were all in the biology class and responded in English. No respondent granted interview in French. The English version of the results is therefore more authentic than any other translated copy.

**Psychological domain**

This study content falls in the domain of social sciences, in the field of education, in the branch of curriculum and evaluation and in the specialty of Psychometrics (as presented under the Ph.D specialties of the department). It intends to develop some standards in tests and testing practices as it tries to put in the aspects of validity, reliability and fairness in the educational measuring instrument which is the test. In so doing it gears to find out if the good analysis of test structure correlates with creative student development such that learning could be centered in the students who should carry their learning as a project for themselves and the society as a whole.

It is a social science because it studies the dynamics of a people and their society and reports them following a laid down principle prescribed by research. It is in the domain of education because it deals with the structure of tests used in the classroom and the relationship of that to the development of creativity in the learners of Biology. It falls under the branch of study; Curriculum and evaluation

in the sense that as a branch of study offered at the University of Yaounde 1, I found interest in the quality of the assessment instrument which is a test and as this can be studied linked to this branch, it automatically is under the branch of study as prescribed by the institution (other specialties of the department include; qualiticians, curriculum developers and docimology).

It is a Psychometric study in the sense that it goes to look at the internal and external qualities of the assessment instrument used to measure psychological traits. This measuring instrument, the test is looked into, in order to assure its validity and reliability, putting it under Psychometrics as opposed to Docimology which is the science of assessment or evaluation. The content in terms of the independent (or predictive) variable which was operationalised dwelled on the following:

1- Test item analysis (TIA)
2- Test items distractor analysis (IDA)
3- Test validity analysis (TVA)
4- Test reliability analysis (TRA) and
5- Test grading analysis.(TGA)

These are to be used to buttress the main variable 'Test Structure Analysis' so as to form research items and carry out the findings by gathering data through the questionnaire to teachers and students of biology. The content of this study also considered the dependent (or criterion) variable; the creative student development which was also researched using five questions on the questionnaire. Relating the independent variables to the dependent variable was based on the theories involved.

This study falls in the context of time within March, 2019 to December, 2021. That is the time the researcher was enrolled into the Ph.D. programme to the

time the supervisor approved the last manuscript concerning the collection and analysis of data. The research data were collected in February 2021. This notion of time comes into this work to provide the reader the context in which to understand the results of the study. This period of time was characterized by COVID-19 and schools were just allowed to function with strict barrier measures in place. Not more than 50 students were allowed into a regular class size and government schools or those with very high student population were asked to attend school in shifts of morning and evening. The regions of the North West and South West Cameroon were at war for secession that left only few functional schools in the regional headquarters. The choice of just one school from each of the regions could be explained by the great malfunctioning of the education system in those areas and the insecurity thereof that could infringe into the results of the research. The understanding of this time frame would present a better understanding and appreciation of the results of the study.

**Philosophical / theoretical domain**

Philosophically this study falls under the domain of philosophy called pragmatism as it seeks to handle issues of practicality and human endeavours in relation to solving real problem-situations of education and the society at large. Theoretically, the test theories included; classical test theory which includes the test score theory, and the item response theory. These theories guide the development of tests used in the classroom and the standardized tests. This study however lays emphasis on the use of these theories in the development and analysis of classroom tests.

It also considers the learning theories related to the development of creativity in students. These include; the cognitivist theories, the gestalts (insight learning) theories and the discovery learning theories. These and other related theories go to lay a foundation on the dependent variable of this work. The autotelic flow theory

serves as a bridge that links both the predictive variable and the criterion variable. It demands a teacher to set the test to flow from the simplest to the most difficult so that the learner may move along it like a game in computer such that where they find any difficulties; they re-strategize to get over the difficulty without abandoning the test or the question. In this light, students will learn to do research even after a test and get the right answer to the question they could not go through during the exam. By these, they will be able to be creative and contribute to societal development and self-development.

**Significance of the Study**

This study is justified as it falls in line with the bid to promote quality education in schools. This could be attained from the classroom where tests are set and taken by students, then corrections are made and the way forward determined.

According to the government's strategic paper on education and training (2005) to expire by 2020, it is very clear that there is need for equal access to schools, quality education and good governance. From statistics gathered from the results of the end of course examinations since 2005, it has been proven that access has been monitored to a certain level as the percentage of students passed; both male and female have been calculated.

Also, the governance aspect has been taken into consideration and is still being addressed in the "Clean School" programme of the Minister of secondary education. Nevertheless, the issue of quality education still remains a preoccupation to the government as no indicators to this have been followed.

According to the United Nations Structural Development Goals (SDG) to expire in the year 2030, goal number four (4) spells out what quality education should entail. According to them, quality education should take into consideration;

sustainable learning, self-employment, research/innovation, equity and human right.

It would be very important therefore for the system to follow all aspects of the quality education such that some indicators like sustainable learning, self-employment, research/innovation and human right could be measured indicators of the output of students who come out of the system. It is hoped that with the analysis of test structure, there will be inputs of values into the learners who will not only sit exams for the certificates or diplomas, but to show their know-how in terms of competencies and mastery of the course work, and be able to apply the knowledge to their context and other contexts.

Specifically, the study will be of significance to the following:

**Teachers**

Teachers would have as their main and basic role to help learners uncover the potentials in them and apply these naturally endowed and acquired potentials to various contexts such that the learners not only become creative at their area of study, but to also apply them to any other context in which they are found.

**The learners**

The learners when conscious of the role of the teacher and his independence in terms of test construction and evaluation, would allocate great respect and commitment to their actions in the classroom such that the outcome of the students' output would not be the success or failure of the teacher but of the learner themselves. They should be conscious in their potentials and seize the bull by the horn when given an opportunity.

**The society**

The society should be aware of the potentials of the learners from the school system and consequently adjust to accommodate them such that a student can participate in community development and own business. Also, that students can be employed into an enterprise immediately they leave school or during holidays.

**Variables of the Study**

Two main variables of the study were:

1- The predictive or independent variable: Test structure analysis (TSA) and
2- The criterion or dependent variable: Creative student development (CSD).

The predictive variable was operationalized into five working variables:

1- Test Item Statistics (TIS)
2- Test Items Distractors Analysis (IDA)
3- Test Validity Analysis (TVA)
4- Test Reliability Analysis (TRA)
5- Test Grading Analysis (TGA)

In all, the criterion or dependent variable which was not operationalized, remained the same. That was: Creative Student Development in Biology (CSD).

Test structure analysis in this study comprised all the stated operationalized variables from 1-5 with each having a series of five items in the questionnaire to further operationalize them and gather useful data from the respondents. The items or questions on the interview were not the same but the guide provided insight for probing into the answers needed in the research.

The creative student development (CSD) as criterion variable though not operationalized also had a series of five questions which enabled the researcher to

gather data for the study. The items in the interview guide were not the same however. All these led to the presentation of a coherent and reliable data in the study.

This chapter serves as a projection to what is expected to be done as the research work goes on. At the end of the research, the researcher intends to develop a test structure table and a test blue print or specification table in Biology and present pertinent analytical factors so that it serves as a model of standardized test and or classroom test. The results thereof were used in this light to pursue the laid down procedure.

Considering that this study is typically in English and considers the English sub system of education, a huge population from the two English speaking regions was to be considered but due to the war in the two regions, only one of the operating schools was taken from Bamenda and Buea to complete for the grassfield cultural zone and the Sawa or coastal zone respectively. Nevertheless the pilot tests were taken from schools in Bamenda which gave the region the place it deserved in the research work. These difficulties and many others are presented in chapter five under the limitations of the study.

# CHAPTER TWO

# REVIEW OF RELATED LITERATURE

This chapter is comprised of a theoretical framework to back the study, the conceptual framework, the Review of Empirical Studies, Cultural Relevance, Summary of Literature and Conclusion. It is based on the works of many predecessors or authors who wrote one or more articles or books in the light of the topic or the contributions of proponents of theories who gave their view points to support knowledge in the domain of my research work. Seven main theories underlay the study; three to back up the independent variable and the other four to back up the dependent variable.

## Conceptual Framework

This comprises an in depth presentation of the main words used in the study that could help in better understanding of the topic and the study as a whole.

### Test

From the teacher's stand point, a test is a way of discovering, by questions or practical activities, what a student knows or what a student or learner can do or is like. This could take the form of a written, oral or practical activity. The teacher then corrects and allocates scores that would be evaluated and taken as answers to know the student's ability in the domain of the test or their performances in terms of achievements. It is from the test scores that decisions are made on the learning of the students.

Test are also considered as something (as a series of questions or exercises) for measuring the skill, knowledge, intelligence, capacities, or aptitudes of an individual or group. This definition from the Miriam Webster Dictionary presents a

test as a tool used in the gathering of information concerning the students or learners. It is from the information or data collected that decisions are made concerning the students.

In line with the above definition, this study considers a test as an instrument or tool used to gather information based on the learners' ability in terms of knowledge, aptitude and competences for final related purposes, taking into consideration the content, context, culture and environmental realities of the learners. A test is to be considered as an instrument for assessment of the learner's level of acquisition of the appropriate knowledge for a given purpose related to the acquisition of creativity for local and extrapolated applications in daily life of the learner.

**Test Analysis**

Test analysis can be seen as a procedure towards looking at an assessment instrument from which test information can be derived and used. The basis of such a test is termed the test basis. The information we need in order to start the test analysis is the test basis which we also use to create our own cases of a test. Walter Kluwer, (2014).

In this study, test analysis is the taking into consideration societal realities, communal realities and individual particularities as a people in the construction of a test so as to meet with or address the needs in a test, and thus teaching/learning context.

**Creativity**

Amabile, (1996, p.33) considered creativity to be 'the responses and products quality could be examined and given some attribute to be creative by observers

considered to be appropriate, and . . . the procedure through which something judged as such, is produced'.

Csikszent M, (1996, p.28) puts creativity as 'an idea, act or product that varies an existing field or domain in other words, that transforms the existing domain to form a new one'. A set of symbolic procedures and rules make up what we call the domain.

Ford, (1995, p.17): According to him, creativity is 'a subjective judgment of the newness and value of a particular outcome from an individual's behaviour or collective behavior considering a specific context'.

Woodman, (1995, p. 293) on his part presents creativity to be 'the creation of an idea, service, useful new product, valuable, process or procedure by some individuals in a production chain within a complex social organisation'.

'A series or variety of definitions on creativity is found as presented above. Most of the researches based on empirical data have defined the concept of creativity as an outcome, with much focus on the production of useful and new ideas concerning services, products, procedures and processes'. Oldham & Cummings, (1996), Zhou (1998).

> "On the basis of this definition, 'researchers have examined creative solutions to learning problems, creative teaching and learning strategies, and creative changes in research and experimentation processes'. West & Anderson, (1996). 'Creative outcomes could range from minor adaptations in products or workflow to major development and breakthroughs on new processes or products'. Mumford & Gustafson, (1988). On his part, Gardner, (1993) contrasted 'little "C" creativity as the type that we all project in our daily lives and big "C" creativity the as the sort of

breakthrough which happens only in very rare occasions'. Csikszentmihalyi, (1996) seemed to be in accordance with this stance. 'In the same similar vein, Boden (1991), distinguished between Historical creativity and Psychological creativity; Nickerson (1999). Whereas Psychological creativity had to do with the generation idea at a personal level, Historical creativity had to do with ideas that are fundamentally new in relation to the whole of human history', Boden, (1991, p. 32). "Creativity is considered to be a useful novelty. Both usefulness and novelty are necessary conditions for a solution or idea to be considered creative. A handful of persons could ask that an idea should contain some elements of novelty or originality to be judged as creative. In the context of creativity at work, novelty however, is not enough to be considered creative: Ideas also must be useful. A new idea or novelty that has not got any potential value is not usual, but also not creative". Zhou & George, (2003). Creative solutions to studying problems, creative student development strategies, and creative development changes in work processes and procedures are all included in this definition. "Practices, services, procedures and processes that are considered to be novel and original, appropriate and potentially useful are said to be creativity". Zhou & Shalley (2003).

"Innovation and creativity are related closely and concepts that are overlapping, but they are not in any way interchangeable. Both creativity and innovation have often been studied differently in isolation using different methodologies and models by researchers. Creativity has been studied in psychology at the individual level, while innovation has been studied in economics and management at the organizational level. Sternberg & Lubart". (1999).

Innovation researchers and organizational scientists define innovation as;

"The intentional introduction and application within a group, role or processes, organization of ideas, products or procedures new to the relevant unit of adoption designed to significantly benefit the group, the individual, the organization or wider society". West & Farr, (p. 9).

While we consider creativity to be 'the production of novel and useful ideas in any domain, innovation is defined as the successful implementation of creative ideas within an organization'. Amabile, (1996). Creativity is important both in itself and also of itself. It could be conceptualized as a necessary pre-condition required for innovation. Therefore, creativity is the seed of innovation. However, "the link between the two is not straightforward and linear". King, (1995, p. 87). Innovative organizations are those that introduce new technologies or management techniques pertaining to products, services, and processes earlier than their competitors.

"Individuals who are creative may influence all stages of the process of innovation, but organizational innovation requires more than creativity. Many factors influence the failure or success of innovation and it depends on them, both inside and outside of the organization, from the relationship of employee to market and regulatory forces. This means that, creativity requires innovative ideas, whereas innovation can be based on ideas that are adopted from external sources. Creativity is therefore a necessary though not sufficient a condition for innovation. It should be noted that the study was concerned solely with creativity including the factors promoting creativity within learners of Biology". King, (1995, p. 87).

We define creativity in this study as the dynamism involved in critical thinking, logical application of ideas, innovative tendencies, proactiveness in the research of facts, and the application of findings to solve actual and future problems in diverse contexts in the society in which the learners live.

**Creative Student**

Special group of students studying a course or programme and developing more ideas and opportunities for themselves and others to be innovative, employed, or improving output for greater performances in skills and income for a better society in which they live.

**Student development**

Training for learners to practice a particular skill or job either before the end of the course (or within the learning process) or subsequently, after learning or out-of-the-school settings.

**Evolution of Tests**

According to Alonge in Boris and Awodun (2012), cited in the African Journal of Theory and Practice of Educational Research (AJTPER) Vol. 1,2015, by the year 155BC, Civil Service Examinations were used to select candidates for employment in the Chinese Imperial Service. These examinations were quite strict and aimed at controlling the learner's ideology and supervising their learning. The authors went further to explain that the idea of using test to assess the performance of learners is therefore said to have begun with the China was the first country to appoint civil servants among the Chinese based on their competitive performance in achievement tests.

With the introduction of the western education by the colonial masters in Cameroon, a similar scenario was put in place. The Germans, British and French in Cameroon despite their varied interests, had the aim of bringing out citizens who could easily communicate with them using a common local language (Duala)as in the case of the Germans, rule themselves through local leaders with whom they could relate (British), or make citizens overseas (French). (Epah, 1988)

These varied approaches to colonialism still had as a method of assessment for placement or employment the use of Tests. Examinations were exclusively run by the foreign Examination Boards from the various countries, using the foreign languages. Success in such examinations gave a candidate the possibility of pursuing further studies abroad or be employed at home or abroad.

In the era of independence, these foreign exams were run in line with the local ones organized by the ministries of education. Gradually, the Ministries took over the organization of all public exams and the government stopped the external exams in Cameroon.

Today the setting of tests and running of examinations in Cameroon is done by both the ministries of education and some examination boards. The GCE Board was put in place in 1992 to cater for the examinations of the English-speaking sub-system of education at the level of the secondary education while the Baccalaureate Board was put in place to cater for the examinations of the French speaking sub-system of education still at the level of the secondary education.

Nevertheless, other examinations like the entrance examinations into first year of the secondary schools, the Teachers Grade One Certificate (CAPIEMP) are organized by the Directorate in charge of Examinations and competitive entrance

examinations of the ministry of secondary Education. The other examinations at the level of Basic and Tertiary education are still under their respective ministries.

**Forms of Tests**

-Computer-delivered test

This is a test which is found in the computer. The candidates get access to the test by codes given to them, which they input to let the test pages open. They write their answers through the keyboard and finally submit the test in soft copy by a click on submit to end the test taking session. Any test not submitted within the required time may be cancelled.

-Paper-delivered test

This is the hard copy of the test found on a paper and the responses are written with pen or a pencil. It is otherwise called paper-pen or paper-pencil test. The questions are found on the paper and the responses are also written on either the same sheet of paper or on a separate paper or on both. At the end of the test, the answer sheet or booklet is forwarded for marking.

**Theoretical Framework**

A theoretical framework according to Henning, Van Ransburg and Smith (2004), as stated in Tebit, (2014), "places research in the subject or discipline in which the researcher is carrying out his study, to enable them to be able to bring out theories about the research work and also to make explicit some of the assumptions of the researcher, about the inter-connectedness of the way things are related in the world". It presents the theory which explains why the problem under study exists: how tests may be analyzed to enable creative student development in Biology.

The research work was based on the Classical test theory (CTT), the Item Response Theory (IRT), the true score theory and the Bloom's taxonomy to back the independent variable (test theories), the Autotelic Flow of Miyali Csikszentmihalyi (1996) creates a basic link to the predictive variable and the criterion variable. The Constructivism Theories (Jean Piaget and Lev Vygotsky), the Gestalts and the Discovery learning theory (Jerome Bruner) were linked to back the dependent variable.

"More than seventy years past, Doctor Frederic Lord made the valuable observation that the observed scores of examinee and true scores are not synonymous with ability scores. Ability scores are test independent and therefore more fundamental whereas true scores and observed scores are test-dependent". Lord, (1953).

The basic idea behind it is that those taking the examinations come to a test administration with levels of ability or scores in relation to the construct under measurement using the test. These scores of ability tend to be test-independent. However, the test scores of the examinee and the corresponding true scores always depend on the selection of the tasks in the assessment, from the domain of assessment tasks over which their ability scores are defined.

> "On difficult tests, candidates will have lower true scores and higher true scores on tests which are easier, but their ability scores will remain the same or constant over any tests that might be built to measure the construct. Of course, abilities may change as a result of instruction and other factors, over time, but at the during an assessment, every examinee will have an ability score well defined with respect to the construct, and it stays invariant or unchanged (independent) over the various assessment tasks samples that might be used in the assessment. This illustration demonstrates the fact that

there is a more fundamental concept than a test taker's true score operating. It is this more fundamental concept that is often of interest in most assessments". Lord, (1952).

As an example; in Computer Adaptive Testing, in which test takers may look at tests as of substantially different difficulties, test-dependent scores such as estimated true scores would be of no value in comparing the test takers. Ability scores or estimated ability scores that are independent of the particular choice of test items would be of value because they would permit fair norm-referenced comparisons.

"Lord (1952-1953), and a host of psychometricians before him; Gulliksen, (1950), had been interested in psychometric theories and models that might lead to descriptions of test-takers that would be independent of the particular choice of items or assessment tasks that were used in a test. Also, some psychometricians had the feeling that measurement practices would be enhanced if item and test analysis could be made sample independent too. One step in that direction in classical measurement is the preference of biserial correlations over point biserial correlations in estimating item discriminating power because the former is more invariant over test takers samples than the latter". Lord & Novick (1968).

"Basically, however, classical item analysis such as item difficulty (proportion correct) and item discrimination (point biserial correlations) and test analysis such as test reliability are dependent on the sample of the test takers from which they are obtained. This is not necessarily a problem, and thousands of excellent tests have been constructed in this way (including most important tests, up to towards the end of the 1960s), although special emphasis has been placed on obtaining suitable samples of the test takers for

obtaining item and test analysis and producing statistically parallel tests for assessment".

"Measurement specialists with interest in value invariant item and person analysis, a special solution lies in the models, concepts and methods associated with item response theory. This has been the point made by Lord in his doctoral thesis and published as a psychometric monograph in 1952, and in an article in 1953 (although at the time the domain of study was known as latent trait theory). In 1957 and 1958, Allen Birnbaum wrote a series of technical reports introducing logistic test models and model parameter estimation". Birnbaum, (1968).

"George Rasch, (1960) published a book describing several Item Response Models; one of which later became known as the Rasch Model (or the one-parameter logistic model), with applications to achievement testing. Later in the 1960s, work by Lord (1968), Lord & Novick (1968) and Wright (1968), brought considerable attention to the field of Item Response Theory. Through the 1970s to 1980s, the measurement journals (Applied Psychological Measurement and the Journal of Educational Measurement) were filled with technical innovations and applications. Books and articles from Hambleton, (1989), Hambleton & Swaminathan (1985), Lord (1980), Wright & Stone (1979) and many other articles and books for practitioners that described Item Response Theory models and their applications were written". Loyd (1988).

Despite the fact that item response theory and its related models and applications have been under experimentation for more than 40 years and extensively verified for the past twenty-five years, classical test theory and related models have been researched, experimented and applied ceaselessly and

successfully for greatly more than sixty years, and many testing programmes today remain firmly founded on classical measurement methods and models.

> "From the year 1993, specialists on measurement had to choose between working within a classical test theory framework or an item response theory framework or even a combination of both with other frameworks. The main purposes of this instructional module were; (1) to focus attention on the similarities and differences between classical test theory and item response theory and related models, concepts and methods, and (2) to consider the procedures in which tests are developed within the two statistic-analytical frameworks. Our main intention was to be descriptive, although we had highlighted advantages and disadvantages of each framework when they had been well documented in the measurement literature". Birnbaum (1968).

The main theories of this study are given detailed presentation with regards to the topic, variables, objectives, research questions and hypotheses as follows:

**Classical Test Theory (CTT)**

The classical test theory comes here to support the independent variable; Test structure analysis. This theory, even though less popular in its use in the development of tests in the classrooms in Cameroonian schools, it has been the basis for the development of assessment instruments in many educational institutions and examination boards for many years since its creation. As presented in the pages ahead, this theory gives parameters which a test as an instrument of educational measurement must address. The CTT predicts outcomes of psychological testing such as the difficulty of the items or the ability of the test takers. Novick (1966).

The main components of Classical Test Theory (CTT) (McDonald 1999, pp. 4-8) are: Classical true-score theory, Common factor theory.

The four major statistics examined or reported in the framework of CTT are:

(1) Item difficulty, (2) Item-test correlation, (3) Reliability coefficient and

(4) Standard error of measurement (SEM)

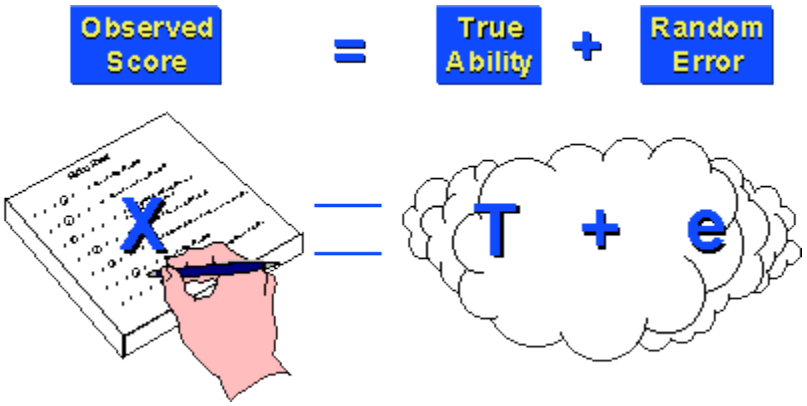
**True Score Theory**

True Score Theory is a theory about measurement. Like all theories, you need to recognize that it is not proven; it is postulated as a model of how the world operates. Like many very powerful models, the true score theory is a very simple one.

Essentially, true score theory maintains that every measurement is an additive composite of two components: true ability (or the true level) of the respondent on that measure; and random error. It is observed that the measurement; the score on the test, the total for a self-esteem instrument, the scale value for a person's weight. We do not observe what is on the right side of the equation (only God knows what those values are!), It is assumed that there are two components to the right side.

Figure 1

*The true score and the observed score equation*



The equation is a simplified but powerful model for classroom measurement.

It is not the only currently used model.

It serves as the foundation of the theory of reliability.

The equation can be used for simulations in the computer to generate scores which were observed with particular properties.

The Theory of True Score

> "The true score equation is represented as X = t + e: X stands for observed score, t for a true score and e stands for the error. The population mean is considered to be zero. The approach of Classical Test Theory to measurement was founded on the true score model. This true score model was based on the observed measurements of a student. The observation was considered as a random variable, comprised of two other main random variables; the true score and the error score, as seen below": Gulliksen, (1950).

"X = T + E shows that X indicates the observed score which has a mean as $\mu_X$ and the as variance $r^2_X$. T which stands for the true score has as the mean $\mu_T$ and the variance as $r^2_T$, finally, E which is the error score has as mean $\mu_E$ and as variance $r^2_E$. The value expected for the error score (E) is 0, which stands for the mean; $\mu_E$ = 0, thus, the value expected for X is equal to the expected value for T. E ($\mu_T$) = E ($\mu_X$)". McDonald (1999).

As observed with the equation presented above, the variance of the observed score indicates the sum of the variance of the true score and the variance of the error score, as shown below:

$r^2_X = r^2_T + r^2_E$

"Considering that true score theory tends to acknowledge measurement error, in most other approaches it generally does not allow different levels of measurement error in view of different ability levels". Lee, Brennan & Kolen (2000).

The use and application of the theory has some assumptions however which include the following:

> "The application of the true and error score model of classical test theory needs the assumption that errors are random and are therefore not correlated with one other and also uncorrelated with the true scores. If those types of errors comprise a combination of several factors, therefore a normal distribution of errors can be predicted with an average or mean equal to zero and an equivalent variance equal to $r^2_E$. Because of random error that occurs, it is further implied that error scores are uncorrelated with any other variables. The assumptions that come up about the error are important in order to interpret the different indicators of reliability. We find out that the classical test theory model does not consider content, or a test item's characteristics however, it makes reference to the relationships between the item and other variables. Therefore, whenever the trait or characteristic is to be measured or if it is measurable by one or more than one test, at that time the equivalent parallel forms of the test or the special form of the test equating methods become necessary for making score comparisons". Gulliksen (1950).

A collection of some of the characteristics of strictly parallel forms of tests comprises of means which are equal, variances and correlations with some other variables. In order to realise such a condition, the properties of the item must be matched across forms. "Reliability is considered as the precision with which a test

score measures achievement. In most cases a higher reliability is desired because we would generally like to have reliability estimates which are greater than 0.85 for tests of high stakes", Holland and Rubin, (1982), for classroom assessment, reliability is accepted when it is greater than 0.7. Ji and Adam (2009, p. 12).Three main methods for estimating the reliability coefficient are recognized, these include:

1. The test-retest reliability (or coefficient of stability)

2. The alternate form or parallel form (that is the coefficient of equivalence)

3. The internal consistency (or coefficient of internal analysis)

**The p-value or item difficulty**

The classical test theory defines the p-values as the proportion of respondents who are responding correctly to, or agreeing with an item, and also those who are used in the determining of an item's difficulty. P-values have a range from 0.00 to 1. The table1below presents some common statistics on item analysis. Columns 2 and 3 of the table1indicate that the p-values and standard deviations have been presented from 30 dichotomously scored, multiple choice, spatial ability items; Spatial Learning Ability Test (SLAT), Embretson (1989), consisting of a sample of 178 persons of ages ranging from 18years to 84years. McCollam (1997) It should be noted that the 10th item seems to be a relatively easy item and item 7 as to be relatively difficult. "Observed p-values are sometimes adjusted for guessing factors in the contexts of multiple choice questions, in the case where the adjusted proportion or ratio of persons passing the item does not seem to be the proportion of people who have the right answer". Crocker and Algina, (1986)

Table 1:

*Sample statistics from 30 items of spatial ability*

| Test item correlation | Item difficulty or p-value | Standard Deviation | Correlation by Point-biserial | Biserial correlation |
|---|---|---|---|---|
| 1 | .567 | .497 | .463 | .583 |
| 2 | .332 | .472 | .319 | .414 |
| 3 | .506 | .501 | .252 | .316 |
| 4 | .371 | .484 | .274 | .350 |
| 5 | .365 | .483 | .420 | .537 |
| 6 | .500 | .501 | .406 | .509 |
| 7 | .253 | .436 | .300 | .408 |
| 8 | .590 | .491 | .366 | .463 |
| 9 | .438 | .498 | .574 | .723 |
| 10 | .725 | .448 | .431 | .576 |
| 11 | .348 | .478 | .441 | .568 |
| 12 | .697 | .461 | .467 | .614 |
| 13 | .365 | .483 | .405 | .519 |
| 14 | .270 | .445 | .211 | .284 |
| 15 | .242 | .429 | .158 | .217 |
| 16 | .258 | .439 | .243 | .329 |
| 17 | .714 | .453 | .425 | .564 |
| 18 | .573 | .496 | .324 | .409 |
| 19 | .444 | .498 | .386 | .486 |
| 20 | .348 | .478 | .427 | .550 |
| 21 | .399 | .491 | .278 | .353 |
| 22 | .449 | .499 | .392 | .492 |
| 23 | .365 | .483 | .231 | .296 |
| 24 | .534 | .500 | .418 | .525 |
| 25 | .781 | .415 | .335 | .469 |
| 26 | .449 | .499 | .378 | .475 |
| 27 | .438 | .498 | .496 | .624 |
| 28 | .371 | .484 | .336 | .429 |
| 29 | .500 | .501 | .377 | .473 |
| 30 | .584 | .494 | .344 | .435 |

Note that Kuder-Richarson 20 = 0.848.

## Correlation between item and test

The discrimination on an item and the item to total score correlations: “Item discrimination can be said to be the correlation between a test item and the total

score. Conceptually, good item discrimination is of evidence when passing an item which positively correlates, with overall individual scores. When vigilant, negative discriminations would indicate faulty item design and items would be modified, eliminated or verified for possible differences in response strategy" Schmidt (1998). Considering the Classical Test Theory, for a set of responses scored dichotomously to a test designed unidimensionally, the effectiveness of an item at discriminating within learners is most of the time acknowledged by point-biserial and biserial correlations.

Biserial correlations are calculated using the assumption that a continuous latent variable which is normally distributed defines the item response. The fourth column of Table1shown above, presents biserial correlations. We can see that the ninth item gives a demonstration of a relatively good discrimination, while item 15 portrays a relatively poor discrimination. Some of the reasons why we do item differences in terms of efficiency are; items which are poorly worded, response strategy differences, weak experimental item design, bias and the multidimensionality of the item.

Point-biserial correlation on the other hand is a direct 'Pearson product moment correlation' of the item response with the total score of the test, for this reason, it is readily available in educational or statistical programmes called standard statistical programme packages like the SPSS or SAS. We however understand that the item difficulty influences the point-biserial correlation such that items that are either very easy or very hard would appear less discriminating. The biserial correlation is slightly higher than the point-biserial correlation for items of moderate difficulty which range from 0.20 to 0.80 but gradually becomes increasingly bigger as item difficulty becomes more extreme.

For example as seen on Table1, item 18 and item 25 have around the same point-biserial correlation but item 25 seems much easier than item 18. We realise that the biserial correlation coefficient for item 25 is 0.469 while the same biserial correlation for item 18 gives 0.409.

Many other classical test theory indicators of discrimination are available for the special case of criterion classification where individuals are placed into different categories from cut scores on test totals. These indicators include the discrimination index, tetrachoric correlations and phi correlations.

The discrimination index calculates the differences in p-value by groups defined from cut scores. Major differences in the p-values among groups give an indication of high item discrimination. Correlations of Phi are product moment correlations with dichotomized total scores and items.

The phi correlation like point-biserial correlations is influenced by item difficulty in such a way that the maximum correlation possible is often substantially below or less than 1. For tetrachoric correlations, we consider them as estimates of the product moment correlation when the two criteria concerned and the items were continuous variables. As seen in the biserial correlation, tetrachorics could be much higher than the direct product-moment correlation coefficient or phi for items which had extreme difficulties or found in samples which were small. Both the phi and the tetrachoric correlations can be calculated to intercorrelate items. Correlations as such may be useful in factor analysis for the determination of the number of dimensions which are measured by a set of items.

Usually, tetrachorics are given more preference over phi correlations due to the fact that they seem not to be biased by extreme p-values. Factor analysis using

tetrachoric correlations can however result in violations in statistics like having negative factor variance estimates.

There exist generalizations of the biserial and the tetrachoric correlations for items that are polytomously scored. Such polytomously scored test items contain multiple categories such as the ones that appear in rating scale items. In the context of ability measurement, the item responses reflect the relative degree of acceptance, or partial correctness in relation to the construct rather than absolute correctness or acceptance in relation to dichotomous test items.

McDonald, (1999) presented that "polyserials are correlations of continuous variables such as total scores with categorical variables like the items. Polychoric correlations are done with two continuous variables divided into at least three separate categories".

We realized that 'adjustments to the item-total correlations can be made by the elimination of the score of the item from the total score in the calculation. This type of correction is called the corrected item-total correlation. The need for such adjustment diminishes in consideration of a number of items relatively large and equally discriminating items' Crocker and Algina (1986).

**Standard error of measurement (SEM)**

> "SEM is defined as the standard deviation of errors of measurement which are considered to be associated with the test scores gotten from a particular group of test takers. SEM could be of help when we want to express the unreliability of individual test scores in a way that is understandable. When we use SEM to compute confidence bands within obtained test scores, some score bands can be used also to explain intra-individual as well as inter-individual differences in the scores. When using approximations when

scores bands are rightly computed, the interpreters become conscious over over-interpretation. We therefore recommend that we use SEMs at various score levels in calculating score bands instead of a single SEM value" Harvill, (1991).

**Limitations of CTT**

Classical test theory is both sample dependent and test dependent McDonald (1999, p.9).

Methods of classical test have been seen to be of great use and are also used widely in test construction and analysis process among practitioners. Hambleton and Jones, (1993), pronounced that, 'the both classical item statistics which are item discrimination and item difficulty which make up the foundation stones of most classical test and item analyses, are dependent on the group (i.e. on the sample). Thus, the p-value and d-value or r-values all depend on the sample of students' from where they are gotten.

When we consider discrimination indices, we find out that higher values tend to be gotten from heterogeneous samples and lower values from homogeneous samples. In the same light, when we look at item difficulty indices, higher values will be found from the test takers' sample of above-average ability and lower values from samples of test takers of low or below-average ability. "That type of sample dependency relationships reduces the overall importance of these analyses" Schumacker (2010).

Classical test theory has another weakness which is that the applications of it are test-based or test dependent. "The difficulty of a test directly affects the resultant test scores. Scores of higher knowledge are directly associated with tests composed of relatively easy items, and scores of low knowledge can be attributed

to a test composed of items that are more difficult. The true score model upon which much of classical test theory is based permits no consideration of the candidates' responses to a specific item. Therefore, basis does not exist that predict how a given candidate will perform on a particular test item" Schumacker (2010). That implies that the ability of the candidate depends on the test item difficulty.

Courville, (2004), scribbled that classical test reliability is an indicator of the quality of a set of test scores; thus, reliability is said to be dependent on the characteristics of the group of test takers, adding to being dependent on the characteristics of the test and the test administration.

A limitation to classical test theory could also be that to compare the performance of different test takers, they must be given the same or parallel items. One other worry of classical test theory could be the inability of it to provide basis for determining how a test taker, in a given group of persons might perform when faced with test items.

Summarily, classical test theory has as assumption the fact that the measurement error could be the same for all test takers or examinees. This is because of the criticisms laid upon classical test theory which has made some test developers to have turned to item response theory.

While trying to employ the standard test development techniques of classical test theory, item analysis is made up of: - determining item parameter that is sample-specific by the use of simple mathematical techniques and moderate sample sizes and also deleting items on the bases of statistical criteria. Standard item analysis techniques include the assessment of item difficulty and discrimination indices and item distractors.

**Selection of an item**

Items are selected on the basis of two characteristics, in addition to concerns for content validity, when applying test development techniques; these are item difficulty and item discrimination. We always make an attempt to choose items with the highest discrimination parameters. Choosing the level of difficulty in a test is mostly guided by the purpose of the test and the planned ability distribution of the group of learners for whom the test is programmed.

An example may be the condition that the purpose of the test stands to select a small sample of high-ability test takers for the purpose to award of a scholarship. Considering this example, items could generally be selected that tend to be quite difficult for the group at large. Some of the norm-referenced achievement tests tend to be commonly designed to differentiate the candidates with respect to their competences in the areas said to be measured; that is, the test is said to be designed to yield a broad range of scores maximizing discriminations among all the candidates taking the test. If a test is designed for the said purpose, items tend to generally be chosen to have a medium level and narrow range of difficulty.

Steps in test development

1. The building up of the test specifications,
2. Arrangement of the test item pool,
3. Testing the items in the classroom,
4. Revise the test items for syntax,
5. Development of the test,
6. Pilot testing exercise of the test,

7. Finishing of the test development exercise,

8. Administration of the test to have norming and technical data,

9. Establishing technical analyses like compilation of norms, standards setting, scores equating, validity and reliability studies,

10. Drawing up of administrative instructions and technical manual,

11. Editing, printing and distribution of the said tests and manuals.

Steps 3, 5, and 9 present where important differences occur in test development using classical test theory and item response theory.

Item statistics depends to a great extent on the characteristics of the candidate sample used in the analysis. An important concern of test developers applying classical test theory is that the test takers sample should be representative of the overall group of persons for whom the test is made.

Heterogeneous samples would generally result in higher estimates of item discrimination indices as measured by point-biserial or biserial correlation coefficients, whereas item difficulty estimates rise and fall with high and low ability students, respectively.

Notwithstanding the difficulty encountered in obtaining a sample representative enough, an advantage of this way to item analysis is that item statistics can be accurately labeled on test taker samples of modest size. To get the detection of poor items for at least norm-referenced tests is quite straightforward and is mainly accomplished by a careful study of item statistics.

An item seen as a poor item is identified through an item difficulty value that is either very high or very low, or a low item-total score correlation. It would be very right to indicate that classical item analysis procedures, together with the

distractors analysis, have the ability to offer the test developer with invaluable knowledge concerning test item quality, regardless of which measurement model was used in the later stages of test development.

**Item Response Theory (IRT)**

The methods of item response theory have two separate traditions. The IRT is often traced to Lord and Novick's 1968 classic book on measurement in the United States of America. Coming before this volume was Lord's 1953 monograph on models of test theory and Birnbaum's methods on the development of estimation.

"The IRT methods also known as latent trait theory in Europe was first introduced by Georg Rasch in 1960. He was a Danish mathematician; he developed a model now known as the 'Rasch model' for a historical statement about IRT" Bock (1997). Since the introduction of the IRT models, test procedures and estimation procedures have rapidly developed. IRT can however include the classical test theory (CTT) as a special case based qualitatively on different principles.

> "The item response theory was first proposed in the field of psychometrics, with the objective for ability assessment. IRT is widely used in education to calibrate and evaluate items in tests, questionnaires, and other instruments and to score tests on their attitudes, abilities, or other latent traits. During the past several decades, educational assessment has used much more IRT-based techniques for tests development. Nowadays, almost all major educational tests like the Graduate Record Examination (GRE) and the Scholastic Aptitude Test (SAT) are developed through the use of item response theory because the methodology could significantly improve

accuracy in measurement and reliability while providing potentially significant reductions in assessment effort and time, especially through computerized adaptive testing (CAT). Of recent, the IRT-based models have also become increasingly popularly used in clinical research, quality-of-life research, and health outcomes" Hays, Morales & Reise (2000).

**The Rasch Model**

We can clearly indicate the components affecting the probability that any individual will give reaction in a particular way to a particular item from a perspective of IRT. A model of measurement expresses the mathematical links between an outcome like a respondent's score on a particular item and the components that impact on the outcome like the qualities of the respondent and qualities of the item.

"Different kinds of models have been developed from the view of the IRT and these models are different from each other in basically two important ways. One of the important differences among the measurement models is in terms of the item parameters or characteristics that are included in the models. Another or second important difference among measurement models is in terms of the format of response option.

The simplest IRT model known today is called the Rasch model or the one-parameter logistic model. Following this said Rasch model, an individual's responses to binary items like the right/wrong, true/false, agree/disagree, is determined by the trait level of the individual and the difficulty of the item. A common way to express the Rasch model is in terms of the probability that an individual with a level of a particular trait will give the correct answer to an item that has a specifically given difficulty. Embretson & Reise, (2000) often presented this as:

$$P(X_{is} = 1|\theta s , \beta i) = \frac{e^{(\theta s - \beta i)}}{1 + e^{(\theta s - \beta i)}}$$

The equation is given the required explanation as follows:

$X_{is}$ represents the response (X) made by subject s to item i.

θs stands for the trait level of the subject s.

βi makes reference to the item difficulty of item i.

$X_{is} = 1$ stands for the right response or a validation of the item.

e is the natural logarithm base of e = 2.7182818 found on most scientific calculators.

Therefore, $P(X_{is} = 1|\theta s , \beta i)$ signifies the probability (P) that subject s will correctly respond to item i. The vertical bar in the statement signifies that that is a conditional probability. The probability that the respondent will respond correctly to the item depends on the subject's trait level (θs) and the item's difficulty (βi). Analysis in an IRT indicates that trait levels and item difficulties are usually scaled on a standardized metric so that their means or averages equal to 0 and the standard deviations equal to 1. Consider these examples in terms of a Biology test.

1- Give the probability that an individual who has an above-average level of Biology ability say a level of Biology ability that is 1 standard deviation above the mean, θs = 1, will answer correctly an item that has a relatively low level of difficulty say $\beta i = -0.5$)?

$$P = \frac{e^{(1-(-.5))}}{1 + e^{(1-(-.5))}} = \frac{e^{(1.5)}}{1 + e^{(1.5)}} = \frac{4.48}{1 + 4.48} = .82.$$

This gives the indication that there is a 0.82 probability that the individual will answer correctly the given item. In other words, there is a high probability; greater than an 80% chance that the individual will correctly

answer the item. This should make great sense because an individual with a high level of ability has been responding to a relatively easy item.

2- Calculate the probability that a student who has an average below the ability level of Biology like a level of ability in Biology that is 1.39 standard deviations below the average or mean, $\theta s = -1.39$ will give the right answer to an item that has a comparatively low level of difficulty like $\beta i = -1.61$)?

$$P = \frac{e^{(-1.39-(-1.61))}}{1 + e^{(-1.39-(-1.61))}} = \frac{e^{(.22)}}{1 + e^{(.22)}} = \frac{1.25}{1 + 1.25} = 0.56$$

The result shows that we have a 0.56 probability that the candidate will give the correct answer to the item. To interpret in other words, it seems to be slightly more than a fifty-fifty chance that the candidate will correctly answer the item. The result should make reasonable sense due to the fact that the candidate's trait level $\theta = -1.39$ is just slightly higher than the difficulty level of the item; $\beta = -1.61$. As a reminder, we should note that the item difficulty level stands for the trait level at which a student will have a fifty-fifty probability of giving the right answer to the item. Due to the fact that the student's trait level is slightly greater than the item's difficulty level, the chance that the student will give the right answer to the item is slightly greater than 0.50.

A relatively more complex model of the Item Response Theory is called the two-parameter logistic model because it is made up of two item parameters. In accordance with the two parameter logistic (2PL) model, a student's response to a binary item can be determined by the individual's trait level, the item difficulty, and the item discrimination. A major difference between the 2PL and the Rasch model is in the putting in of the item discrimination parameter which can be presented as:

$$P(X_{is} = 1|\theta s\ ,\ \beta i\ ,\ \alpha_i) = \frac{e^{(\alpha i\ (\theta s - \beta i\ ))}}{1 + e^{(\alpha i\ (\theta s - \beta i\ ))}},$$

In which case $\alpha_i$ represents the discrimination of item i, with higher values that represent more discriminating items. The two parameter logistic model says that the probability of a respondent giving the right answer to an item is conditional upon the respondent's trait level ($\theta s$), the item's difficulty ($\beta i$), and the item's discrimination ($\alpha_i$). Let's reconsider the items "How much calories are in two eggs?" and "What is the energy level in an orange?" Assuming that the two items have equal difficulty level (say, $\beta = -0.5$). Then we also consider that both have different discrimination values say $\alpha_1 = 0.5$ and $\alpha_2 = 2$.

What would be chance that Wase who has an above average level ability in Biology say, a level of Biology ability that is 1 standard deviation above the average, $\theta = 1$ will give the correct answer to item 1?

$$P = \frac{e^{(0.5(1-(-0.5)))}}{1 + e^{(0.5(1-(-0.5)))}} = \frac{e^{(0.75)}}{1 + e^{(0.75)}} \quad \frac{= 2.12}{1 + 2.12} \quad = 0.68.$$

What will then be the probability that Belle, who has an average level ability in Biology ($\theta = 0$), will give the right answer to item 1?

$$P = \frac{e^{(0.5(0-(-0.5)))}}{1 + e^{(0.5(0-(-0.5)))}} \quad \frac{= e(0.25)}{1 + e(0.25)} \quad = \frac{1.28}{1 + 1.28} \quad = 0.56 \ .$$

Identify what makes the difference. Wase's level of Biological ability is one standard deviation greater than Belle's, but her chance of answering the item rightly is only 0.12 greater than Belle's. This presents a relatively huge difference in the trait level that is one standard deviation apart, but a seemingly little difference in the likelihood of giving the right answer to the item.

Let's now take into consideration the probabilities that Wase and Belle Likely will give the right answer to item number two (2).

Wase: $P = \frac{e^{(2(1-(-0.5)))}}{1 + e^{(2(1-(-0.5)))}} = \frac{e^{(3)}}{1 + e^{(3)}} = \frac{20.09}{1 + 20.09} = 0.95$

Belle: $P = \frac{e^{(2(0-(-0.5)))}}{1 + e^{(2(0-(-0.5)))}} = \frac{e^{(1)}}{1 + e^{(1)}} = \frac{2.27}{1 + 2.27} = 0.73$

The difference in responses for item 2 should be noted. Wase has got 0.95 probability of giving the right answer to the item, while Belle has got a lesser score of 0.73 as a probability of giving the correct answer to the item. The disparity between the students' ability in Biological is still one standard deviation, but Wase's probability of giving the correct answer to item 2 is 0.22 greater than that of Belle. Compared with item one (1), we find out that the item 2 being the item with the greater value of discrimination, draws up a clearer distinction among candidates having different trait levels. Like seen, the two parameter logistic (2PL) model is an extension of the Rasch model (that is the one parameter logistic (1PL) model), however, there are other models that are expansions of the 2PL model. We cannot be taken by surprise to find out that the three-parameter logistic model (3PL) puts in another totally different item parameter. A discussion of this model will not be done other than to bring out the fact that the third parameter is considered as an adjustment for guessing. Summarily, the 1PL model, 2PL model and 3PL model represent Item Response Theory measurement models that are different with respect to the number of item parameters that are included in the models. There is a minimum of one additional way in which IRT measurement models diverse from one another.

Another alternative way in which IRT models differ is in terms of the response option format. To this extent, we have talked about 1PL, 2PL and

3PL models that have been designed to be used for binary outcomes as the response option. Many tests, questionnaires and inventories in the behavioral sciences however, include more than two response options. As example, many personality questionnaires include self-relevant statements like; "I enjoy having conversation with friends" and respondents are given more than three or just three response options like *strongly disagree, disagree, neutral, agree, strongly agree*. These types of items are known as a polytomous items, and they require models of IRT that are different from those demanded by binary items. Such models as the graded response model Samejima (1969) and the partial credit model) are polytomous IRT models. Despite the fact that these models differ in terms of the response options that they can be able to accommodate, they are founded on similar general principles as the models designed for binary items; which is to say that, they reflect the idea that a candidate's response to an item can be determined by the person's trait level and by item properties like discrimination and difficulty. The Rasch model is said to be one of the most greatly used models of the IRT in the various applications of IRT". Masters (1982

**The Basic IRT Models Extended**

The early IRT models like the Rasch model and the two-parameter model, concentrate basically on the analyses of dichotomous responses which have a unique or single latent trait. Diverse expansions of these basic IRT models have been developed for better and more flexible modeling in different situations. The list that follows puts forward some generalized or extended IRT models and their capabilities:

- Graded response models (GRM) which analyze rating scales and ordinal responses

- Three and four parameter models which analyze test items that have ceiling and guessing parameters in the response curves

- Multidimensional IRT models which analyze test items that can be explained by more than one factor or latent trait

- Multiple groups IRT models which analyze test items in independent groups to study invariance or differential item functioning

- Confirmatory IRT models which analyze test items that have hypothesized relationships with the latent factors

These extensions or generalizations of the IRT models are not mutually exclusive. They can be combined to address the complexity of the data and to test the substantive theory in practical applications. De Ayala (2009) and Embretson & Reise (2000).

This item response theory or latent trait theory or strong true score theory is basically aimed for the design, analysis and scoring of tests or questionnaires. It relates individual's performances on a test item and the test taker's level of performance in relation to the other students Frederic, Rasch & Lazasfeld (1950-1960s).

"Item response theory is generally a theory of statistics centred about test taker item and test performance, and how performance is related to the abilities that are measured by the items in the given test. Item responses theory can take discrete variables or continuous variables and can be scored dichotomously or polychotomously; the categories of the item score can be unordered or ordered; there can be many abilities or one ability that underly the test performance and there are many ways or models in which the relationship between item responses and the underlying competence or competences can be specified. Amidst the

general structure of IRT, many models have been formulated and applied to real test data". Hambleton (1989).

**Analysis of an item**

In the process of employing item response theory, we find that item analysis involves; the determination of sample-invariant-item parameters employing relatively complex mathematical techniques and large sample sizes and making use of goodness-of-fit criteria to detect an item that does not fit the specified response model. The property of sample-invariance found within IRT signifies that those who develop the test do not need a sample that is representative of the test taker population to calibrate the test items. However, they do need a heterogeneous and huge test taker sample to insure proper estimation of item parameter. Even when the test taker samples differ, the developer of the test is able to use the principles inherent to IRT to give the estimate the same ICC regardless of the test taker sample used in the item calibration process. The developer of a test using IRT is however, faced with another different difficulty. Due to the fact that IRT requires bigger sample sizes in order to obtain good item parameter estimates, the developer of test must ensure that the sample of test taker is sizable enough to guarantee accurate item calibration.

The procedure for detection of poor items in a test using the item response theory is not as straightforward as when using the classical test theory. Test items are evaluated generally in terms of their goodness-of-fit to a model using a statistical test or an analysis of residuals. It is important to emphasize that an adequate fit of model-to-data is essential for successful item analysis; otherwise, items may appear poor as an artifact of poor model fit. Hambleton, Swaminathan, & Rogers (1991) and Hambleton & Murray (1983).

Poor items are usually identified through a consideration of their discrimination indices (the value of ai will be a low positive or even negative) and difficulty indices (items should be neither too easy nor too difficult for the group of examinees to be assessed).

**Item selection**

As is the case with classical test theory, item response theory also bases item selection on the intended purpose of the test. However, the final selection of items will depend on the information they contribute to the overall information supplied by the test. A particularly useful feature of the item information functions used in IRT test development is that they permit the test developer to determine the contribution of each test item to the test information function independently of other items in the test. Lord (1977) outlined a procedure, originally conceptualized by Birnbaum (1968), for the use of item information functions in the test building process. Basically, this procedure entails that a test developer takes the following four steps:

1- Describe the shape of the desired test information function over the desired range of abilities. Lord (1977) calls this the target information function.

2- Sort out items with item information functions which will cover up the hard-to-fill areas found under the target information function.

3- After the addition of each item to the test, compute the test information function for the test items selected.

4- We keep on sorting out test items up to the point that the test information function approximates the target information function to a satisfactory degree. Example of this is; suppose a developer of test wished to build a test to fill the target information function.

The step number one would be to specify the given information function as the target information function. The step that followed would be to include an item to the test, covering up the hard-to-fill areas first. The test information function should be calculated again after the inclusion of each item in order to bring out the specific information needed to complete the test.

Through the following of this procedure, the developer of test would create the ideal test to match up the target information function. Considerations of content validation are checked through monitoring during the process of item selection. Such a procedure gives room for the test developer to construct a test that will precisely fulfill any set of desired test specifications. It is therefore possible to construct a test that "discriminates" well at any particular region on the ability continuum. That is to say, if we have a good idea of the ability of a group of examinees test items can be selected so as to maximize test information in the region of ability spanned by the examinees being tested. Of course, this optimum selection of test items will contribute substantially to the precision with which ability scores are estimated.

Furthermore, with criterion-referenced tests, it is common to observe lower test performance on a pretest than on a posttest. Given this knowledge, a test instructor should select easier items for the pretest and more difficult items for the posttest. Then, for both testing administrations, measurement precision will have been maximized in the ability region where the test taker would most likely be located.

Moreover, because items on both tests measure the same ability, and ability estimates are independent of the particular choice of test items, the instructor can measure growth by subtracting the pretest ability estimate from the posttest ability estimate.

Item response theory directly underlies the independent variable of the study; test structure analysis, specifically in the test item analysis and test distractors analysis.

**The Taxonomy of Learning Domains; Bloom (1956)**

Doctor Benjamin Bloom an educational psychologist in collaboration with Edward Furst, Max Englehart, David Krathwohl and Walter Hill, (1956), together wrote and published a structural framework on how to categorize educational objectives: Taxonomy of Educational Objectives. This framework which is familiarly known as Bloom's taxonomy came to the lime light for the purpose of promoting higher ways of thinking or thought in education like the analysis and evaluation of processes, procedures, principles and concepts instead of just rote learning or remembering of facts. This taxonomy turned out to be one of the best-known approaches to classifying questions. Since this study has as preoccupation; the analysis of test structure, the researcher finds this taxonomy appropriate in addressing the test development issues. It has served a guide in the development of the biology test attached to the appendix. The three domains of the taxonomy are presented below.

**The Three Domains of Learning**

Bloom and his committee (1956) identified three domains of educational activities or learning. These domains according to them were:

- **Cognitive**: mental skills

- **Affective**: growth in feelings or emotional areas

- **Psychomotor**: manual or physical skills

Found in every domain are multiple levels of learning that evolve from more basic, surface level learning to more complex, deeper level learning. This level of learning that a teacher strives to inculcate on learners will differ across learning experiences that depend on:

- The characteristics or nature of the experience

- The levels of developmental of the learners participating and

- The intensity and duration of the experience.

It should be noted here that the above domains/categories are still the same as what instructional trainers, educators and designers often refer to as KSA which stands for 'Knowledge for cognitive, Skills for psychomotor, and Attitudes for affective'. The taxonomy of learning behaviours may be considered of as the goals of the learning process. Meaning that, after an episode of learning, the learner should have gained a new knowledge, attitude and skill.

Whereas the constituted group produced an explicit compilation to foster the affective and cognitive domains, it left out the psychomotor domain. This was explained to be because they had much lesser experience in teaching manual skills inside the college class since these were based at the higher education level. However, there have been research by other psychologists and researchers that resulted in their coming up with psychomotor models.

It is important that when writing learning objectives, one should first think about and know which domain(s) is relevant to the learning experience that is being designed.

**A- The Cognitive Domain of Learning**

The cognitive domain involves knowledge and the development of intellectual skills (Bloom, 1956). It deals with how we acquire, process, and use knowledge. It is the "thinking" domain and includes the recognition or recall of procedural patterns, concepts and specific facts that serve in the development of intellectual competences and skills. We have six major categories of cognitive processes, beginning with the simplest to the most complex. Benjamin Bloom (1956) and his colleagues presented six levels of cognition; these are:

Knowledge; recall, recognition or rote memorization of facts or ideas

Comprehension; making cognizance or understanding of the meaning of the facts

Application; appropriate usage of the ideas facts or rules provided or available

Analysis**;** sorting out or breaking down into component parts, the information gotten

Synthesis**;** putting together or combining ideas, information or facts to make a new whole

Evaluation**;** giving value judgment or proposing an opinion towards the given information or situation

Table 2

*Guiding questions per cognitive level*

| Taxon | Attributes | Guiding words | Questions/items |
|---|---|---|---|
| **Knowledge** | Shows previously learned content by facts recalling, terms, basic concepts responses. | who, what, why, when, omit, where, which, choose, find, how, define, label, show, spell, list, match, name, relate, tell, recall, select | What is ...? How is ...? When did _____ happen? How would you explain ...? Can you recall ...? How would you show ...? Can you select ...? Who were the main ...? List three ...? Which one ...? Who was ...? |
| **Comprehension** | Understanding of ideas and facts, organizing, comparing, interpreting, stating main ideas | compare, contrast, demonstrate, explain, illustrate, outline, relate, rephrase, translate, summarize, show, classify | Classify the type of ...? Compare ...? State or interpret in your own words ...? Rephrase the meaning ...? What facts, What is the main idea of ...? Which statements support ...? What is happening . . . what is meant . . .? What can you say about ...? Which is the best answer ...? |
| **Application** | Solving problems by applying acquired techniques and rules in a different way | build, construct, develop, interview, make use of, organize, experiment with, plan, solve, utilize, model. | Give the use ...? How would you solve _____ using what you have learned? Organize _____ to show ...? Show your understanding of ...? What approach would you use to ...? What elements would you choose to change ...? What questions would you ask in an interview with ...? |
| **Analysis** | Breaking information into parts by examining and identifying reasons or causes; drawing inferences and finding evidence to support generalisations. | analyse, categorize, classify, compare, contrast, discover, dissect, examine, distinguish, distinction, relationships, function, inference, assumption. | Present the parts of ...? What inference can you make ...? What conclusions can you draw ...? How would you classify ...? How would you categorize ...? Can you identify the difference parts ...? What evidence can you find ...? What is the relationship between ...? Make a distinction between ...? What is the function of ...? Justify ...? |
| **Synthesis** | Gathering information in a different way by combining elements in a new arrangement or proposing other possible solutions | build, combine, compose, construct, create, design, develop, estimate, formulate, invent, originate, propose, solve, discuss, modify, improve, adapt, minimize, maximize, delete, theorise, elaborate, test, improve | Solve ...? Improve ...? Elaborate on the reason ...? Propose an alternative ...? Invent ...? How would you adapt _____ to create a different ...? What could be done to minimize (maximize) ...? How would you design ...? What could be combined to improve (change) ...? How would you test ...? Can you formulate a theory for ...? How would you estimate the results for ...? Can you construct a model that would change ...? Can you think of an original way for the ...? |
| **Evaluation** | Defending and presenting opinions by making judgmental decisions about data, validity/reliability of ideas or quality of work based on a set of criteria | conclude, criticize, decide, determine, dispute, evaluate, judge, justify, measure, compare, rate, recommend, agree, interpret, appraise, prioritize, opinion, ,support, criteria, prove, assess, estimate, influence, deduct | With the outcomes ...? What is your opinion of ...? How would you prove ...? Can you assess the value or importance of ...? What would you recommend ...? How would you evaluate ...? How could you determine ...? What choice would you have made ...? What would you select ...? How would you prioritize ...? What judgment would you make about ...? How would you justify ...? What data was used to make the conclusion ...? Why was it better that ...? How would you prioritize the facts ...? How would you compare the ideas ...? |

*Source: www.learningtheories.com*

This table elaborates the key words used at each level of the taxonomy and the question tag format for easy setting.

Table 3

*Given examples of elements on the Bloom's taxonomy*

| Taxon | Attributes | Keywords | Sample objective | Sample activity | Sample item |
|---|---|---|---|---|---|
| **Knowledge** | Rote memorization and recognition or recall of facts | recite, list, define, match, name, quote, identify, recognize, label, recall, | 'By the end of the lesson, the learners will be able to define photosynthesis' | Students are grouped-up to perform simple experiments to the class presenting how photosynthesis occurs. | This question on an exam or homework can be used. "list three factors that control photosynthesis." |
| **Comprehension** | Understanding the meaning of the facts. | explain, describe, restate, summarize, interpret, discuss, paraphrase, | "By the end of this chapter, the student will be able to restate three principles of transpiration in their own words." | Pair students and have each pair think of words that define transpiration. After a few minutes the pairs volunteer some of their definitions and present them on the board. | A student is assigned to write a simple phrase that explains what transpiration means in her own words |
| **Application** | Use of the, rules, facts or ideas correctly | predict, solve, apply, illustrate, use, demonstrate, calculate, determine, model | "By the end of the week, the student will be able to calculate the number of molecules per substance in the equation of photosynthesis." | After presenting the photosynthesis equation in class, the students pair up for just a few minutes and practice it so that they feel comfortable with it before calculating. | During a test, present an equation and ask they "Calculate the number of molecules per element." |
| **Analysis** | information broken down into components | classify, outline, break down, categorize, analyze, diagram, illustrate | "By the end of this course, the student will be able to differentiate between respiration and gaseous exchange. | Present the content to the students and ask the them to categorize some main words as either respiration or gaseous exchange then have them explain in details why they categorized them the way they did. | Students are given a home work to outline the basic principles of respiration and gaseous exchange, point out the differences between the two and their relatedness. |
| **Synthesis** | Combination of parts to make a new whole | formulate, modify, build, invent, compose, generate, derive, develop design, create, | By the end of the chapter, the student will be able to design an original homework problem dealing with the principle of conservation of energy in plants." | each discussion is tied to the previous knowledge, in order to help the students join all the special classroom sessions into a unified topic/theory. | The students are given a project in which they have to design an original homework problem that deal with the principle of conservation of energy in plants. |
| **Evaluation** | Making judgment on the worth or value of information or ideas | support, choose, determine, relate, defend grade, compare, , judge, contrast, argue, support, convince, justify, select, evaluate | "By the end of the term, the student will be able to determine whether using conservation of energy would be more appropriate for feeding or reproduction or both" | Different groups of students solve the same problem using various methods and have each group present the pros and cons of the method they chose. | During a test, describe a system and ask the students which method they would use to solve the problem and why |

Source: www.learningtheories.com

The elaboration of the objectives, activities and assessment needed at each level of the taxonomy for creative students development.

## Writing Objectives Using Bloom's Taxonomy

The cognitive domain of Bloom's taxonomy gained its popularity and stood the test of time from 1956 till the mid-1990s when Lorin Anderson, a former student of Bloom, and David Krathwohl revisited the cognitive domain in the mid-nineties and made some changes. One of the most glaring of these changes was changing the names in the six categories from noun to verb forms. Today the renaming of the six levels of the taxonomy and reviewing of them gave the following:

## Presentation of the Original Bloom's Taxonomy and the Revised Form

1. Remember implies recognizing and recalling
2. Understand deals with exemplifying, classifying, interpreting, summarizing, explaining, inferring, comparing
3. Apply involves executing, implementing
4. Analyze expects differentiating, organizing, attributing
5. Evaluate involves checking, critiquing or criticizing
6. Create entails generating, planning, producing

'Remember' stage or level has replaced 'knowledge', 'evaluate' has replaced 'synthesize' and 'create' has replaced 'evaluate' at the 6$^{th}$ and highest level of the cognitive taxonomy. This study clarifies that the students are to be creative in their studies and subsequently become creative in their jobs.

Table 4

*Verbs on the revised cognitive domain*

| TAXON/LEVEL | SAMPLE VERBS |
|---|---|
| **Remembering**: Retrieve or recall previous learned information. | **Examples**: Recite procedure. State elements involved in the photosynthetic process.<br><br>**Key words**: describes, identifies, labels, lists, matches, outlines, recalls, recognizes, names, defines, reproduces, selects, knows, states |
| **Understanding**: Comprehension of the meaning, interpolation and interpretation of problems. State problem in your own words. | **Examples**: Explain in your own words the steps involved in a complex task.<br><br>**Main words**: comprehend, convert, defend, distinguish, estimate, explain, generalize, infer, interpret, paraphrase, summarize, translate |
| **Applying**: Use a concept in a new context. Apply what was learned in the classroom into new situations in the field. | **Example**: Apply laws of statistics to evaluate the reliability of a written test.<br><br>**Key words**: change, construct, demonstrate, discover, manipulate, modify, predict, prepare, produce, relate, show, solve |
| **Analyzing**: Separate material into component parts so that its organisational form can be understood. Differentiate between facts and inferences. | **Examples**: Recognize logical fallacies in reasoning. Gather information from a department and select the required tasks for training.<br><br>**Key Words**: break down, compare/contrast, diagrams, deconstruct, differentiate, distinguish, illustrate, infer, outline, relate, separate |
| **Evaluating**: Make judgments about values of ideas or materials. | **Examples**: Choose the most effective equation. Recruit the most qualified candidate. Explain and justify a new product.<br><br>**Key words**: appraise, compare/contrast, conclude, criticize, defend, describe, discriminate, justify, relate, summarize, support |
| **Creating**: Build a structure or pattern from a variety of elements. Assemble to form a whole with emphasis on creating a new outcome or structure. | **Examples**: Design a new machine to perform a specific task. Integrate competences from several sources to solve a problem. Revise and process to improve outcome.<br><br>**Key Words**: combine, compile, compose, devise, design, generate, modify, organize, plan, reconstruct, relate, reorganize, revise, rewrite, summarize |

The table presents verbs for possible achievement of goals if used in the setting of tests.

Krathwohl and Anderson as shown in their taxonomy (revised) also presented a different taxonomy made up of the types of knowledge to be used in cognitive reasoning. These were:

- Factual knowledge; this included knowledge on the terminology and knowledge on the specific elements and details.
- Conceptual knowledge; here, they highlighted knowledge of categories and classifications, knowledge of generalizations and principles, and knowledge of structures, models and theories.
- Procedural knowledge; it comprised of knowledge of subject specific algorithms, skills and knowledge of subject specific methods and techniques, knowledge of the criteria to determine when expected to use the appropriate procedures.
- Metacognitive knowledge; knowledge on cognitive tasks, strategic knowledge, in addition to the convenient or right conditional and contextual knowledge and self-knowledge.

In analyzing test structures, teachers are expected to consider the level of the cognition if the students and use the right word, verb or question tag in the setting of the test questions. This will put the child in the right psychological level ant to be able to do things in their way.

**A) The Affective Domain**

The Affective Domain is one of the three domains in Bloom's taxonomy. The way individuals respond emotionally and their ability to feel the joy or pain of other living things is described by skills found in the affective domain. The objectives in the affective domain typically target the growth and awareness in feelings, attitudes, and emotion. The affective domain incorporates the way in which things are dealt with emotionally by us like values, feelings, appreciation, motivations, attitudes and enthusiasms. These five main categories are presented from the simplest behaviour to the most complex behaviour.

## Categories or Levels of the Affective Domain

**1- Receive:** The learners must be open to experience and willing to hear. Awareness, willingness to hear, selected attention. Examples are; take interest in laboratory session, listen to the lab instructor or trainer, learning experience, take notes, make time for learning experiment and participate actively.

Main words used here include asks, acknowledge, attentive, dutiful, courteous, gives, follows, understands, listens.

**2- Respond**: React and participate actively by the learners. Take part and respond to a given circumstance. Outcomes of learning may put emphasis on compliance in responding, satisfaction in responding or motivation, or willingness to respond. Examples include the participation in classroom discussions, giving a presentation, and questioning new models, ideals and concepts. In a bid to get full understanding of them, know the rules on safety and practice them.

Key words are; assists, answers, aids, conforms, complies, discusses, helps, greets, performs, labels, tells, presents

**3- Value**: The value or worth an individual attaches to a particular phenomenon, behavior or object. This outstretches from simple acceptance to the more complex commitment state. The aspect of valuing puts its base on the internalisation of the set of specified values while clues towards these values are expressed in the student's overt behaviour which is most at times identifiable. Examples are like to decide the worth, value and relevance of ideas experimented to accept or be attached to particular action or stance.

Key words: challenge, debate, confront, argue, refute, persuade, justify, and criticize.

**4- Organise or conceptualise values**: values are organised into priorities through the contrasting of different values, resolution of conflicts among them and creating a unique value system. The emphasis is on relating, comparing and synthesising values such as to quantify and qualify personal views, state personal reasons and position, state self and beliefs.

Key words: develop, defend, formulate, modify, prioritize, relate, reconcile, arrange, contrast, compare and build.

**5- Internalise or characterise values**: In this case the student adopts a philosophy and belief system. The behaviour is consistent, pervasive, predictable and most valuable characteristic of the student. Instructional objectives are dealing with the learner's general adjustment patterns; that is emotional social and personal. Examples**:** Shows independence and self-reliance when doing an experiment, displays teamwork; cooperates in group activities. The use of objective approaches in solving problems, displaying of a professional commitment on a daily basis to ethical practice, revision of judgments and changing of behaviour in light of novel evidence and valuing of humans for what they are, not how they look.

Key words: performs, discriminates, acts, displays, influences, modifies, qualifies, revises, verifies, questions, serves, solves.

Bloom’s affective domain can be a bit trickier to understand for some individuals than some other people. The differences between the various levels, particularlly between number 3, 4 and 5 levels are subtle and not very clear as the separations somewhere else in the Taxonomy.

### B) The Psychomotor Domain

The Psychomotor Domain was ostensibly established to address skills development relating to manual tasks and physical movement. The psychomotor domain lays emphasis on skills which are characterized by imitation, manipulation, development of precision, articulation and naturalization. In the psychomotor domain, skills describe the competence of being able to manipulate physically a instrument or tool like a hand lens or a burette. The Psychomotor objectives most often concentrate on change and development in behaviour and skills.

Subcategories for skills in the psychomotor domain were never created by Bloom and his colleagues. However, other educators have created their own psychomotor taxonomies since then. On the bases of Dave (1970) version of the Psychomotor Domain: 'Developing and Writing Behavioral Objectives', the theory was first put forward at a conference in Berlin in 1967, hence the Dave's model attributed to 1967 or 1970. According to Dave, the following levels are explained.

**1- Imitation:** Here the learners copy the action of another; observe and replicate. Examples of activity or demonstration and evidence to be measured for instance are that a learner can watch the trainer or instructor and repeat the action, activity or process. Key words include verbs that describe the activity to be trained or measured at each level like replicate, copy, follow, adhere and repeat.

**2- Manipulation:** The behaviour described here reproduces activity from instruction or memory. Examples are that the learner can perform a type of task from verbal or written instruction. Key words include build, re-create, perform, implement, execute.

**3- Precision**: Here learners are able to execute skills independent of help and reliably. Students carry out an activity or task with expertise and to high quality

without instruction or assistance; able to demonstrate an activity to other students. Key words are complete, demonstrate, show, control, calibrate and perfect.

**4- Articulation:** At this level the learners integrate and adapt expertise to fulfill a non-standard objective. Examples are combined and related associated activities to develop methods to meet varying, new requirements. Key words; combine, construct, solve, integrate, coordinate, adapt, formulate, develop, master and modify.

**5- Naturalization:** This refers to unconscious, automated mastery of related skills and activity at strategic level. Examples of activities are; approach, defined aim and strategy for use of activities to meet strategic need. Key words are; specify, design, manage, project-manage and invent.

Bloom's taxonomy is a basic consideration in this study in that it is a guide in the selection of words used in the development of test items depending on the test purpose. Despite the elaborate nature of these taxonomies to address the cognitive, psycho-motor and socio-affective domains of the students' learning, many teachers ignore them and just set tests based on the examination syllabuses such that students could easily pass their exams. This study serves as a clarion call on all the teachers especially of biology to consider setting tests following the various domains such that the cognition, socio-affective life and the psychomotor are developed holistically for creative student development.

**The Autotelic Flow theory**

**The Flow concepts**

'The theory Flow has become interesting to researchers in positive psychology like Mihaly Csikszentmihalyi and Jacob Getzels while they were on studies in the creative process during the 1960s'; Getzels and Csikszentmihalyi, (1976). The Flow theory is also considered as the 'zone' and is the mental state of

operation in which an individual carrying out an exercise or activity is fully concentrated, immersed or overwhelmed in a feeling of full involvement, energised focus and enjoyment in the process of carrying out the exercise or activity Wikipedia.(9 jan. 2018).

In education, Flow is an optimal psychological state which individual learner experience while involved or engaged in an exercise or activity which is both appropriately challenging to someone's skill level, sometimes culminating into concentrated focus and immersion on a task. This can bring about or result to high levels of personal satisfaction, deep learning, job satisfaction and personal fulfillment and success.

**Characteristics of Flow**

Eight components of the flow experience are described by Csikszentmihalyi. The first three of them are main prerequisites, while the five others address the subjective experience during an exercise or activity in flow:

1- Goals clarity and direct or immediate feedback or response

As identified in the arts or many sports disciplines, a player of tennis understands exactly what is needed in order to be victorious in a game. The regulations and rules are clearly defined. In any action taken, failure or success is immediately perceived. Biological science like in the arts and sports are classic activities of flow.

2- A high concentration level on a given content of study or field

This gives room for an individual's consciousness to lead them deeply into the exercise or activity. On the contrary, there are oftentimes contradictory and chaotic requests on daily basis which may cause dissatisfaction and confusion.

3- Skills and challenges should be balanced

The right degree of challenge to a student's ability is provided by the difficulty of a task. A very difficult piece of laboratory experiment will leave a biology student disappointed and frustrated, a very easy to do experiment leads to routine and boredom. In this light, flow takes place in range between 'very much' and 'very little'.

The relationship between requirements and capabilities

4- The feeling that one is in control

The feeling of heightened control over one's actions in an exercise is a characteristic of flow. The 'control' expression is misunderstood easily. It can put many students off by its association with nervous attention or compulsive domination. None of these qualities is found in 'control' in flow. It consists of a state of relaxation and security with the total absence of worry; the paradox understood as 'control without controlling' in Zen Buddhism.

5- Effortlessness

Flow constitutes ease and flexibility; everything goes effortlessly and harmoniously. A laboratory experiment or a personal field work performance may look strenuous from the outside but if in fact the student is in flow, he/she experiences no particular strain. The activity runs smoothly, guided by an inner logic. All necessary decisions arise spontaneously from the demands of the activity without any deliberate reflection.

6- Perception of time is altered

Someone's normal perception of the time aspect is on standby while in a deep state of flow. We feel that time has been condensed. A couple of hours is felt like ten

minutes or expanded in a way that a few seconds are felt as if minutes have past: Reasons why we consider flow-mode as ‘timeless’.

7- Consciousness and action are melted together

A state in which there is no room for fear, worry, self-conscious or distraction rumination is created due to complete immersion or involvement in the exercise. Those performing do not feel like they are separated from their exercises or actions; they have the sensation of being united with their performances. This state of being; feeling united with the actions can extend to an individual’s natural environment as well as to a whole population of humans doing team work or working together; termed ‘team flow’.

8- The flow-experiences quality being autotelic.

The greek word ‘autos’ means ‘self’ and ‘telos’ means ‘goal’. Achieving the goal of an exercise or activity is not only rewarding but the activity or exercise on its own is fulfilling. Therefore flow is regarded as ‘immediate return on investment’.

**Csikszentmihalyi's theory of flow**

Flow has been Csikszentmihalyi's best-known theory. Flow is described by the state of engagement and concentration that can be attained when completing or carrying out an exercise or task that challenges someone's abilities. ... In order to be engaged in a state of flow, several conditions must be met.

Three things were defined by Csikszentmihalyi to be what individuals experience while they are in the cognitive flow states:

- Extreme focus on an exercise or task,

- A sense of actively being in control,

- Distortion of the experience of time

Flow can also be defined as an act of running or moving smoothly like a water movement or the continuous movement of stories, ideas, emotions or knowledge. Examples of Flow include a steady arrangement through the construction of a research paper and the movement of the stream.

Key steps towards benefiting and achieving from flow include:

- Select from the job or exercise you prefer or love

- Go for an important activity or task

- Be assured that the task is challenging enough but not very difficult

- Work on your quiet or peak time

- Put aside every aspect of distraction

- Ensure that you get focused on the chosen task for a possible long time

- Get pleasure for yourself

- Don't stop doing practice.

Nine levels involved in the achievement of Flow and happiness in one's learning or work:

- Have you ever been immersed in your own learning or work as much as you losing track of the notion of time? Being overtaken by an activity in such a manner when it can be scarce for most individuals, it is a condition of being called 'flow'.

- From what I have as experience, flow is among the keys to happiness during learning and at work and also a favourable side benefit is that flow does not only lower the stress level but also improves on your output and products.

- Flow is a notion which is too much mentioned and used nowadays and something a majority of us have been part of at a given moment or another.
- Why is flow important here, what is flow and how do we achieve flow on a constant basis for improved productivity and happiness at during learning and at work.

In simple term, flow is a condition of the mind that you attain while you are fully immersed in an activity or task, forgetting about the world outside or around. Flow is a conceptual notion put forward by the positive psychologist Mihalyi Csikszentmihalyi and nowadays we are likely to read about flow on blogs and in many sorts of magazines in many walks of life.

To consider that you are in the flow state, you;

Find yourself completely focusing on the activity carried out or the task at hand

Lose control of yourself; forget about others and about the environment around you

Forget about the notion of time

Content yourself in happiness and stay in control

Finally transform into a creative and productive person.

I love flow because it carries the too ‘Zen concept’ of completely being in the moment and makes it applicable to work activities and tasks. It is a concept that I have discussed about a lot on the Zen habits of ‘being in the moment’, completely focalising on one activity or task and looking for a sensation of happiness and calmness in your learning or work; flow seems just that.

I can conclude that the ability to carry out a single task is among the solutions to the right productivity as contrary to multiple task; the type of productivity in which you clear off 15-20 items from your to-do list; even though

that could be motivating, wherein you engage in switching one task to another throughout the day and keep being busy on a regular basis.

Productivity the true sense of it is the type wherein somebody actually attains their goals, wherein they accomplish valuable and long-lasting projects. Being a researcher, that might mean researching on one or more memorable and important topics instead of 10 or 40 less important cases that persons would tend to forget few minutes after going through them. This implies making key projects work rather than responding to a list of mails, getting so many phone calls made, taking part in a series of meetings and perusing paper works throughout the day. Meaning that the closing up of some major deals and also meaning, quality in the place of quantity.

Having learned to focalise on those types of important projects and tasks, Flow is the manner you get to finalise them. You forget yourself in the challenging and important activities or tasks and in the place of being constantly intercepted by small issues like calls, emails and coworkers; you are now able to focalise on the long enough tasks to actually go through them.

Being immersed in these activities or tasks, one enjoys him/herself more. Stress is reduced when improving on quality products. Important stuff is completed instead of finishing or completing just anything. Things are acquired instead of just keeping oneself busy. Flow is among the keys to all of these.

**Happiness at work in relation to flow**

How can this mystical state of being be achieve? Is there a need to chant or meditate anything? Never; although meditation can improve one's ability to concentrate, one does not need to. Flow is in no way anything mystical, it is really practical and achieving flow is in no way mysterious. It takes practice to reach

flow, but you will get better at it. Below are some major steps to acquiring and achieving gains from flow:

Choose an activity that you like. If you are afraid of an activity or a task, you would have a hard time immersing yourself in it. If your task is composed of issues you do not love, you risk considering to find another task. Or considering seeking programmes you like to do within your current task or work. Be sure that what task or activity ever that you choose will be something you can be passionate about.

Deal with an important activity or task. There is an activity you like that is simple and not important and also there is a job you like that will give a long-term impact on your life and career. Select the former as it will be a much better use of one's time and of Flow.

Ensure it to be challenging but not so difficult. If a task is very simple, you will be able to go through and finish it with little thought or effort. An exercise should be challenging enough to require someone's full concentration. However, if it is very difficult, one will find it hard to lose oneself in it, as one will spend most of their concentration time just trying to find out how to do it either that, or one will end-up in discouragement. It may need some trial and error method to figure out tasks of the appropriate level of hardness.

Look for an ultimate and quiet time. This involves two stages assembled into a single stage. One would first want to find a quiet time which will enable one to be able to focus on an activity. I usually choose the morning hours before I get caught in the daily hustle of that may culminate to a busy but boring one. Early morning when one just got up from bed could be the right time, or it could be early at school or job side when most students or workers have not yet reached or are still one the coffee desk and trying to put things in place. Otherwise, one could try to do so during the hour of lunch or lunch break when students are usually out of the

classroom or office desk. The evening time may be ideal for some persons or other lucky ones might do their exercise at any time of the day when one can find a quiet area in which to work. No matter the time one chooses, it should really be a convenient energy time for them. It will not be a good time to go for flow after lunch when one gets tired. Settle for a time when one has much more energy and the ability to concentrate.

Put away anything that could be a distraction. Besides looking for a quiet place and time to carry out an activity, one would be at ease putting away all other sorts of distractions. This involves turning off pieces of music that could distract excerpt the one that helps one to stay focused, putting off of telephones, electronic mails and other notifications like Twitter and Google, and any other thing that might come up or bring up some noise to interfere with one's thoughts. It will also be of help to clear off my desk even when that implies piling miscellaneous paper copies within a file folder and to consider sorting through them later on. However, there is nothing on my desk these days but I have not been working like this even though on a serious note, a cleared desk helps.

One would have to master being focused on such a task for a long time possible even though there is need for regular practice. Begin on the activity or task one has selected and stay focused on it for a possible length of time. We may have some difficulties at first if we have the habit of regularly swapping from one task or activity to another but we must continue in our attempts to re-focalize on the task we are carrying out, each time we move away from it. This will make us better in it. If your focus can be fixed on the task without distractions having very well selected the task or activity that one loves, considered important and challenging, one should then be immersed in the flow phenomenon.

It is an exciting experience to lose oneself in the flow phenomenon, to me; it is a moment of enjoyment to experience flow. It is a great event to have the ability to plunge oneself into an activity considered worthwhile; to take giant strides towards realising an important activity or project and to do something someone considers passionate to them. Make up time to evaluate how one feels surely understanding the fact that it will be difficult to evaluate it when in Flow already.

Practice makes perfect. Practising is a necessary action to take. Make some practice at each stage or step, after finding a convenient and quiet time, for one to putting away distractions, to selecting the appropriate activity, especially maintaining someone's focus on an activity for a long time possible. Nevertheless, whenever one fails, use it as a lesson to learn. When success comes, also draw out some lessons from it; what was done correctly in the process? The more practice exercises, the better you get towards the flow.

Accept the wins. In addition to the happiness gotten from reaching the flow, one will also be fulfilled with everything concerning work. Important activities will be carried out. More activities will more often be accomplished instead of beginning and frequently being interrupted. All of these are highly rewarding and satisfying. Make out time to appreciate such and also to keep on practicing it daily. "To keep on being concentrated for a considerable length of time is important in order to handle difficult achievement." Russell (2016)

**Time factor: How long does it take to get into Flow?**

Much time is required to attain the flow state. Many people approximate about 30 minutes time to keep focusing on an activity or task to reach full time flow, though this varies based on the individual, the time of the day and the activity to be carried out. The most important of things to note is that it takes some time to reach flow.

Being in a state of flow is also known informally as finding oneself in the zone. This is what positive psychology. it is the state of mental operation wherein an individual carrying out an activity is plunged fully in a sensation of calorized focus, total involvement and enjoyment in the process of the activity.

**How to get into Flow state when writing**

Seven easy means of getting into a flow state while writing

All distractions should be eliminated because they are always found in the surrounding environment.

Many individuals love listening to music while writing or working. They could listen to instrumental music on an earphone.

Stay hydrated.

Supplement your nutrition.

Don't depend on stimulants.

Take a nature break.

Fit in exercise.

In life flow is the art of letting things happen. Flow deals with the state of mental operation where someone in an exercise is totally plunged in a sensation of energised focus total implication and victory in the process of the exercise. Flow however, can also be applicable to one's entire lifespan and while that occurs, one must have traced their path.

**The Cognitive Constructivism Theory**

Constructivism views that learners make sense from what they do on their own. It propounds that learners come to knowledge by recognizing the meaning of what is

found in the environment. The event in the environment could be anything: a test, a chemical reaction, a sentence read from a novel, the instructions of a teacher in the classroom. This constructivism places emphasis on the value of the active implication of the students in the building up of knowledge for themselves. Learners are instructed to make use of the previous knowledge and notions to help them in the acquisition of new information. Other terms used to depict constructivists view of learning includes; generative learning, situated learning and authentic instruction. The roots of constructivism can be traced from John Dewey (Learning by Doing) and the progressive educators to Jean Piaget and Vygotsky, then to Jerome Bruner and delivery Learning.

During test structure analysis, the test is supposed to consider diverse interests of the learners that would make them exploit their various talents and the environment in motivating them to work by themselves and to see results for themselves by personal or self-evaluation.

**Constructivism: Piaget (1896-1980)**

Here, Piaget as a cognitive cognitivist focused on development wherein the accommodation mechanisms, equilibrium and assimilation remain keys to the process of development which is built on adaptation. It is via these mechanisms that people build up novel knowledge from the experiences they have got. While students learn or assimilate novel knowledge, it implies that they have inculcated experiences that are new within the old scheme or framework. This usually happens with the existence of harmony within that which has been learned, amidst the existent knowledge and disharmony is when there is lack of understanding. While this occurs then comes cognitive advancement. Accommodation comes as a mechanism to reframe someone's mental perception of experiences of the outside environment to accommodate novel experiences.

Pedagogic mechanisms that propagate active learning are what correlate with Constructivism. Real learning to Piaget is for the learners to bring forth solutions individually. His centralized location with regards to learning is discovery learning wherein the student interiorises from hands-on activities than from inactive or passive participations. The worthiness of first-hand and direct information is very important for the teachers to be abreast with. In order to demystify learning, the environment for knowledge acquisition has to be organised in such a manner that will give room for students to evolve in skills and concepts on their own and at their speed following each one's individual aptitude. Engaging in active learning process as such will take us to profound learning framework.

The opportunities for knowledge acquisition that will enhance the development of cognitive abilities focalise on studying instead of the final outcome via the encouragement of the students to pose and respond to questions, explore and manipulate the outside surrounding. Promote student/student teaching questions to peer and the teacher. Pedagogues therefore are demanded to have concern in the mediation of mental processes giving that knowledge acquisition happens within a wider procedure of development which considers a continuum of progressive intellectual interaction and cumulative organization of that which has been acquired.

If learners become familiar with all the aspects of test development; instructions, coherence in form, structure of test and clear test purpose, they will easily assimilate, accommodate and equilibrate whatever is going on during lessons and take active part and personal initiatives permitting them to guide their own activities in learning thereby developing creativity.

**Social Constructivist Theory- Lev Vygotsky,** (1896-1934)

Athough Vygotsky is given the credit to be the 'father' of social-constructivist theory; his finding was centered on the idea of Jean Piaget on the offspring of man as an active knowledge acquirer. Piaget (1959) mentioned by Verenikina (2010), which is usually referred to as the 'roots of constructivism' Jones & Araje (2002). Whereas Piaget focalised on steps of development of the child and personal building up of knowledge referred to as 'individual constructive theory' Vygotsky found out the greater socio-cultural context.

Vygotsky and his collaborators in Russia during the 1920s and 1930s Steiner & Mahn (1996) systematically applied Socio-cultural processes to acquiring knowledge and development. Vygotsky brought up a multi-layered, rich theory from where he examined a series of subjects which included the psychology of art, thought and language and learning and development including focusing on the education of learners having special needs; Steiner & Mahn (1996).

**Social-Constructivist Theory; the Vygotsky approach**

Vygotsky mentioned that the "human mind is developped through an individual's manipulation process with the environment and is characteristic of the relationship that exists among subject and the object" Verenikina, (2010). He discovers a representative role in humans' understanding of the environment and of themselves. These roles are characterised as 'tools' Turuk (2008).

Moreover, Vygotsky stood for the fact that humans beings do not directly act on the physical environment without the use of instruments as intermediary. These instruments could be some art ware which could be like a symbol or signs made by humans based on particular historical and cultural conditions bearing with them the specificities of the said culture Turuk (2008). Thus, he debates that mental

procedures would be comprehended only when one understands the instruments and signs that mediate them Verenikina (2010).

Following Vygotsky, a kid is dependent completely on other persons within the early stages as the socio-cultural milieu keeps on to present to the kid a variety of activities and necessities plunging the kid in his/her environment. Turuk (2008). Particularly, these persons could be the child's parents who instruct him/her on what to undertake, how to go about it as well as what to avoid, thus, initiating the kid's activity. Parents are considered as those who represent the culture and the conduct via which the culture passes into the children, make them actualised instructions primordially through language. Vygotsky went ahead to explain that infants personalise these social and cultural inheritances by the acquisition of knowledge via interactions and contacts with persons as the first step we refer to as inter-psychological plane and later assimilated and internalised of this knowledge culminating one's private importance to it also we refer to as intra-psychological plane Turuk, (2008). This movement from communal to private property is not seen simply as a copy; rather it is considered as a transformation of that which had been studied via interaction to become private issues. Vygotsky in addition, holds that in educational institutions also learners do not copy the instructor's' abilities but they transform that which the instructors present to them within the processes of appropriation Turuk (2008).

The social development theory of Vygotsky is the output of the Russian psychologist called Lev Vygotsky (1896-1934). His finding was greatly not known to the Western world till it was made public in the year 1962.

This theory of Vygotsky serves as one among others of the foundations of the constructivism theory. It states three main topics with regards to social interaction,

the more knowledgeable other (MKO) and the zone of proximal development (ZPD).

**Social Interactions**

Social interactions have some primordial roles in the cognitive development process. Contrary to the understanding of the development of the child by Jean Piaget where development is said to come before knowledge acquisition, Vygotsky instead thought that social knowledge acquisition is encountered before development. He stands that: "Every function in the cultural development of the child occurs two times: 1st on the social level and afterwards on the individual level; again 1st between persons; inter-psychological and later within the infant-intra-psychological".

- The more knowledgeable other (MKO)

The more knowledgeable other depicts whoever that has a more reasonable understanding or a greater ability level than the student in relations to a specifically given activity, procedure or concept. The MKO is thought of normally, to be a coach, teacher or elderly adult but the MKO could still be co-learner, a youthful individual or computers too.

**The ZPD- Zone of Proximal Development**

The ZPD is the stretch between a learner's ability to carry out an activity under the supervision of an adult and with other learners' cooperation and the learner's ability to independently solve the problem. Learning takes place within this zone as put forward by Vygotsky.

Vygotsky based his work on the relationships between persons and the socio-cultural environment which they manipulate and interchange in experiences they explain to each other. In accordance with Vygotsky, human beings utilise

instruments that emanate from a culture such as the one spoken and written to interact with their social milieu. Initially, learners create the gadgets to mainly be used as social roles, means to pass on message about their needs. Vygotsky held in mind that the internal appropriation of these tools gave room to more elevated thinking abilities.

Vygotsky went on to say that, 'good learning' takes place in the Zone of Proximal Development wherein he distinguished between the learners' potential and actual stages of development.

Vygotsky defines the Zone of Proximal Development (ZPD) as "the distance between the present level of development as has been decided by solving problem independently and the stage of potential development as has been programmed via the solving of problem based on the guidance of the elder or in accordance with peers who are more capable" Schunk (2012).

Puntambekar & Hübscher (2005) as cited in Schunk (2012) on their part said that the ZPD indicates the quantity of possible knowledge acquired by a learner in possession of the right teaching conditions. The principal explanation to the exposition of the concept of the ZPD by Vygotsky was as a result of his disagreement with two practical aspects in psychology of education Turuk (2008). Aspect number 1 is learner assessment on abilities of the intellect and the number 2 is the evaluation of the practices of instruction. He had in mind that the issuing of test should not depend on the actual level of the learner's acquisitions but it is of more vital importance to also consider the potential development of the learner, Verenikina (2010).

He claims that the present developmental stage that is given consideration as the stage of independent performance, does not appropriately give a description of development. It instead presents "what is developed already or acquired, it is a

‘yesterday of development’. The stage of aided performance gives an indication that what someone can acquire in the closest future, what is evolving like potential stage, ‘tomorrow of development’, what someone ‘can be’ Verenikina, (2010). Thus, ZPD can be given a description as the gap between what someone can do with help and without help.

Cole & Cole (2001) in their work cited in Verenikina, (2010) say that the word ‘proximal’ is synonymous to ‘nearby’, meaning that the provision of assistance goes a little bit above students’ actual competence complementing and improving on their preexisting abilities. This is grossly a test of a learner’s readiness in the light of development or intellectual stage in a specified area of study and it indicates the relationship between learning and development Schunk, (2012).

In the ZPD, a leader and a student put their efforts in cooperation on an activity which the student found not easy to carry out unassisted due to the level of difficulty. It also mirrors the notion of collective action where in referring to Bruner (1984), cited in Schunk (2012), those who are more knowledgeable or are more able to do something, share the said knowledge and ability in order to complete an activity with others who are less knowledgeable. Rogoff, (1986) cited in Shunk, (2012) stipulates that a reasonable lot of guided participation is needed while working in the ZPD and students put in private personal understandings to social cooperation and build up meaningful ideas by introducing these understandings including the experiences in the given context.

“As an example, consider that an instructor like ‘Munja’ and a kid like ‘Wase’ would work on an activity like drawing an image of mummy, daddy, and Wase working something out in cooperation in the house. Wase put into the activity her understandings of what the persons and the house seem to be and of the

kinds of things they might engage on, in addition to knowledge of how to go about drawing and making picture images. Munja puts in similar understandings and knowledge of the necessary conditions to work on variety of activities. Let's consider that they chose to make a picture image of three of them working in the compound. Munja could draw up a picture of father clearing grass, mother trying to trim the hedges, and Wase raking the cut grass. If Wase was to draw her image in front of the father, Munja might explain that Wase must be behind the father to rake up the grass left behind by his clearing. In the course of the interaction, Wase modifies her convictions about working in the compound based on her present understanding and on the new knowledge she constructs" Schunk (2012).

**Educational Applications of Social-Constructivists Theory**

Moll, (2001), Karpov & Haywood, (1998) believe that the ideas of Vygotsky sold themselves to many applications in education marking the domain of self-regulation as the greatly influenced Schunk (2012). One of the applications that reflect the theory of Vygotsky is 'reciprocal teaching'. It consists of interactive discussions between the teacher and a little group of Learners. First, the teacher simulates the actions. Later the teacher and learners take turns as if they were the teacher. In the course of the presentation of practical work results, if the learners practice asking questions, then to understand their understanding level, the instructor can put in place a question-asking strategy in the sequence of teaching. As students develop skills gradually, 'reciprocal teaching' involves the principle of social interaction and ZPD of the perspective of Vygotsky. Schunk, (2012).

Another domain in which the ideas of Vygotsky fit is the peer collaboration. Ratner et al (2002) and Bruner (1984), state that it resembles the ideals of collective activity Schunk (2012). The social interactions shared while learners carry out a task collectively serve as a function of instruction. This strategy is basically utilised

in the studying of Biological science, Mathematics, other sciences and language arts which proves to the realised effect of the social world at the time of learning.

Someother field of application of Vygotsky's theory is apprenticeships as they pop up in cultural institutions such as agencies and schools that help in the transformation students' cognitive development. Apprentices do their operations inside a ZPD on the task as basically their activities depend on tasks above their capabilities. These apprentices enhance an equally distributed understanding of valuable procedures by collaborating with people of great expertise and integrating this with their present understandings Schunk (2012).

**Vygotsky's Social Development Theory Applications in education**

Transmissionist or instructionist model have traditionally been adopted by many schools in which a teacher or instructor 'transmits' information to learners. Contrarily, Vygotsky's theory propagates the knowledge acquisition contexts in which learners take a primordial position during learning. The roles of the pedagogue and learner are thereof moved, as a pedagogue would cooperate with their learners in a bid to assist in the facilitation of meaningful construction in learners. Studying thereof comes to be an experience reciprocity for the learners and instructors. Sciences, Second Edition, (2015)

The Socio-constructivist theories focalise on the making of reality and the way individual persons perceive the environment. These theories are in relation with the humanistic theories in that they were influenced by philosophy including phenomenological notions of that which is reality and that which is constructed socially. The foundation of the present day social constructivist theories emanates from Berger & Luckman (1966). They explored the creation of reality and the influence of personal meaning based on the experiences of life, cultural and societal expectations, norms, and rules that they named 'social constructionism.' From then,

many more theorists have modified social constructionism into three other particular theories, which lay emphasis on either collective or personal strengths in the construction of reality Teater, (2014). Socio- constructionism states that what is real is composed through the utilisation of speech during interactions with others and is primordially impacted by past events, social domains and cultural factors. Constructivism stands that reality is manufactured more through someone's biological strains through the processes of development, cognitive build-ups, and the human mind. Social constructivism puts together both socio- constructionism and constructivism by putting the emphasis of reality construction on both societal and biological factors. There exists a combined emphasis on the fact that the construction of reality is being impacted by both nurture and nature.

Social constructivist theories have been the backbone of many other theories used usually inside the practice of social work. ‘Symbolic interactionism’ focused on three basic principles of language, thought, and meaning. The theory propounded that humans make use of symbols like roles, rules and words to provide meaning and to make sense of the world around. This meaning is transferred to other persons via language. Human beings are thought to interact with one another and the society and they allocate meaning and symbols to these kinds of relationships and interactions DeLamater & Myers (2011). Further inter-actions are thus, based on the kinds of meanings and symbols each one has tagged on that situation or relationship therefore, people take action on what they hold at heart against what is objectively true. The theory of communication stands that “persons cannot be deprived from communicating” and that “every behaviour is communication and every communication impacts behaviour” Watzlawick et al (1967). The Role theory looks into how individuals act out socially defined roles like father, wife, sister, teacher, manager, and mother; their competence to hold on

to the expectations of the society of acceptable and unacceptable kinds of behaviour for the specific role, DeLamater and Myers (2011).

Social constructivist theories are of great use to the practice of social work because they make provision of a theoretical base for the understanding of how realities and perceptions of the environment are person-specific and created from a combined work of interactions with the society, the historic structures of the society, rules, norms, and culture and the meanings that people give to that type of interactions. Those theories are primordial in use in the assessment and intervention levels of practice in social work. Applying the socio-constructivist theories in assessment considers a social worker assuming a curiosity position with the customer utilising language to try to understand the customer's reality and perception of the environment and to acknowledge that two customers do not have similar reality of perception of the milieu despite living following the same experiences.

The stage of intervention includes a social worker putting in interventions which might explore the experiences of the customer and meanings with attempt to rebuild thinking or viewpoints that are culminating to the presentation of a problem into someone that are more accommodative. Interventions also consider the social worker making a challenge to assumptions of the society or social constructs which are blocking the customer from growing up and developing. Those types of interventions usually applied in social work practice are narrative therapies; solution-focused brief therapy, social constructivist approach, anti-oppressive, cultural competency, advocacy, anti-racist and anti-discriminatory practices.

Research on "Human Learning" is unstoppable that keeps on to develop and to expand. Individuals have reached the agreement to the fact that learning is

important but the viewpoints on the processes, causes, and the consequences of learning differ from one person to the other, Schunk (2012).

As research works from a variety of theoretical traditions put forward their ideas and hypotheses in fundamental and applied situations, the results of their research provide some improvements in learning and teaching but to relay these research data, it is expected to have a theory comprising of a set of principles which are acceptable scientifically and provided to give explanation of a phenomenon. The theories put in place a structure for the interpretation of observations in the environment and operate as linkages between education and research, Suppes (1974, mentioned in Schunk (2012). In addition, research data is regarded as simply a disorganised collection of information if it is not founded on any theory. Thus, for the researchers on "Human Learning" to depend on it, there are varieties of "Learning theories".

"Learning theories give accounts based on empirical data of the variables that impact on the process of learning and put forward explanations of the manners in which that impact happens" Alzaghoul (2012). It aids one to also understand the particularly complicated process of learning as it gives a description on the way persons and animals get knowledge. There exist three major categories or philosophical structures below which learning theories belong such as cognitivism, constructivism and behaviourism Alzaghoul, (2012).

The theories on behaviourism basically focalise on the objectively perceivable elements of learning but cognitive theories search above behaviour and explain learning based on the brain. In other words, constructivism sees learning as a process whereby the student in participation, build up or construct novel notions or concepts, Alzaghoul (2012). There are a series of theorists included in all large categories explaining various viewpoints on the main concept.

Vygotsky placed emphasis on the fact that adults and children together act as active agents in the process of development of the child. Cole & Cole (2001) quoted in Verenikina (2010), write that the development in this example is constructed together. Both the instructor and the learner are considered as active members in the knowledge acquisition of children when referring to teaching. It is very necessary for the teacher to intervene in the learning of the children but it is the quality of the interaction between the teacher and learner that is viewed as crucial in that process of learning; Tharp & Gallimore (1998), cited in Verenikina (2010).

The theory lays emphasis on the importance of that which the student comes with in whatever situation of learning as an active maker of meaning and solver of a problem Turuk (2008). It accepts the nature of dynamism of the interaction between learners, teachers and duties and makes provision of a stance of learning as taken from interactions with others. Ellis (2000) mentioned in Maturuk (2008) stipulates that the theory of Vygotsky makes some assumptions that learning emanates not 'through interaction' but 'in interaction'. Students start first in succeeding to perform a new duty aided by any other individual and then interiorize the task such that they could carry it out by themselves. As such, social interaction is considered to mediate in knowledge acquisition.

The theory moreover extrapolates to add that interactions that mediate learning successfully are such in which the students scaffold the new duties Turuk (2008). The following list identifies the key points of the theory.

- Social interactions are valuable; knowledge is cooperatively constructed among two or more persons.

- Developing an interior representation termed internalization of actions and mental operations that occur in social interactions is the manner through which Self-regulation is developed.

- The cultural transmission of tools; that is language and symbols leads to human development.

- The most critical tool is language which emanates from social speech, to personal speech and to inner (covert) speech.

- The difference between that which infants can do by themselves, and that which they can do with the help from peers or adults defines the zone of proximal development (ZPD). The interactions of the peers and adults with the learner in the ZPD encourage the development of cognition. Meece (2002) quoted in Schunk, (2012).

The work of Vygotsky has formed the social constructivism foundation in educational settings Jones & Araje (2002). Particularly, Vygotsky emphasised on the role the social context or of others in learning which has forced educational experts into re-examining the level to which learning is a personal process. Moreover, they have laid emphasis on the role the bigger community and the role of significant others in the learning process. Then, the Social constructivism has got the highest effect on teaching and curriculum design due to the fact that they seem to be the most all-round-fit for integration within the current educational approaches, Jones & Araje (2002).

Vygotsky's fervent view shows that students build up meaning from reality and actively acquire what are being taught in their learning environment. Constructivism therefore means that learning includes creating, constructing, developing and inventing someone's personal knowledge and meaning. The teacher

is a facilitator in his role as one who gives information and ensures the organization of activities of pedagogy for students to uncover their own learning Liu & Chen (2010). Furthermore, Vygotsky's ZPD serves as foundation of many more concepts and approaches.

**Gestalt (or Insight Learning) Theory- Perception Patterns and Relationship**

The learning theory called Gestalt has its origin from Germany. Three theorists of German origin propounded the theory;Wolfgang Kohler, Max Wertheimer and Kurt Koffka who had their inspiration from the publishing and ideologies of the one who gave the name to the learning theory. A theorist in learning; Graf Christian von Ehrenfels who had taken the holistic approach to Knowledge acquisition by propagating the ideology that learning happens while learners have the ability to understand in its entirety, a conception, instead of analyzing down into separate parts.

"Gestalt", as a term originates from a word in German that approximately has as meaning form or pattern. Basically the theory of Gestalt stands that the 'whole' is greater than the 'sum of its parts'; meaning that learning entails much than just commanding voluntary feedbacks from students.

Just like it is said in different theories of learning, the Gestalt theory owns legislations of organization via which it should operate. Those laws or legislations organisation exist already in the human structure of the mind and the way they structured perceptions. The Gestalt theorists propound that the experiences and perceptions of students poses a significant effect on the manner in which they go about learning.

Phenomenology is a peculiarity of Gestalt, which is the study of the way individuals display learning through introspection of the experiences they have

lived and consciousness. When the teaching exercise is in relation with the reality of experiences of life, learning occurs in its best. The brain of human beings poses the capacity to build a scheme of the stimuli made by those experiences of life. The procedure of making schemes or maps is known as "isomorphism." Isomorphism can be practiced in test development in that as the teacher regularly sets tests following the right practices and adapted to the level and environment of the learners, they develop familiar experiences and will organize their learning activities consciously with fixed objectives to attain.

In an automatic manner, the brain tries to fabricate a whole image every time it views just an area of an image. This presents the law number one of organization termed the "factor of closure" which is not just applicable to pictures but also applicable to sounds, feelings and thoughts.

Gestalt theory holds that the brain of a human draws up aspects of learning which are forwarded near to one another as an entity or whole in the place of separate parts. This law of organization is termed the "factor of proximity," which is most often found in learning domains like music and reading wherein letters and words or notes of music do not give sense when standing alone but turn out to a whole story or song when mapped out together by the brain of humans.

The following law of organization of the theory of Gestalt is the "factor of similarity," that holds that learning is facilitated if similar groups are connected with one another and compared with other groups which put forward ideas that differ. This kind of Gestalt learning causes students to develop and improve critical thinking skills.

While making observation of items found beside us, it is not strange for the eye to wave away emptiness or space and holes only to find whole objects instead. That

type of organizational law is termed the "figure-ground effect". In the process where new ideas and thoughts are interiorized, the brain appears to create links, or "traces," which form representations of the connections which take place between conceptions and ideas and also pictures. This organizational law is called the "trace theory."

The theory of Gestalt puts its fundamental stress on the processes of cognition of a greater order making the student to make use of higher skills of problem solving. They are obliged to consider the concepts put forward to them and do research for the fundamental similarities which liaise them to one another into a binding whole. Following such a manner, students have the ability for the determination of particular relationships between the perceptions and ideas put forward.

The theory of learning by Gestalt puts forward the importance of the presentation of knowledge or pictures which present loopholes and factors that do not fit squarely within the whole image. This kind of knowledge acquisition demands the student to utilize skills of critical thinking and skills of problem solving instead of placing off responses using the remote memory, the student should make an examination and deliberation so as to find out the responses they are looking for. This will promote sustainable learning which will culminate with eminent competency development.

During the presentation of information from the teachers to the learners with the use of the Gestalt learning theory, say during test development, the teachers should make sure that their teaching strategies consider the use of the organizational laws such that each aspect be seen clearly and its role obvious in the whole test.

The Gestalt theory propounds that a person is considered a whole being and the teaching strategies put in place for teaching and testing the individuals assist to the

discovery of whether there is a thing which is stopping mentally their reasoning from studying some other new information. Testing procedures serve for the presentation of difficulties in general and to try the removal of whatever mental blockage from the student so as to allow new information to be preserved.

The design of test development procedures that consider the student's history and present perceptions and experiences constitute the way to teach new information. Within the Gestalt theory of learning, if the students encounter testing information or notions which are unorganized, the cognition puts it in order in a bid to ensure that the student recognizes and applies the notions being tested.

**Discovery Learning Theory: Jerome Bruner**

"The first work of Jerome Bruner in the domain of cognitive psychology laid focus on the series of decisions taken by the concerned as an entity of the strategies in solving problems in situations of experiment. The work of Bruner in cognitive psychology guided us to the zeal in the development of cognition of infants and similar aspects of instruction and in the 1960s he brought forward a theory of cognitive growth. The theories of Bruner lay emphasis on the experiential and environmental conditions that influence every member's particular pattern of development. He argues that the intellectual ability of human being evolves in levels from childhood or infancy to adulthood by a stage-by-stage progression in the way the mind is utilized and how it has impacted experimental psychologists and educators all over the world. In particular, Bruner has interest in language and other ways that the human thought is represented. Bruner made a definition on three of the modes of representation or "symbolizing," human thought in one of his best-known research articles. The enactive mode comprises human motor abilities and involves activities like the use of tools. The iconic mode refers to sensory competences. To end with, the symbolic mode considers reasoning and is specified

in the use of language which carries a pivot role in Bruner's cognitive theories and development. He gave it the term "a medium not just for the representation of experience but for transforming it also."

This implies that in test structure analysis, the aspects that influence the motor activities, sensory capacities and reasoning should be considered and the language, to the reach of the learners; such that no dictionary or interpreter may be needed to enable the students answer the questions without bias.

The view that the learner becomes an active participant in the process of education by Bruner has got a wide acceptance.

Ormrod, (1995, p. 442) gives a description of discovery learning to be "an approach to teaching via which learners interact with their surrounding by manipulating and exploring objects fighting with interrogations and controversies, or performing experiments."

> “Discovery learning is based on inquiry, it is constructivist learning theory which occurs in problem-solving situations wherein the student draws from historic experiences and existing knowledge to fish out information and relationships and novel facts to be studied. It is foreseen through this process that learners could more probably remember notions and knowledge uncovered by themselves. Models which are founded on discovery learning model comprise: case-based learning, guided discovery, problem-based learning, simulation-based learning and incidental learning. Those who put forward this theory have in mind that discovery learning has many strong points: promoting motivation, encouraging active engagement, independence, responsibility, autonomy and creativity development and problem solving skills” Learning theories.com.

Bruner's theory depicts the framework of knowledge. He explained it by stipulating that the specialist in curriculum and the educator "must precise the means through which a collection of instruction must be organised such that it could be most timely understood by the student." This became the idea considered among the most valuable concepts credited to Bruner. The explanation was given as: "Any problem, idea or collection of knowledge can be provided in a simple way enough such that any peculiar student can get abreast with it in a form which is recognizable."

According to Bruner, teaching should direct the student over the content so that there can be improvement in the ability of the student to "transform, grasp and transfer" that which is understood. Generally making sequence must flow from concrete, hands-on or enactive to visual or iconic and to symbolic; mathematical symbols or descriptions in words. This sequence will however be of dependence on the learner's symbolic system and style of learning.

Bruner proposes that to move from external gifts (extrinsic rewards) like praise from a teacher, to internal zeal (intrinsic rewards) peculiar to the solving of problems or to understand concepts is highly desired. To him, learning is dependent on knowledge of the findings if it could be of use for correction. The learner seriously needs feedback for the development of knowledge. The instructor can produce a required connection to the student in the provision of feedback at the start and then provide help to the student for the development of strategies for the obtaining of feedback on their own.

In contribution to the above layout, Patsula, (1999) proposes that, "With reference to Bruner, the teacher should attempt to encourage learners put forward hypotheses, provide decisions and uncover principles by themselves Kearsley (1994b). The task of the instructor is to "translate the substance to be studied into

an appropriate format to the student's present state of understanding" and the organization of it in a spiral form in order that the learner continually constructs on that which they have already studied."

Bruner (1966) as quoted in Kearsley (1994b) says that "an instructional theory should look into the issues that follow: The most effective sequencing wherein one presents the material, the manner in which a collection of knowledge could be structured in order to let it more readily assimilated by the learner."

Kearsley (1994) summarizes by employing the principles that follow:

"-Instruction should comprise the experiences and contexts which make the learners anxious and with the ability to study considered as readiness.

-Instruction should be structured such that it could easily be assimilated by the learners; termed spiral organization.

-Instruction must be constructed to ease the extension and or close up the gaps; like going beyond the information given."

This work of Bruner is supported by the experiential learning theory which postulates that in the absence of reflection, we would just keep on repeating our errors; MacGill & Beaty,(1995).

In his research Kolb, cited in Molindo, (2008); course document found out that "people study following four means with the probability of one mode of learning being developed more than the other. The experiential learning cycle presents that learning is;

- done via concrete experience

- acquired by reflection and observation

- obtained via abstract conceptualization

- and via active experimentation".

Following the variety or difference in the styles of learning; Honey & Mumford (1986) quoted in MacGill & Beaty (1995, p.177), founded on Kolb's research, outlined four styles of learning:

- "Those who enjoy the experience itself; activists,

- Those that spend a great deal of time and effort reflecting: reflector

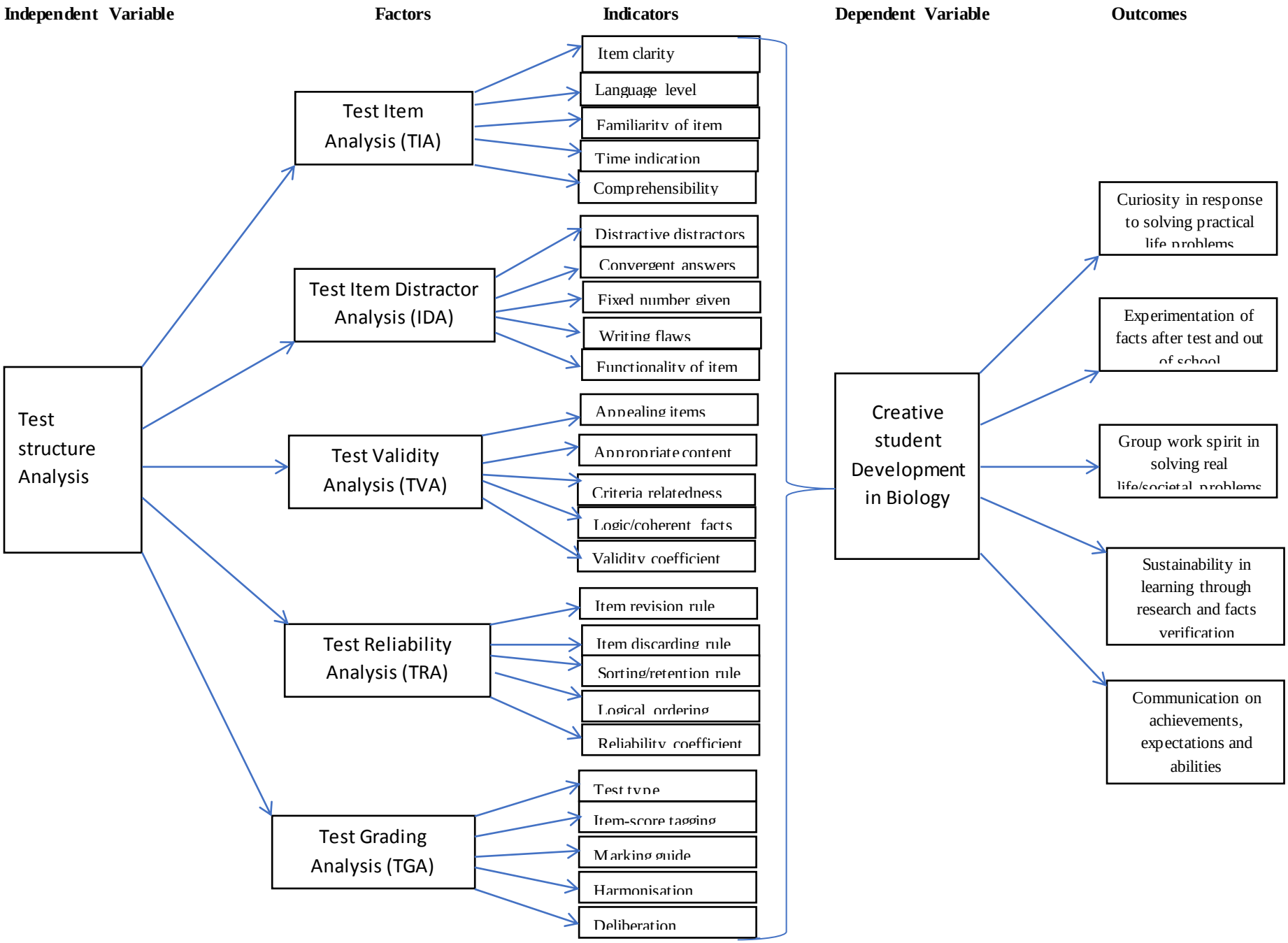
- Theorist (good at making connections and abstracting ideas from experience),

- Those that enjoy the planning stage: pragmatist"

Honey & Mumford debate that "learning is ensured when thinking about the style we use in learning such that we can construct on strongholds and engage towards minimizing weaknesses to improve-on the quality of learning."

With this variety of styles of learning, it is obvious that one manner of developing a test cannot meet all the needs of the learners since each learner may learn in their own way. According to the Cognitive Evaluation Theory (or Theory of Self-Perception), when talking about task, we do the evaluation of it in the light of how good it would meet our requirements to have the feeling of competence and in control. When we have the impression that we will have the ability to get through the task, we will be motivated intrinsically to finish it, with no requirement of further external motivation.

Figure 2: *Conceptual Diagram of Variables, Factors, indicators and outcomes*



**Review Related Literature based on Empirical Studies**

Lord, (1952), states that using the item difficulties and inter-item correlations as its tools, the test scores theory yielded the conclusions as follows: (1) the regression of test score on ability is evidently curvilinear; (2) the standard error of measurement is normally the smallest for those test takers for whom a test is least discriminating; (3) errors of measurement have a binomial instead of normal distribution; (4) the highest discrimination at a said level of ability happens when everyone of the items are of 50% difficulty; (5) a free response test with no extraordinarily high item inter-correlations and with everyone of the items of 50% difficulty happens to be more discriminating for practically every one of the test takers than a test characterized by a diversity of item difficulties; (6) "the shape of the frequency distribution of test scores or of true scores unnecessarily represent the shape of the frequency distribution of ability." Some empirical studies considering seven short tests revealed very good agreement with the theory.

Atumasa, (2018) "Test development practices for sustainable learning competences in biology in secondary schools in Meme division." Was the title of the master's dissertation done in the University of Bamenda in the year 2018? The objectives were as follows: The purpose of this study was to find out the test development practices that would enhance sustainable learning competencies in learners. Below are the specific objectives that guide study:

1- To find out any relationship between the characteristics of test audience during test development and sustainable learning competencies in learners.

2- To find out how the process of developing of a test plan relate to sustainable learning competencies in learners. The researcher administered the

questionnaires to 150 students and 50 teachers. The simple random sampling technique has been used in choosing the sample. The data collected was analyzed using the statistical package for social sciences (SPSS) version 20 and the bivariate correlation procedure or The Pearson Product-Moment Correlation Coefficient was used to arrive at correlation coefficients. It was found that of the four factors of the independent variables correlated with the dependent variable, all had a significant correlation. This indicated that there is a significant relationship between test development practices and the development of sustainable learning competencies in learners. Therefore the characteristics of test audience, the development of test plan, the process of item development and the writing of test administration instructions all proved significant either at the 0.05 (2-tailed) level or the 0.0 1(2-tailed) level of significance. The researcher recommended that teachers should scrupulously consider the specificities of the test audience in developing the test, develop a test specification table to serve as base for the composition of test items and ensure the writing of test administration instructions before administering any test. With the strict follow-up of these practices, students will be able to have an insight of what a test may be and study in line with the objectives of the testing process. Coherence and consistency in the test development and testing processes will give a sense of critical and reflective thinking to learners and they may develop sound judgments about their works and efforts they put in during studies.

3- To determine how the process of item development relates with sustainable learning competencies in learners.

4- To find out how the process of writing the test administration instructions during test development correlates with sustainable learning competencies in learners.

## Review of Related Literature Based on the Variables of the Study

The review of literature at this stage is based on the variables of the study. There are two main variables; the independent variable (Test structure analysis) and the dependent variable (Creative student development in Biology). Since the main independent variable has been operationalized, it comprised five research variables; test item analysis, test item distractors analysis, test validity analysis, test reliability analysis and test grading analysis), each linked to the dependent variable to form a hypothesis. It is the linked variables which were used to review literature at this stage. The reviewed parts were as follows:

## Test Item Analysis and Creative Student Development in Biology

Item analysis is a methodology of relooking into items on a test, both qualitatively and statistically to make sure that they all meet minimum quality-control criteria. The difference between qualitative review and statistical analysis is that the qualitative review uses the expertise of content experts and test review boards to identify items that appear not to meet minimum quality-control criteria. Such qualitative review is necessary when doing item development when data is not available for quantitative analysis.

A statistical analysis like the item analysis is carried out after items have been administered and real world data are available for analysis. Statistical analysis also assists in the identification of items that may have slipped through item review boards; item defects can be hard to identify. The objective of qualitative and statistical review is the same; to identify problematic items on the test. Problematic items are also called “bad” or “misfitting” items. Items may be problematic as a result of one or more of the reasons that follow:

- Badly conceived items can cause learners to get confused while giving responses to them.

- Graphics; diagrams, pictures or other information developed in the items may not be clearly representative or may actually be misleading.

- Response to items may not be clear or correct and a distractor could potentially qualify to be the right answer.

- Items may have distractors that many candidates can obviously consider as wrong, increasing the ambiguity of the candidates to guess the answer that is correct.

- Test items may be a representation of a different content area other than that measured by the rest of the test commonly known as multidimensionality.

- The presence of bias factor for or against a gender, ethnic or particular sub-group may be seen in the item or the distractors.

- Bias for or against the context or environment in which the test takers are found.

Item analysis comprises of a procedure that permits one to find out the responses to the questions that follow:

a- Do the test items function as expected?

b- Do the test items contain the appropriate difficulty level?

c- Are the test item void of irrelevant indicators and other clues or defects?

d- Are the distractors in multiple choice items type effective?

**The use of item analysis data:**

a- The provision of a base for effective discussion of the test result class

b- The provision of a base for remediation in class works

c- The increase in test construction skill

d- The improvement of classroom discussion.

**The Procedure on Item Analysis:**

The procedure for item analysis provides a particular emphasis on the level of item difficulty and the power of item discrimination.

**The procedural steps of item analysis are as follows:**

1. The ranking of the test papers from the highest to lowest score.

2. Take out 27% highest test score papers and 27% lowest test score papers.

As an example; if a test is administered on 60 students, 27% will be 16 so we select 16 test papers from highest scores and 16 test papers from lowest scores.

3. The other test papers (middle scores) are kept aside as they are not needed in the item analysis.

4. Put in tabular form the number of candidates in the upper and lower group who selected each alternative response for each item of the test. This can be carried out on the back of the test paper or a separate test item card may be used.

This area of the review centred on the possible links between test item statistics and the development of creativity in students during studies and after studies for job placements. Previously in our educational system in Cameroon, teachers neglected this aspect in dealing with tests. Little or nothing could be done to a test item after writing it down or copied from somewhere for the adaptability to the learners or content tested. With the coming of the examination boards in 1992, criteria or standards for test development were put in place and bit by bit teachers got impregnated in the ways and considerations of treating items for tests. Terms such as moderation, deliberations, pilot testing came into play. This has mostly been applied by conscious and conscientious teachers who teach examination classes or who are selected examiners of the GCE.

Today with the regular demand for educational reforms by the stakeholders, civil societies, the government and the international bodies, with the need for quality education as demanded by the sustainable development goals of the United Nations, much is being attempted towards the promotion of quality test items that would boost the learning domains of cognition, psychomotor and socio-affective domains of learners such that they develop creativity as they study and after studies.

It is hoped that with the development of this study, the educational stakeholders could see the direct link or relationship between test item statistics and creativity in learners. The following reviews will throw light on the relationship between test item statistics and creative student development in learners of Biology.

With the phenomenon of teachers teaching following the test questions of previous examination sessions, teaching to the test has become so rampant. According to Nenty and Fotegang (2014), teaching to the test focuses on contents

on which items of the previous examinations has been developed to assess hence limiting the curriculum to content and cognitive behaviour that are easily measurable. By teaching to the test, the development of low-order cognitive skills among learners tend to perpetuate itself because given the general inability of public examination to develop items to measure higher-order cognitive skills, teaching to the test means teaching students for the development of lower-order skills s called for in public examinations. For example, according to Jerald (2006): In Biology, learners who regularly have been having only drill exercises in test-like items have not got the opportunity to have a mastery of a particular skill or concept and most probably cannot give correct answers to questions that assess the same skills or concepts in a different way. As example, we found in one study that in a centre which highly depended on item drilling, 83% of the candidates chose the right response to a multiple-choice item written as "87-24=". But only 66% could give the right response to the open-ended item "subtract 24 from 87". From the observation put forward by Nenty and Fotegang (2014), teaching to the test exaggerates performance at the expense of learning. Implementing a distorted curriculum, that is, assessment-driven syllabus, results in education's inability to develop in the learners all the skills and knowledge deemed essential by the society for growth and development. In this light, sustainable student development is not assured.

This phenomenon in Rwanda also applies to Cameroon, as many teachers teach to the test. As such only the lower levels of the Bloom's taxonomy are mostly measured (knowledge, comprehension and application). The ability to analyse, synthesise and evaluate or create remains still to be addressed in the learners.

## Test Item Distractor Analysis and Creative Student Development in Biology

This section of the review involves mostly the multiple-choice questions with three or more alternative answers from which the right and correct answer should be selected or chosen. The manners in which these alternative answers are presented have a lot to influence the creativity of the students. From the content covered through the objectives of the test to the purpose of the test, the right answer to a particular question should not be as different from the distractors as to give a clear view of the obviousness of the right and correct answer.

In other words, the wrong answers or other alternatives of the right answer should turn around the same concept, content and objective such that the respondent must have mastered them before being able to distinguish between them and having the correct one. In this way the competency and creativity of a student can be measured in a given concept content and objective.

For distractors to be in content this study expected that the domain taught should not be over extrapolated to different cultures, environment or similar domains in different subjects. For example, including cactus plant and date plant as distractors to leafy plants or answers to succulent fruits producers in a school typically in the forest zone, this would put the learners out of the rails as they might have never seen or read about the plants, thus not being familiar with them. This example takes into consideration the context, culture and the environmental realities of the place of testing where the students have also studied. The failing of the students in the tests would therefore be caused by many of the listed factors.

## Test Validity Analysis and Creative Student Development in Biology

The development of classroom tests in our schools from experience has been neglecting the aspect of validity. Many teachers go for past question or quickly set

and administer tests without ensuring the validity of such tests. The issues of item statistics, item discrimination, and item difficulty or distractor analysis are far from being put in place.

Validity tells us whether the test is measuring what it purports to measure. For example, a test on microscopy is supposed to measure student competency in viewing and drawing the representation seen. But if it includes worded descriptions, a test as such may serve as a challenge for candidates with poor skills in English language. In effect it would not be simply measuring skills in microscopy but also skills in English language. Therefore, the test would measure the two; microscopy and English language skills for one subgroup of candidates, microscopy alone for the rest of the candidates for whom language does not pose any problem. This presents a difference from the stated goal of being a test on microscopy alone for all the candidates.

Test validity also goes in the sense of ensuring that what the society has validated through the curriculum be tested from the curriculum and not from examination syllabuses. That is the process of education and therefore of testing is valid to the extent that it is curriculum-driven. If on the contrary, the test is based on the set assessment syllabus, then it is cheating the society of the values validated in the curriculum, it tends to narrow the curriculum content to only that which could be assessed, given teacher's inability to develop valid test for complex human behaviour, and hence to contents that have been measured previously. It therefore distorts the curriculum, hijacks education intentions or cheats the society of her intentions. It is the curriculum, not the assessment syllabus that should serve as the criterion in any attempt to assess the validity of education in any society. Nenty, H and Fetogang, (2014)

**Face validity**

This is validity in its simplest form. Whether subjectively or objectively verified, test items or question should appear to measure what is expected of them to measure. Akoko, (2016).

Most often the test items set by some teachers tend to scare the test takers in that just on seeing the items, they either feel out of place in that they were not familiar with the content area or that the test was for a superior or inferior group of learners. With this many learners lose the zeal or enthusiasm to take part in the test. Or tend to lose concentration. The items could also be out of context if unfamiliar organisms or unfamiliar environmental issues are involved. This would limit the students in applying what they had learned in their contexts thus killing their spirit of creativity.

**Criterion validity**

There are two sub types of this validity; the concurrent validity and the predictive validity. The criteria used here depend on the two types. Criterion validity finds out whether the test gives results in agreement with other measures of the same thing.

Concurrent validity involves comparison of the new test with another established test. Predictive validity verifies whether the test predicts some future events (e.g., exam results). It is of advantage to the learners if they are aware that a test takes the form, they are familiar with, and that the results would be published sooner or later. With this knowledge, students are at ease with the test and could easily be boosted in their creativity in order to ensure success and create new records in the future in terms of performances or other outcomes.

**Construct validity**

This validity verifies if the test taps the concepts studied. The concepts should be very common as those studied or familiar in context. Just like the structure of a cell, this validity ensures that the presentation of the test be that which the learners have been used to, in order to boot their attention and given the zeal to answer the questions. The face validity and the predictive validity form part of the construct validity.

**Ecological validity**

This validity verifies whether the test involves naturally (or spontaneously) occurring behaviour (or attitudes etc.). These aspects if taken into consideration in the test will naturally incite curiosity in the learners and making them feel at ease as the test addresses what they are habitually found in, and involves spontaneous behaviours which they could tend to modify or innovate for better results to the satisfaction of every stakeholder. If laboratory studies are instead done in the natural settings, there are spontaneous attitudes and aptitudes that may lead to creativity for the student in either modifying the prevailing conditions or the activities carried out, thus developing creative tendencies.

**Population validity**

This validity is verified if the right sample or sampling is considered in the test. If because of large population estimation, a test is set, only to find out that the population is not large but sparsely populated, the estimation may affect the test takers who are few. This would limit their spirit of creativity in that some aspects like time may play highly on their psyche.

## Test Reliability Analysis and Creative Student Development in Biology

Reliability gives us information on whether a test is probably going to provide similar findings if administered to the same group of examinees several times. One other character of reliability is that the items of the test should have similar behavior with diverse populations of test takers which generally mean that the items should have approximately the same ranking when sorted by their p-values which are indicators of "item difficulty." The items on a biology test administered to Form five students in GBHS Bayelle in the South West region should show similar p-value rankings when administered to Form five students in St. Mikael Academy of science and Arts (SMASA) Bamenda North West region. The administration of the same test the next year to another cohort of form five students should once more show similar rankings. Great variations in the rank of item where one year it was the easiest item; the following year it was the most difficult on the same test would expose a problem with the two items and by implication, the test itself.

## Test Grading Analysis and Creative Student Development in Biology

Test grading is that part of testing or evaluation in which the learners or those who have taken the test are totally exempted in any stage of it. Nonetheless the process needs the consideration of many contextual issues in which the test takers found themselves in order to provide adequate and truthful report. In this light, the test graders have to involve the test takers passively. This can be seen in situations in which in the course of the learning process, no forecast of any sort was predicted to affect the testing, but in the course if testing many abnormalities crop up. A good example is the coming of the COVID-19 pandemic in which learners within a few weeks had to learn to wear face masks which they had never adapted to wear. The impact of this phenomenon may be ignored during the

writing phase, but the overall performance may be affected, thus the need to consider the possible impacts of such circumstances during the grading. The impact can be addressed in many ways: Reducing the grading range for various grades, reducing the lower marks for those below the pass line, or deliberating on general cutoff points for successful candidates. We could also consider the highly hit zones and the psychological effects therein in terms of the effect related to testing and the outcomes, thus, grading in favour of the learners. This brings in the theory of true score.

With these possible inputs in test grading the learners' ability to be creative even in times of crisis or pandemics would be greatly improved that they may not only count on the guidance of the teachers but also on their personal abilities to make things happen positively in their favour. Critical thinking, logical analysis, adaptability and innovation would be eminent situations.

The war situation in the North west and south west regions, in which learners found themselves too was a factor to consider during grading of tests. Students become creative if they can apply what they learn in Biology to the real-life situation to address their personal needs and that of a larger population in a bid to find solutions to them. In this light, learners had to adapt to the prevailing conditions of solving their problem of food scarcity by indulging in practical Biology and boosting crop production to feed the dying population.

When this type of adaptation occurs where learning has taken place, we expect that test structure would not be the same and the grading of tests would not take the same protocol like that of other regions where no war or crises were present at the time of studying and development of the tests. A practical example was in the year 2000 when I was to graduate from the GTTC Bangem, I made a small garden of mixed cropping where there was none before. It was fenced to

prevent animals from entering. The flourishing of the crops drew the attention of the passersby including my teacher in charge of agriculture. This practice gave me an 'A' grade in the final course examination without me being evaluated at the farming site of the examination. Instead, I was asked to direct others on how to go about theirs for similar successes.

**Cultural Relevance**

This research work is a pertinent work for African countries as a whole. Africans have their own way of learning peculiar to them and their varied cultures and different from others elsewhere, even to those Africans who have travelled and studying abroad. It is therefore, time that we addressed the issues of African education in their tests, such that they are nurtured in what they know best and are familiar with, than what is imported from others. A proper use of the outcome of this research work will go a long way to transform our African society, instead of being transformed to suit other societies abroad.

**Appraisal of the Chapter**

Chapter two has handled everything concerning the literature review in relation to the topic of this research. It began with the major theories that guided the work to the review of the past works in relation to the various independent and dependent variables related, to the conceptual framework and some empirical studies of the research. To back the independent variable:" Test Structure Analysis", the researcher used the following theories: The Classical Test Theory (CTT), the Item Response Theory (IRT) and the True Score Theory (TST) even though found as component of the Classical test theory. The Classical Test Theory (CTT) which predicts outcomes of psychological testing such as the difficulty of the items or the ability of the test takers. Novick, (1966). The main components of

Classical Test Theory (CTT), McDonald, (1999, pp. 4-8) are: Classical true-score theory, Common factor theory.

The four major statistics examined or reported in the framework of CTT are: Item difficulty, Item-test correlation, Reliability coefficient and the Standard error of measurement (SEM). The Item Response Theory is aimed at measuring mostly the psychological traits like students' abilities, attitude and many other psychological traits as reported from the responses collected from the items in the test. The true score theory has been used to postulate to the examination or testing stakeholders that there could be a score that could present the reality of the learners taking into consideration their different abilities and the possible error that could not be neglected.

To back up the dependent variable; "Creative Student Development in Biology", the chapter has exhaustively explained the theory of motivation by Abraham Maslow in all its forms and applications in the classroom, the socio-constructivist theory by Vygotsky, the Flow theory by Miyali and the self-determination theory which emphasizes on intrinsic and extrinsic motivation going to buttress the theory of motivation. This chapter thus considered three theories to the credit of the independent variable and four to the credit of the dependent variable. However, the motivation theory and the autotelic flow theory both have served to lay down the foundation of the work in backing both the predictive variable and the criterion variable. Thus of the six main theories used, two are for the independent or predictive variable, two for the criterion or dependent variable and two for both variables.

The theories for this study are mostly imported theories which may one way impart Africa, but the researcher has adapted them to suit the African learning context in general, the Cameroonian context in particular, taking into consideration

the various regional and cultural contexts to be specific. It is from such a study that we have built and developed practices and models pertinent to our African and Cameroonian contexts.

# CHAPTER THREE

## RESEARCH METHODOLOGY AND DESIGN

This chapter comprises the introduction, research design, area of the study, sample size and sampling technique, instrument of data collection, validation of the instrument, reliability of the instrument, administration of the instrument, method of data collection and analysis, ethical consideration, synoptic table, operational definition of key variables and the conclusion.

This study used both qualitative and quantitative research method, considered in research as mixed method. The researcher decided to use the two methods in order to provide evidences as gathered from respondents in the field in terms of their general views in the descriptive part and to also provide empirical statistics of what is actually taking place on the ground. Particularly the researcher chose the qualitative approach because the researcher intended to understand and describe how the creative student development could be influenced by the analysis of a test structure. According to Creswell (2003, p.179-181), qualitative researchers are interested in meaning; that is, to discover and understand how people construct meaning out of the way in which they perceive their lives. Thus, given the nature of this research, a qualitative approach would help us unpack the manner in which test structure analysis could solve the problem of students' creativity development in Biology during school and after school.

Strydom in De Vos et al. (2011, p.202) stresses, that, an important task in a qualitative research is deciding on those to participate in the research. The researcher chose only those participants who would be able to supply the requisite information; be willing to participate in research and ready to share information Morse and Richards, (2002, p.20).

The collection of data is the vehicle with which researchers collect information to answer the research questions and defend conclusions and recommendations based on the findings from the research, Mertens in Mahlo, (2006, p. 31). For Creswell, (2003, p. 181), it employs multiple methods that are both interactive and humanistic, in this case observation, semi structured interviews and document analysis, to establish a link between reality and the theoretical assumptions, Mouton and Marias, (2003, p. 156). The researcher made use of in-depth face to face interviews to enable him to gather the information about how the teachers set and analyze test structures for the development of creativity in learners. One-on-one interviews not exceeding thirty minutes were conducted at the respondents' convenience until data saturation was reached.

Qualitative analysis was employed throughout the data collection process where the analysis commenced with viewing of all the data and then dividing the data into smaller and more meaningful units, Henning et al, (2004, p.127). The one compared the material within the subcategories to discover connections between the categories and themes.

In this study, the constant comparative method was used. The method entails: familiarization and immersion, coding into themes, inducing themes, elaboration and interpretation and checking. The above methods, the motivation for the choice of the methods and how the researcher went about using them has not been discussed in detailed because the study on Psychometrics demands more of quantitative measurements than qualitative measurements. Thus, more emphasis has been laid in the quantitative report.

The quantitative method was also used in this study because the researcher by the domain of study intended to provide numerical data that would give facts and figures as collected in the field. These facts and figures are what make the

research to be verified as people may change but certain facts in terms of the interfering variables may remain constant. The researcher used the questionnaire to collect certain data that permitted him to interpret numerically the relationship between the analysis of test structure and the creative student development. A research in these same areas and in the view of verifying the data could be carried out subsequently to ensure the validity of the work. Considering that the researcher's domain of research falls in the field of educational measurement and precisely Psychometrics, emphasis would be put in the reporting of the two methods; qualitative and quantitative data because both were collected with due respect of their approaches and procedures peculiar to each.

**Research Design**

According to Mbua N. (2003) research design is the blue print, methodology or plan of activities that the researcher uses in carrying out investigation in a given area of problem. In other words, research design refers to the researcher's plan on how to proceed as in his study. From the above definitions, we deduce that research design is the plan and strategy put in place to carry out the research project. Amin (2005) noted that a research design is necessary because it guides the entire research process so as to yield maximum fruits and reduce cost or expenditures in terms of effort, time and money. Without a research design, it would be difficult to set meaningful boundaries.

The research design used for this study is the survey research design. Surveys are used to gather data from a sample of a population at a particular time and the findings are later inferred on the entire population of the study. Surveys are used to obtain information on peoples' attitudes, practices, concerns, preferences or interests.

The researcher used the survey design because the opinions of the participants were necessary to obtain information on the various variables of the study. David and Sutton (2004) express the view that survey designs are more suitable when dealing with the opinions of respondents. Moreover survey design is suitable for collecting data from a large population Bhattacherjee, (2012) thus it was appropriate to use the survey research design considering the fact that the research is dealing with a large sampled population of 1603 students and 36 teachers drawn from a larger group respectively.

**Area of the study**

The study covered the whole of Cameroon, taking into consideration the representation of the main geographical and cultural areas. It covered the four main geographical and cultural zones namely; the sudano-sahelian zone at the north, the savanna or grass field zone of the west, the coastal or littoral zone of the south west and the equatorial forest zone of the south, centre and the east.

This study covered particularly Yaounde in the centre region of Cameroon and its urban peripheries. It comprised all the bilingual secondary schools and English speaking secondary schools in the subdivisions of Yaounde 1,2,3,4,5,6 and 7 of the Mfoundi division, the English speaking secondary schools or Bilingual secondary schools of the Bafoussam 1,2 and 3 subdivisions of the Mifi Divisions, St. Paul's college Nkwen-Bamenda, the English speaking secondary schools or bilingual secondary schools of the Douala 1,2,3,4,5 and 6 sub divisions of the Wouri Division and Bilingual Grammar School Molyko Buea and the English speaking secondary schools or bilingual secondary schools of the Ngaoundere 1,2 and 3 subdivisions of the Vina division.

Yaounde is the National Political capital of Cameroon. It is also the divisional headquarter of the Mfoundi division found in the Centre region of Cameroon. It has developed in infrastructure to encroach into subdivisions of other divisions surrounding it. A consideration of Yaounde, leaving these extensions will therefore have an effect on the results as many learners leave from one area to the other in quest of the right schools that offer the quality they expect. Yaounde (Mfoundi division) is bounded to the North west by the Obala sub division of the Lekie division, to the North-East by the Soa sub division of the Mefou & Afamba division, to the South East by the Mfou sub division still of the Mefou and Afamba division, to the South west by the Mbankomo sub division of the Mefou & Akono division. Yaounde (Mfoundi division has a total surface area of 297.8 Km2 as of 2005.

Douala is the economic capital of Cameroon; it is the regional headquarter of the Littoral region, it is the largest city only comparable to Yaoundde in terms of population and size. It is found at the coast. Douala is entirely the Wouri division of the Littoral region. The Wouri division is bounded to the north by the Moungo division, to the east by the Nkam division, to the south by the Sanaga maritime division and to the west by the Atlantic Ocean.

Bafoussam is the regional headquarters of the west region and the divisional headquarters of the Mifi division. It is made up of the Bafoussam 1,2 and 3 Subdivisions. It is bounded to the north by the Bamboutos division, to the east by the noun division, to the south by the Koungki division and to the west by the Menoua division.

Ngaoundere is the headquarters of the Adamawa region, it is the seat of the Vina division. It is made up of Ngaoundere 1, 2 and 3 subdivisions. The Vina division is bounded to the north and east by the Mayo Rey division, to the west by

the Faro and Faro & Deo divisions, to the south by the Djerem and the Mbere divisions.

Bamenda is the regional headquarter of the North West region and the divisional headquarter of the Mezam division. The school of interest here is the St Paul's Bilingual Comprehensive College Nkwen-Bamenda. The Mezam Division is bounded to the North by the Menchum and Boyo divisions, to the west by the Momo division, to the south by the Lebialem division of the South West region and the Bamboutos division of the West Region, and to the East by the Ngoketungia division. Bamenda is bounded to the north by the Bafut and Tubah sub divisions, to the west by the Mbengwi subdivision, to the south by the Bali sub division and to the east by the Santa sub division.

Buea is the Capital of the South West Region of Cameroon. It is found in the Fako division whose headquarter is Limbe. Buea is the only regional capital which is not a divisional headquarters. It is bounded to the north by the Muyuka sub division, to the west by the Mbonge sub division, to the south by the Limbe 1 subdivision and to the east by the Tiko sub division.

Cameroon is found between the west and central Africa. It is bounded to the north by the Lake Chad, to the west by Nigeria, to the north east by Chad, to the east by the Central African Republic, to the south by the Congo republic, Gabon and Equatorial Guinea; to the south-west by the Atlantic Ocean. The map of Cameroon showing the various cultural spheres is attached to the appendix.

The map presents the various areas of Cameroon which constituted the population of the study. The areas are endowed with geographical specificities and cultural peculiarities that led to the various naming as:

Sudano-sahelian sphere comprises the Adamawa, North and Far North regions, with Ngaoundere, Garoua and Maroua respectively as their regional headquarters. Ngaoundere was used as a representative of this sphere.

The Fang-Beti sphere includes the Centre, East and South regions with Yaounde, Bertoua and Ebolowa respectively as the regional capitals. Yaounde was used as the representative town of this sphere.

The Grass field sphere includes the North West and the West regions with Bamenda and Bafoussam respectively as the regional capitals. Both towns were used in the population of the study; one school from Bamenda and two schools from Bafoussam. This was so because only few schools were functional in the town of Bamenda due to the war situation in the two English speaking regions.

Finally the Sawa sphere made up of the Littoral and the South West regions with Douala and Buea respectively as the regional capitals. Both towns contributed to the population of the study with Buea providing one school and Douala providing two schools. This was so because only few schools were functional in the town of Buea due to the war situation in the two English speaking regions.

This study therefore covered the whole of Cameroon taking into consideration the differences in cultural background, environmental background, climatic background and specificities that may influence learning. The Sudano-Sahelian zone predominantly dominated by the fulbes, the matakams, the massas, the toupouris, the moundangs and the guizigas culturally use pipes and horns to make music and their biodiversity is made up of semi-arid flora and fauna adapted to the dry and hot climatic conditions sometimes in the day and extreme cold in the nights. The study of Biology in this area would not be the same as that of other areas.

The grass field area on the other hand as the name indicates, is predominated by huge grasses mostly at the top of the undulating hills and their slopes while some valleys are covered with thick savanna forests. The humidity of the area and the varying of the climate up and down the slopes give the area a unique feature peculiar only to them. The fauna and flora therefore vary as the topography varies making the study of Biology particular to the area and different from others.

The Coastal or Sawa zone from all indication is full of variety of diversified flora and fauna, both aquatic and terrestrial. The presence of a tropical rain forest and the mangroves makes the area typically different from others when it comes to the study of Biology.

The Fang-Beti or Centre –South zone is characterized by a dense tropical rain forest, on a practically low lying plain. This area is less mountainous and full of fresh water bodies in the dense equatorial forest. The flora and fauna species are particular to the area especially as the area stretches to other neighbouring countries. The study of Biology in this zone cannot be the same as in others.

The differences in the geography of these areas in Cameroon permitted the researcher to carry out sample populations from them so as to have a global picture of it in Cameroon and propose a way forward. The particular schools used as experimental grounds in this study were grouped into government schools, denominational (or mission) schools and the lay private schools. The schools chosen from the various regions were as follows:

Adamawa Region

Government Bilingual High School (GBHS) Ngaoundere (Government institution), the MAZENOD Bilingual college Ngaoundere (Catholic Mission institution) and Grace Bilingual College Ngaoundere (Lay private institution).

Centre Region

Government Bilingual High School Etoug-Ebe-Yaounde (Government institution), Saint Benedict Bilingual College Mvolye-Yaounde (Catholic mission institution) and Queensway Bilingual College Nkolmesseng-Yaounde (Lay private institution).

Littoral/South West Regions

Bilingual Grammar school (BGS) Molyko-Buea (Government institution), Sacred Heart College (SHC) New Bell-Douala (Catholic Mission institution) and National Bilingual College (Nabicol) Bonaberi-Douala

North West/ West Regions

Government Bilingual High School (GBHS) Gouache-Bafoussam (Government institution), Saint Paul's Bilingual College Nkwen-Bamenda (Catholic Mission institution), Reunification Bilingual College Bafoussam, (Lay private institution)

The government, the mission and the lay private institutions each had three schools drawn from the four main cultural and geographical areas of Cameroon. Altogether six out of the ten regions of the country were used in the study. It should be noted that the schools in the various towns were also located in different sub divisions or administrative set-ups of the towns, making the population a heterogenous one.

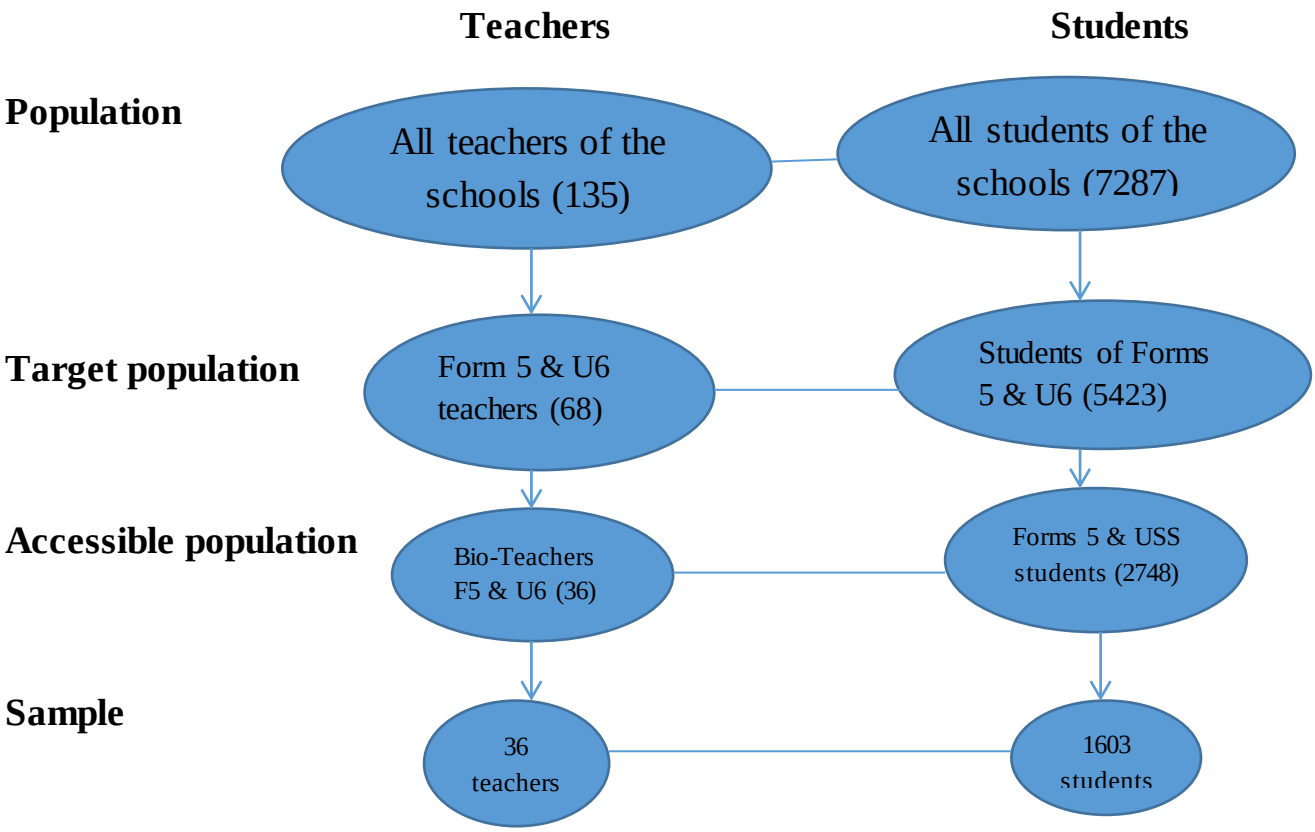
The map of Cameroon showing the various cultural zones and the regions involved has been drawn and annexed to the appendix to this work.

**Population of the study**

Geographically, a population of a place is the number of individuals living in a given area. This study did not consider the general population, but was interested in the population that has direct impact to the study, though the results would be of great interest to the entire population. The population of the study was discussed under three different types; the target population, the accessible population and the sample. This study considered as general population, all the teachers and students of the selected English secondary schools and colleges, as well as the bilingual secondary schools and bilingual colleges in the various cultural/geographical spheres used in the study. The population of this study was further presented under the target population from which we got the accessible population and from the accessible population was drawn the sample. The target population comprised the teachers of Biology of forms five and upper sixth classes of all the selected schools in the selected towns where the study was carried out, and all the form five and upper sixth students of the selected schools. The accessible population was made up of the science students of form five and upper sixth science who offered to study biology, and the teachers who taught biology in the school. The sample was drawn from the form five science classes and upper sixth science Biology students. The distribution of the population of the study to the target, the accessible population and the sample is illustrated in the figure below:

Figure 3

*Operationalization of the population to the study*




***Source:*** The Researcher

Figure 3 illustrates that the population of the study was the largest' followed by the target population and lastly the accessible population. It was from the accessible population that the sample was drawn using the simple random sampling technique. This has been depicted by the sizes of the circles which get smaller as we move from the population down to the sample.

The presentation of the population of the study could also be shown in the table that follows in figures.

Table 5

*The population of students and teachers of the study*

| **Region** | **Division** | **Town** | **No. Students** | | **No. Teachers** | **Total** | |
|---|---|---|---|---|---|---|---|
| | | | A | O | | S | T |
| Adamawa | Vina | Ngaoundere | 313 | 403 | 25 | 716 | 25 |
| Centre | Mfoundi | Yaounde | 820 | 1216 | 39 | 2036 | 39 |
| Littoral/SW | Wouri/Fako | Douala/Buea | 1452 | 1336 | 38 | 2788 | 38 |
| West/NW | Mifi/Mezam | Bafoussam/Bamenda | 916 | 1330 | 33 | 2247 | 33 |

From table 5, the advance level students are represented by the letter 'A', and the ordinary level students by the letter 'O'. 'S' represents the total number of students while 'T' represents the total number of teachers. The four cultural and geographical zones are represented by the regions and their corresponding divisions and the towns of the research.. The 'Sawa' sphere (Douala/Buea) with 2788 students was the first in terms of student population but second in terms of teacher population with 38 teachers The 'grass field' sphere ( Bafoussam/Bamenda) had the second position in terms of student population considered in the study, with 2247 students and third largest population of teachers; 33 teachers. The 'fang-Beti' sphere (Yaounde) was third in terms of the student population but first in terms of teacher population with 39 teachers. The 'sudano-sahelian' sphere (Ngaoundere) is the fourth in both student and teacher population; 716 students and 25 teachers, as presented in the study.

Table 6

*The population per region, town and school*

| Region | Town | School type | No. of students O | No. of students A | Total number of students O+A | No. of teachers |
|---|---|---|---|---|---|---|
| Adamawa | Ngaoundere | Gov't | 249 | 187 | 436 | 12 |
| | | Mission | 90 | 71 | 161 | 08 |
| | | Lay private | 64 | 55 | 119 | 05 |
| Centre | Yaounde | Gov't | 758 | 481 | 1239 | 24 |
| | | Mission | 303 | 224 | 527 | 09 |
| | | Lay private | 155 | 115 | 270 | 06 |
| Littor/SW | D'la/Buea | Gov't | 999 | 640 | 1639 | 23 |
| | | Mission | 77 | 104 | 181 | 09 |
| | | Lay private | 260 | 208 | 468 | 06 |
| West/NW | B'sam/B'da | Gov't | 557 | 394 | 951 | 19 |
| | | Mission | 328 | 272 | 600 | 08 |
| | | Lay private | 446 | 250 | 696 | 06 |
| Total | | | 4286 | 3001 | 7,287 | 135 |

Table 6 is an elaborate presentation of table 3; the population of the study, to include the type of school and the population input to the study. Each type of school is presented in the third column and the number of student population in the advance level 'A' and the ordinary level 'O'.

**Population distribution according to the selected towns/Regions**

The study was carried out in selected towns; especially the headquarters of the regions chosen for the study. The cosmopolitan nature of these towns enabled the schools to be suitable for the study because learners came from the various

geographical areas of the region and diverse ethnic groups to study Biology in the schools which could also be said to be better equipped for biological studies and evaluation than the others. The schools selected in the regional headquarters are well indicated in the presentations that follow;

**Adamawa**

According to the statistics collected from the GCE A/ Levels results 2020 from the GCE Board on 2nd of October 2020 there were twelve schools that served as centres for the GCE. Six were in Ngaoundere while the other six distributed to the other divisions and sub divisions of the Adamawa region. All the three schools selected offered Biology to the students. The population of biology students was as follows: first cycle 403 students, second cycle 313 students. The population of teachers was twenty five (25) teachers for both first and second cycles. The following schools constituted the population of the study:

- Government Bilingual High School (GBHS) Ngaoundere

- Grace Bilingual Secondary School - Ngaoundere

- Mazenod Bilingual College – Ngaoundere

Table 7

*The population from schools in Ngaoundere*

| **Type of school** | **Number of schools** | **Number of students** | **Number of teachers** |
|---|---|---|---|
| Government schools | 01 | 436 | 12 |
| Lay private schools | 01 | 161 | 08 |
| Denominational (Mission) schools | 01 | 119 | 05 |

Table 7 simplifies table 5 by sorting the region of Adamawa (Ngaoundere) in presenting the type of schools and the number of schools type hosting the population of the study.

**Centre Region**

According the statistics collected on the 2nd of October, 2020 from the GCE Board publication of the A-level GCE examination results, there were about one hundred and thirty schools which served as the GCE examination centres in the Centre, South and East regions, with about ninety-six centres in Yaounde alone. All the three schools selected for the sample all offered Biology to the students and thus, the teachers who are active examiners of the exam and/or markers. The population of biology students was as follows: first cycle 1216 students, second cycle 820 students. The population of teachers was thirty-nine (39) teachers for both first and second cycles. The following schools constituted the population of the study:

- Government Bilingual High School (GBHS) Etoug-Ebe-Yaounde

- St. Benedict College Mvolye-Yaounde

- Queensway International College Yaounde

Table 8

*Shows the distribution of the population according to the type of school*

| Type of school | Number of schools | Number of students | Number of teachers |
|---|---|---|---|
| Government schools | 01 | 1239 | 24 |
| Lay private schools | 01 | 527 | 09 |
| Denominational (Mission) schools | 01 | 270 | 06 |

Table 8 is a presentation of the various types of school and the corresponding number of each type of the school that hosted the population of the study.

**Littoral/South West Regions**

According to the statistics collected from the 2020 GCE A-Level results on the 2nd October.2020, eighty-two schools served as centres for the GCE A-level 2020 examination in the South West Region. In the Littoral Region, one hundred and seven schools were noted as centres for the 2020 A-level GCE examination. All the schools selected for the study offered Biology to the students and therefore had Biology teachers who possibly mark the examination and/or set tests for it. The population of biology students was as follows: first cycle 1336 students, second cycle 1452 students. The population of teachers was thirty-eight (38) teachers for both first and second cycles. The following schools constituted the population of the study:

- Bilingual Grammar School Molyko-Buea

- Sacred Heart College Douala

- National Bilingual College Douala

Table 9

*Population according to the type of schools in Douala/Buea*

| **Type of school** | **Number of schools** | **Number of students** | **Number of teachers** |
|---|---|---|---|
| Government schools | 01 | 163 | 23 |
| Lay private schools | 01 | 181 | 09 |
| Denominational (Mission) schools | 01 | 468 | 06 |

Table 9 presents the types of schools and the corresponding number of the type of school that hosted the population of the study.

**West/North West Regions**

According the statistics collected on the 2nd October 2020 from the GCE 2020 A-Level results, there were twenty-six schools which served as GCE centre in the North West region and Forty-three in the West region. One school from Bamenda in the North West region contributed to the population of the study because schools were not fully functional in the region and only few schools were open in Bamenda centre, while two schools from Bafoussam contributed to the population of the study because many of the students in the North West as a whole and Bamenda in particular had run to attend school in Bafoussam. The population of biology students was as follows: first cycle 1331 students, second cycle 916 students. The population of teachers was thirty three (33) teachers for both first and second cycles. The following schools constituted the population of the study:

- Government Bilingual High School (GBHS) Gouache - Bafoussam

- St. Paul's Bilingual Comprehensive College Nkwen-Bamenda

- Reunification Bilingual College – Bafoussam

Table 10

*Population according to the type of school in Bafoussam/Bamenda*

| Type of school | Number of schools | Number of students | Number of teachers |
|---|---|---|---|
| Government schools | 01 | 951 | 19 |
| Lay private schools | 01 | 600 | 08 |
| Denominational (Mission) schools | 01 | 696 | 06 |

Table 10 presents the population according to the type of school and the number of schools per type of school involved in the study.

**The Target Population**

The target population is a sub set of the population as a whole. While there was the population of teachers in the various schools that teach different subjects and classes, the target population is limited to a particular group of teachers; teachers of particular classes. In the same light the population of students who study in various subjects and classes was reduced particular to those who were in the forms five and the upper sixth classes.

The target population therefore comprised all the teachers of the selected schools intervening in the form five and upper sixth classes. This is a sub set of the total population of teachers in the school. It also comprised all the form five and upper sixth students of the selected schools. This also as a sub set of all the students of the school.

Table 11

*The target population in the types of schools*

| Region | Town | School type | Number of students F5 | Number of students U6 | Total number of students | No. of teachers |
|---|---|---|---|---|---|---|
| Ad. | Ngaoundere | Government | 249 | 56 | 305 | 05 |
| | | Mission | 90 | 21 | 111 | 04 |
| | | Lay private | 64 | 16 | 79 | 04 |
| W/NW | Baf'sam/B'da | Government | 557 | 157 | 714 | 05 |
| | | Mission | 328 | 108 | 436 | 04 |
| | | \Lay private | 446 | 100 | 546 | 03 |
| Lit/SW | D'la/Buea | Government | 999 | 228 | 1227 | 13 |
| | | Mission | 77 | 41 | 118 | 04 |
| | | Lay private | 260 | 83 | 343 | 03 |
| Centre | Yaounde | Government | 758 | 192 | 950 | 14 |
| | | Mission | 303 | 89 | 392 | 06 |
| | | Lay private | 155 | 46 | 201 | 03 |
| | | | 4286 | 1137 | 5423 | 68 |

Table 11 presents the target population of the study. There were 4286 form five students and 1137 upper sixth students in all the selected schools in the study, giving a total of 5423 students that made up the target population. Sixty-eight (68) teachers

## The accessible population

Accessible population is an extract of the target population in that it considers a lesser number of individuals from the target population used in the study. It is from the target population that we have access to the accessible

population. Below is the presentation of the accessible population in the various regional capitals and the type of schools involved.

Table 12

T*he accessible population of the study*

| Region | Town | School type | Number of students F 5 | Number of students U6 | Total number of students | No. of teachers |
|---|---|---|---|---|---|---|
| Ad | Ngaounere | Government | 148 | 48 | 196 | 03 |
| | | Denominational | 73 | 19 | 92 | 02 |
| | | Lay private | 52 | 14 | 66 | 01 |
| W/NW | Baf'sam/B'da | Government | 226 | 108 | 334 | 03 |
| | | Denominational | 175 | 86 | 261 | 02 |
| | | Lay private | 205 | 80 | 285 | 02 |
| Lit/SW | Douala/Buea | Government | 278 | 144 | 422 | 08 |
| | | Denominational | 63 | 36 | 99 | 03 |
| | | Lay private | 115 | 70 | 225 | 02 |
| Centre | Yaounde | Government | 254 | 127 | 381 | 06 |
| | | Denominational | 169 | 70 | 239 | 02 |
| | | Lay private | 108 | 40 | 148 | 02 |
| | | | 1866 | 842 | 2,748 | 36 |

From the presentation of table 12, a total of 2748 students constituted the accessible population of the study, 36 teachers of biology were accessible for the study. From these, 1603 students were drawn for the sample and all the 36 teachers constituted the sample.

## Sample size and sampling technique

The sample is drawn from the accessible population using a particular sampling technique appropriate to the type of study involved. From the accessible

population, a sample of 1603 students was drawn using the simple random sampling technique and 36 teachers of biology were drawn from the accessible teacher population using the snow ball sampling technique because the teachers were not all present at the same time and so the recruitment of one teacher into the study by their responses to the questionnaire led to the teachers bringing another biology teacher of the school to respond to the questionnaire, making it snow ball technique.

The sample was drawn from government schools, denominational schools and lay-private schools in the towns of Yaounde, Douala/Buea, Bafoussam / Bamenda and Ngaoundere. These schools were among those which served as examination centres for the GCE A-Level and O-level 2020 and which offered Biology to the students. The Krejcie and Morgan table was used to help in the number of participants at the time of the study. Students were drawn from the form five and upper sixth classes and the teachers from those who teach the classes.

The number of students selected for the sample was based on the available number of students in the classroom at the time of the research. The Krejcie and Morgan table was used as reference each time of the exercise. Where the number of students was greater than the needed sample, the random sampling technique was applied where N pieces of paper with numbers on them were shared. Those who picked the pieces of paper with numbers formed the sample and were given questionnaire to answer while those who picked empty pieces of paper were exempted from answering the questionnaire. A total of 1603 student participants were selected using this method and they formed the sample of students.

For the sample of teachers, most of the schools had just the required number of teachers used at the time of the research. Since all teachers do not teach Biology at the same time or day, I gave questionnaire to those present who in turn gave to

others later and all were collected already filled on the next day. In some other schools, I gave the questionnaire to the teachers to fill at their convenience and to collect them later still at their convenience. A total sample of 36 teachers was gotten at the end of the process in all the selected schools involved in the study.

I used the range of given population size (N) and the sample size(S) as given by Krejcie and Morgan (1970).

The following table shows the distribution of the population to the sample of students:

Table 13

*Distribution of the sample of students*

| **Region** | **Type of School** | **Population** | **Target Population** | **Accessible population** | **Sample** |
|---|---|---|---|---|---|
| Adamawa | Gov't | 436 | 305 | 196 | 127 |
| | Missi | 161 | 111 | 92 | 73 |
| | Lay | 119 | 80 | 66 | 56 |
| Centre | Gov't | 1239 | 950 | 381 | 191 |
| | Missi | 527 | 392 | 239 | 144 |
| | Lay | 270 | 201 | 148 | 103 |
| Litt/SW | Gov't | 1639 | 1127 | 422 | 201 |
| | Missi | 181 | 118 | 99 | 76 |
| | Lay | 468 | 343 | 185 | 140 |
| West/NW | Gov't | 951 | 714 | 334 | 175 |
| | Missi | 600 | 436 | 261 | 155 |
| | Lay | 696 | 546 | 285 | 162 |
| 04 | 09 | 7287 | 5423 | 2708 | 1,603 |

The table 13 presented the sample as originated from the various populations from the target population through the accessible population to the sample as gotten from each region and school. The student sample was clearly calculated to form the total sample used in the study.

Table 14

*Distribution of the sample of teachers*

| **Region** | **Type of School** | **Population** | **Target Population** | **Accessible population** | **Sample** |
|---|---|---|---|---|---|
| Adamawa | Gov't | 32 | 05 | 03 | 03 |
| | Missi | 12 | 04 | 02 | 02 |
| | Lay | 24 | 04 | 01 | 01 |
| Centre | Gov't | 106 | 14 | 06 | 06 |
| | Missi | 39 | 06 | 02 | 02 |
| | Lay | 48 | 03 | 02 | 02 |
| Litt/SW | Gov't | 73 | 13 | 08 | 08 |
| | Missi | 48 | 04 | 03 | 03 |
| | Lay | 45 | 03 | 02 | 02 |
| West/NW | Gov't | 43 | 05 | 03 | 03 |
| | Missi | 27 | 04 | 02 | 02 |
| | Lay | 26 | 03 | 02 | 02 |
| 04 | 09 | 523 | 68 | 36 | 36 |

The sample of biology teachers as was drawn from the target population through the accessible population to the sample as shown in the table 14. The sampling technique used was the snow ball sampling technique.

**Research Instrument Design**

The various research instruments used in the study; questionnaire to students and questionnaire to teachers, semi-structured interview and observational

checklist were all designed by the researcher according to the needs of the research such that qualitative and quantitative information could be gathered and the data analyzed accordingly. A presentation of the various instruments and their designs were done as follows:

The researcher used varied data collection instruments to gather the required data for the study. To get information from the teachers and the students, the questionnaire was used. The questionnaire for the teachers was different from that of the students, but both of them had items to address the test structure analysis and the creative student development. Both questionnaires were made up of nine parts: The preamble which comprised my personal information and some instructions for the respondents, the section on demographic information, the section on instructions on how to answer the questions and the sections of the questions lettered from B to H. Each section had five statements which expected the respondent to place a tick corresponding to a number from 1 to 10 which rated their stand on how much they agreed or disagreed about the statement based on each hypothesis or research question. The numbers were under the headings as follows; 1,2 for strongly disagree (SD), 3,4,5 for somewhat disagree (SWD), 6,7,8 for somewhat agree (SWA), and 9,10 for strongly agree (SA). The section A needed personal and particular information from the respondents. They had to fill the blank boxes with an "X" in indicating the correct information pertaining to them. Both the teachers and the students answered closed questions on the same variable; test structure analysis as the predictive or independent variable while the questions on the dependent or criterion variable; creative student development in Biology, followed the same pattern.

Semi-structured interview was used in order to permit the researcher probe into the answers of the respondents in order to have further information and

clarification on the subject in question. The information given was recorded using the phone recorder and some written with pen on paper. Most of the information got from interviews was based on the enrollments of teachers and students, the presence of a body in charge of test development, analysis and banking, deliberation council, test validation system, approval system, test administration system and departmental councils. Also, information was got on the number of teachers who teach just the first cycle, who teach just the second cycle and those who teach both cycles, their qualifications and longevity in service or experience.

The check list was also used to record observable aspects that could influence the analysis of test structures and that could impact on student's creativity development. Other environmental aspects that could affect the study, like farms, fields, laboratories among other observable aspects were recorded during observation.

For the purpose of clarity and mitigation of possible errors, the following instruments were used to collect both quantitative and qualitative data.

**Questionnaire**

The questionnaire for this study was for the students and for the teachers. It was designed to gather numerical data to be used for correlation. The one for the students were similar to that of the teachers such that the information given by the students would be validated or confirmed by the teachers or vice versa. Some sort of checking of the responses given by both respondents to ensure clarity in the responses and objectivity.

The ten-point Likert scale was used to provide options for closed responses from the respondents. Being a study in Psychometrics, the researcher expected more elaborate numerical responses to ensure numerical precisions in the analysis

of data. Thus, the ten-point Likert scale was used with more numerical and elaborate numerical responses. The questionnaire served as the main instrument of data collection. Being a psychometric study, quantitative report that concerns numbers took a toll over the qualitative report. Nevertheless, the issues reported quantitatively were not exactly those reported qualitatively. The interview and the observation were reported qualitatively in this study.

The questionnaire is an instrument used to collect quantitative data from the sample in the study. The questionnaire used in this study was made up of nine sections separately divided. The first part was introducing the researcher to the respondents, the second part labeled section 'A' was on demographic information from the respondents. The third part was the instructions to the respondents and the fourth part labeled B-G comprised the research statements to bee responded to.

The section A addressed the demographic information of the respondents in the form of spaces to be filled or boxes with options to tick, the sections B to G were tables of five questions each, addressed the individual predictive or independent variables of the study and the section G made up of five questions also addressed the criterion or dependent variable.

**Interviews**

The interview designed was the semi-structural interview. It was made up of items which served as leading or guiding questions, such that an extrapolation in the response could make the interview more elaborate or explicit. The semi-structured interview was designed to gather information from the heads of departments, pedagogic heads and principals who could give an insight into the issues to be addressed concerning test structure analysis and creative student development.

The interview used here was a semi-structured interview. Interviews are means of collecting information verbally from some respondents that would give it better than written. Interviews give qualitative information about an eye witness account of an event or a subjective analysis of one's personal view point or opinion about an observation which they took active or passive part. This information got from interviews go a long way to complement what was not handled in the questionnaire or to throw more light on what could not objectively be collected using the questionnaire. The semi structured interview was used as pre-prepared questions were laid down to guide the respondents and permitted them give responses in the light of the required data and not to let them go out of the range of the required responses. Semi structured interviews also serve to save time as the respondents are expected to go straight to the point by avoiding going off topic or adding unnecessary facts. The semi-structured interview asked questions in line with the theories used in the test development, items and/or test banking facilities, distractors conditions of consideration, validity and reliability analysis, test grading practices and relationship of the above with creative student development in Biology.

**Observation**

The observational checklist was designed in such a way as to find out was available in favour of a good test structure analysis and that could have a correlation with creative student development. It took the form of an inventory to support the predictive and criterion variables. This was more of a field work or hands-on sort of activity where the presence of the researcher was imperative and collaboration with the school authorities and the participants was at its reality and optimum level.

Observation is the process whereby the researcher gets first-hand information by actively taking part in the event (active observer) or passively taking part (passive observer). In the case of this study, both types were carried out as the researcher took active part in many of the school's meetings with the staff, departmental meeting, test writing sessions and marking guide setting and correction of the items and the test administration. The researcher also used all the human senses to get information from a distance or close to the information site. Check lists were also used to get an inventory of the required aspects for the study. The school site and location, presence of laboratories, farm land and fence or enclosure among others were observed. The checklist verified the existence of item banks, types of tests given, the tagging of validity or reliability coefficients on tests, marking guides, score sheets, mark allocations on tests and mark sheets among others.

**Validity of the instrument**

"The degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests", standards for Educational and Psychological Testing (AERA/NCME, 1999), gave this as a definition of validity in this context.

The various types of validity were ensured in the various instruments used in the gathering of data for analysis in the study, from the internal validity to the external validity. The various types of validity carried out included the face validity, construct validity, content validity, ecological validity and criterion validity. The procedures were as follows:

**Face validity**

Face validity is when a test appears to measure what it is supposed to measure at the sight of the test items. In this light the face validity of the questionnaire and the

other instruments used for this study were grouped under their various variables and the five variables and the criterion variable all had five items investigating on them. At first sight thus, an in relation to the topic of the thesis and the variables of the study, the face validity was ensured.

**Content validity**

The extent to which the instruments represented all facets of the study considering the variables used defined content validity. In this case the researcher ensured that all the variables of the study were covered using five items that explored the variable of both the predictive and criterion variable. In doing this the researcher ensured that the content validity was considered.

**Construct validity**

Construct validity involves whether the scale used, adequately measures the construct earmarked to be measured. In this case the ten-point Likert scale was used to provide a wide range of responses to be chosen by participants. Each items of the construct (variable) involved had answers ranging from 1, 2, 3- strongly disagree to 8, 9, 10 – strongly agree. We had 4, 5 – somewhat disagree and 6, 7- somewhat agree. Since each construct had five items, the construct validity in the study was ensured.

**Ecological validity**

This is an example of the external validity. It measures how generalizable findings are to the real world, such as situations or settings typical of everyday life. This validity was ensured in the choice of words used as he moved from one region to another in gathering the data. It is one of the wishes of the researcher that the development of tests and their analysis consider ecological validity in the selection of items for a test and the testing conditions.

**Criterion validity**

This included the predictive validity and the concurrent validity. It involves how accurately a test measures the outcome it was designed to measure. The concurrent validity measures test and criterion validity in the present, while predictive validity measures those in the future. By specifying the questions of the predictive and criterion variables, the researcher ensured the criterion validity in the choice of the items asked during the research.

The exercise of validation of the instruments could not be assured without the expertise of the supervisors and the carrying out of the pilot test. This part of the validation is presented as follows:

The instruments constructed by the researcher were proofread by some course mates peer review for clarity and language shortcomings, proper arrangement of the items and the representation of the test in relation to the domain; test structure analysis and creative student development respectively. This was to ensure the face validity before taking the test to the supervisors.

The content validity at this stage was done as follows: three of the test scripts were given to three different colleagues who on their individual time and corners approved or disapproved each item in the questionnaire. After collecting the questionnaire from each of them, the content validity index was calculated first per item then per research question.

For the one per item, out of the three mates considered as judges, if two accepted an item, the validity would be 2/3 giving the index for the question to be 0.67. If out of the three judges, only one validated an item, it would be 1/3, giving 0.33. If all of them validated an item, the validity of the item would be 1. If none validated an item, it would be zero (0) validity. An average of the validity indices;

for the five items on each research question was calculated. That is the average of the five indices was got by adding say 0.67+ 0.67+1+0.33+1= 3.67/5 =0.73. No item had an index of zero. This was done on all the five test questions and the items for the dependent variable.

The validity of the questionnaire was thus got by adding these average indices to have the final index of the questionnaire. The index was got as follows: 0.73+0.80+0.73+1+1=4.26/5=0.85. A validity index of 0.85 indicated that the test is 85% valid.

The co-supervisor also verified the phase validity and content validity. He checked the structure of the questionnaire and looked at the questions in relation to the variables of the study and the expected responses to be gotten. The ten-point Likert scale was proposed to be used to ensure regular interval in the scale. The researcher corrected the questionnaire and took them back to the co-supervisor for approval before being taken to the supervisor for final approval in a meeting with all the supervisees.

A pilot test was then carried out to find out if the respondents of similar characteristics but limited number could respond to the questions without constraints or complications and in order to calculate the reliability. Five copies of the questionnaire were given to five students of Biology of form five and five also to five Upper Sixth Biology students in Government Bilingual High School (G.B. H.S.) Bayelle-Nkwen – Bamenda were chosen at random and five to five other students of Biology of form five and five other students of Biology of Upper Sixth Science of St Mikael Academy of Sciences and Arts (SMASA) Bamenda. The questionnaire were directly given to the students by the researcher and collected all some few hours later, five days later the same questionnaire were given to the same persons for responses the all collected again, having a return rate of 100%. The

feedbacks from the respondents were studied and they showed that the instruments (questionnaire) used was at the level of understanding of the students, thus valid to measure the aspects for which they were constructed.

The same pilot test was carried out on five Biology teachers from GBHS Bayelle-Nkwen-Bamenda and two Biology teachers of the St. Mikael Academy of Sciences and Arts Bamenda. The responses were clear and they agreed the items were clear enough for them to understand. This ensured the face and content validities of the instrument.

**Reliability of Instrument**

A test- retest reliability using the questionnaire to students was carried out on the same population used for the pilot test and the reliability coefficient came out to be 0.94 signifying that the instrument was reliable in terms of its stability; that is, it was consistent in what it was meant to measure at different points in time.

This coefficient was calculated as follows:

Table 15

*Shows the computed scores used in the calculation of the test-retest reliability coefficient.*

| **n** | **X** | **Y** | **xy** | **$x^2$** | **$y^2$** |
|---|---|---|---|---|---|
| 1 | 48 | 49 | 2352 | 2304 | 2401 |
| 2 | 42 | 44 | 1848 | 1764 | 1936 |
| 3 | 46 | 48 | 2208 | 2116 | 2304 |
| 4 | 44 | 46 | 2024 | 1936 | 2116 |
| 5 | 46 | 46 | 2116 | 2116 | 2116 |
| 6 | 49 | 49 | 2401 | 2401 | 2401 |
| 7 | 43 | 44 | 1892 | 1849 | 1936 |
| 8 | 42 | 44 | 1848 | 1764 | 1936 |
| 9 | 46 | 46 | 2116 | 2116 | 2116 |
| 10 | 45 | 46 | 2070 | 2025 | 2116 |
| 11 | 43 | 44 | 1892 | 1849 | 1936 |
| 12 | 44 | 46 | 2024 | 1936 | 2116 |
| 13 | 46 | 46 | 2116 | 2116 | 2116 |
| 14 | 49 | 49 | 2401 | 2401 | 2401 |
| 15 | 48 | 49 | 2352 | 2304 | 2401 |
| 16 | 42 | 44 | 1848 | 1764 | 1936 |
| 17 | 46 | 48 | 2208 | 2116 | 2304 |
| 18 | 44 | 46 | 2024 | 1936 | 2116 |
| 19 | 46 | 46 | 2116 | 2116 | 2116 |
| 20 | 49 | 49 | 2401 | 2401 | 2401 |
| ∑ | 908 | 929 | 42257 | 41330 | 43221 |

From the table 15, n is the number of cases used in the pilot test (20 students), x is the score of the first testing, y is the score of the second testing (retest), xy is the multiplication of the two scores, $x^2$ is the square of the first score and $y^2$ is the square of the second score. These values were used in the test-retest formula as follows:

$$r_{xy} = \frac{n(\sum xy) - \sum x \sum y}{\sqrt{n(\sum x^2)-(\sum x)^2}\ \sqrt{n(\sum y^2)-(\sum y)^2}}$$

$$= \frac{20(42257)-(908)(929)}{\sqrt{20\ (41330\ )-(908)^2}\ \sqrt{20(43221)-(929)^2}}$$

$$= \frac{845140-843532}{\sqrt{826600-824464}\ \sqrt{864420-863041}}$$

$$= \frac{1608}{\sqrt{2136}\quad \sqrt{1379}}$$

$$= \frac{1608}{(46.22)(37.13)}$$

$$= \frac{1608}{1716.15}$$

$$= 0.94$$

$$r_{xy} = 0.94$$

It should be noted that the test- retest reliability served to establish the stability of the instrument in measuring what it was meant to measure. The calculated score of 0.94 therefore indicated that the instrument was highly stable thus very reliable in the pursuit of the study.

Coefficient of determination (Coefficient of common variance) is the square of the reliability coefficient. If therefore the reliability coefficient was 0.94, then the coefficient of determination ($r^2$) would be $(0.94)^2$ which gave 0.88. This value multiplied by 100 would give 88. This means 88% of the variances were common. That is 88% of the variances of the first administration of the test were enhanced by the variances of the second administration of the same test and vice versa.

**Administration; interview, observation and collection of data**

The direct delivery technique (DDT) was used in administering the questionnaires. After the validation of the instruments, the researcher took the authorisation to research from the Dean of the Faculty of Education of the University of Yaoundé 1 which he presented to the head of each of the schools from which the samples were drawn and in which the pilot test was conducted. He presented the authorisation and asked for permission to enter the school for the research exercise. The principals all signed a hand written application for authorisation to research in their school or endorsed it with the school stamp. In most cases, I was sent to meet the Vice principal who sent me to the head of department of Biology. They called a meeting with the teachers concerned and then met each teacher sampled and gave him the questionnaire after I had explained my mission verbally and the reasons for the research. He then gave some time for them to respond individually. He later went round to collect from each one. The teachers and the administrators of each of the schools helped in informing the students to accept me and collaborate in the carrying out of my research work. The exercise for the students was also done as each sampled student received the questionnaire and was asked to respond to the questions just like in an examination room. After that the students each gave back their already filled questionnaire directly to the teacher. A return rate of 100% was obtained for both the teachers and students.

With the aim of examining the effects of the various test structure analysis on creative student development, the researcher took a sample one thousand six hundred and three (1603) students, thirty-six (36) teachers drawn randomly from some schools in some six towns in six regions of the country: The Adamawa-Ngaoundere, the west and North West-Bafoussam/Bamenda, the Littoral/South

West-Douala/Buea and the Centre-Yaounde. These schools were selected using the stratified random sampling technique under lay private, denominational and government schools. Also, the schools were those which served as centres for the GCE and offered Biology in Both cycles. The sample was drawn from the form five and the upper sixth students of particular schools of these groups. The researcher designed questionnaires for the teachers and for the students.

After the validation of the questionnaires by my peer group members and supervisors, the instrument was taken for a pilot test to some 20 students in some two schools in Bamenda. Ten students each from the lay private secondary school: St. Michael Academy of Science and Arts (SMASA), Nkwen-Bamenda, and Government Bilingual High School Bayelle-Nkwen-Bamenda, for government school. These questionnaires were administered to both groups and the responses there of, were analyzed. These questionnaires were based on the research questions stated above, for the predictive or independent variables on one hand and on the criterion or dependent variable on the other hand. The researcher also made semi-structured interviews to the heads of departments, the school principals and others personnel. There was also an observation check list to find out the qualities of certain tests from the test banks where they existed and the conditions under which they existed. The researcher also observed the conditions under which tests were given as a participant observer in some cases and as a passive observer in others.

With the help of some class teachers and some discipline masters, the questionnaires were administered directly. That is the direct delivery method was used. The researcher with the aid of some school administrators or teachers arranged the learners two per desk, one at each end or three per desk, to avoid any collaboration or influence from each one of them. The modalities and purpose of the exercise were explained and each sampled case was given the questionnaire to

answer within a time frame. After the allocated time the questionnaires were collected by us, ensuring that all the sections and questions were answered. Those of the teachers were given to them to respond individually at their conveniences. The return rates of the questionnaires were 100% for both teachers and students.

**Method of Data collection and Analysis**

The data for this study was collected using three major data collecting instruments. These included the Questionnaire to students and the questionnaire to teachers. The semi-structured interview and the observational checklist were also used. The researcher went to the field to collect data by presenting the authorization from the dean of the faculty of education of the university of Yaounde 1 accompanied by a hand written application for permission to enter the school and do research. The school authority accepted by stamping the application and handing me to the person to take me round to the classes and the other staff members for recognition. The samples of students were selected and the questionnaire administered directly and then collected immediately each student finished. The checklist items were verified in the process of moving round the campus and directives asked during interviews for the things or aspects I could not directly observe. The interviews took place in the offices of the vice principals and principals, in the school science laboratory for the heads of departments and in the staff rooms for some teachers of biology. The analysis of the data was done as explained below:

The data collected using the questionnaires were analyzed in the following manner: using the hand, the researcher made a tally of the responses given by the teachers for each response, on one hand. On the other hand, the same thing was done to the responses given by the students. The sum of the similar responses got for each question was put together and the final score of responses for teachers and for students were put into the computer using the excel spread sheet and the

Special Package for Social Sciences (SPSS). Similarly, the indicators of the dependent variable also got by questionnaire were tallied and put also in the SPSS for analysis. The researcher used the Pearson product Moment correlation analysis to find out the correlation between the predictive or independent and the criterion or dependent variables. The outcome of the analysis was reported using the American Psychological Association (APA) style -7$^{th}$ edition.

The qualitative statistics was presented in the form of the means, medians and standard deviations calculated using the appropriate statistics in the SPSS.

Table 16

*Recapitulative table of hypotheses, variables, indicators and Modalities*

| Topic | General Hypothesis /General Question | Specific Questions | Specific Hypotheses | Variables | Indicators | Modalities |
|---|---|---|---|---|---|---|
| Test Structure Analysis and creative Student development in Biology in secondary schools in Cameroon | There is a significant relationship between test structure analysis and creative Student development in Biology | 1-How does test item analysis enhances creative student development? | Ha1-There is a significant relationship between test item analysis and creative Student development in learners. | IV: Test Item Analysis (TIA)<br>DV: Creative Student Development (CSD) | IVs-Item difficulty index<br>-Item discrimination<br>-response time<br>-comprehensibility<br>-responsiveness<br>DVs-curiosity<br>-experimentation<br>-collectiveness<br>-learnability<br>-communication | 1,2 =strongly disagree (SD)<br>3,4,5=Somewhat Disagree (SWD)<br>6,7,8 =Somewhat Agree (SWA)<br>9,10 = strongly Agree (SA) |
| | What relationship exist between test structure analysis and creative Student development in Biology? | 2-What relationship does the test item distractor analysis have with creative student Development? | Ha2- There is a significant relationship between test item distractor analysis and Creative Student development in learners | IV: Item Distracter Analysis (IDA)<br>DV: Creative Student Development | IVs-Number of items<br>-Total distractors<br>-Functional distractors<br>-Nonfunctional distractors<br>-cognitive level of item<br>DVs- seen above | |
| | | 3-How does test validity analysis influence creative student development? | Ha3- There is a significant relationship between test validity Analysis and creative Student development in learners. | IV: Test Validity Analysis (TVA)<br>DV: Creative Student Development | IVs-face validity<br>-content validity<br>-criterion validity<br>-construct validity<br>-validity coefficient<br>DVs-seen above | |
| | | 4-In what way does the test reliability analysis affect creative student Development? | Ha4- There is a significant relationship between test reliability analysis and creative Student development in learners. | IV: Test Reliability Analysis (TRA)<br>DV: Creative Student Development | IVs-reliability coefficient<br>-item revision<br>-Item discarding<br>-item retaining<br>-Item ordering<br>DVs-seen above | |
| | | 5-What relationship exists between test grading and creative student development? | Ha5- There is a significant relationship between test grading and creative Student development in learners. | IV: Test Grading Analysis (TGA)<br>DV: Creative Student Development | IVs-Test Type<br>-item scoring<br>-marking guide<br>-harmonization<br>-Deliberation<br>DVs-seen above | |

Table 16 is a synoptic table of all the hypotheses, variables and indicators of the study. It gives a vivid summary of the main parts of this study to permit easy comprehension. From the main topic of the thesis, the table shows the main alternative hypothesis and the main question of the study. The table shows all the five research questions and specific hypotheses. The various predictive or dependent variables operationalized are presented and the indicators of the responses and the modalities are all part of the table called the synoptic table. A good understanding of the table would give a vivid run down of this research work.

## Ethical Considerations

Ethical concerns are vital issues involved in research. The consent of each and every one to take part in a research constitutes a whole procedure without which vital information and reliable data may not be gathered. It is for this reason that this part of the research write-up comes in. Below is the procedure described as concerns the various levels in which consent was needed.

"Ethical guidelines serve as standards and a basis upon which each researcher ought to evaluate his or her own conduct, and the guidelines should be internalized in the personality of the researcher" De Vos, (2011, p. 57). "In educational research, ethics is concerned with ensuring that the interests and wellbeing of people are not harmed as a result of the research being done" Lankshear and knobel, (2004). It is in line with these that the researcher endeavoured to bear in mind several measures throughout the study.

## Consent Clearance

Before commencing this research, I got consent from the Dean of the Faculty of Education by writing a consent letter of application. The researcher personally went to all the schools to inform the principals and participating students and

teachers as well as school other available staff members about the nature and rationale of the study and how they would be involved.

**Voluntary Participation**

On entering the classroom and presented to the students, the researcher took over to explain his mission to the students who accepted to cooperate and answer the questionnaire. The reason for some not to answer the questionnaire was just because they did not fall in the sample due to the random sampling procedure. All those who were sampled voluntarily responded to the questionnaire and handed back to the researcher without compulsion.

Participants were informed to take part in this research out of their free will. They were also informed of their right or freedom to leave the research site if they in anyway felt uncomfortable to respond to the interviews or questionnaire. However, none of such happened while I was in the field.

**Harm to participants**

The administration of the questionnaire required just the pen or the pencil to give the right answers. Since it was a typical academic exercise explained to them, harm was far-fetched and they willfully took active part with no casualties.

Participants were assured that they would not be exposed to any danger or harm either physically or psychologically.

**Anonymity and confidentiality**

Anonymity and confidentiality were maintained throughout the course of the research. After the explanation of the aim of the research to the respondents, they were asked to be anonymous and write only their school and classes; information needed to identify the various respondents and during analysis. Many respondents

however thought that writing their names could have an effect but no name has been used so far in the analysis and no information was personalized as to denigrate someone or implicate someone. Letters of the alphabet were used to identify all participants. No real names were used when I was writing up the research. Also, interview transcripts and documents were kept confidential. No one else other than me, had access all the above documents. Participants were told in advance that the private data identifying participants would not be reported publicly.

**Feed back**

Many respondents asked about getting feedback about the research which they thought is very interesting and pertinent. The responses were that as an academic exercise, the procedures for feedback could only come after the publication of the thesis or during and after the defense.

All participants had the right to get feedback about the end of the research and conclusions arrived at either during the defense process or during seminars or from the library when published.

The main methods, techniques and procedures of this study have been presented in chapter three. This chapter serves the basis of what has been carried out by the researcher in his own expertise in bringing additional knowledge to what already has existed. Contrary to the works of chapter one and two based mainly on the laid down literature and the works of others, chapter three brings out the methods, techniques and the operational procedures followed to the realization of this thesis. The analysis of the data collected has been presented in chapter four and the detailed results also presented in both qualitative and quantitative research methods.

# CHAPTER FOUR

## PRESENTATION AND ANALYSIS OF FINDINGS

The relationship between test structure analysis and the creative student development was the main point of this study. The purpose was to find out if the analysis of test structure has a significant relationship with creative student development in Biology in secondary schools in Cameroon. The main topic was broken down into variables; the independent variable was "test structure analysis", while the dependent variable was "creative student development in Biology". The independent variable was operationalized into sub variables as follows: test item statistics (TIS), test item distractor analysis (IDA), test validity analysis (TVA), test reliability analysis (TRA) and test grading analysis (TGA). Five items per independent variable permitted the researcher to gather information about each from the teacher and student respondents. The dependent variable; creative student development in Biology also had five questions (items) that permitted me to measure the variable in the population of teachers and students.

The researcher used research objectives, research questions and research hypotheses to operationalize the topic. In this chapter however, the research hypotheses were used to present the results of the findings of the research work. The following research hypotheses were therefore used to carry out the study:

As a general research hypothesis; the test structure analysis has a significant relationship with the problem of creative student development in Biology in secondary schools in Cameroon.

The alternative hypotheses that follow, guided the study:

Ha1-Test item statistics to a significant extent relates with creative student development in Biology.

Ha2- There is a significant relationship between test item distractor analysis and Creative student development in Biology.

Ha3- To a significant level, test validity analysis relates with creative student development in Biology.

Ha4-Significantly, test reliability analysis relates with creative student development in Biology.

Ha5- There is a significant level of relationship between test grading analysis and creative student development in Biology.

The analyses presented in this chapter were performed using the data collected on the dates as follows:

Ngaoundere; February, 2021

Douala/Buea; February, 2021

Bafoussam/Bamenda; February, 2021

Yaounde; February, 2021

Four sub sections make up this chapter. The demographic information for the first sub section, the second sub section is the presentation of the responses to the research questions, the third sub section is based on the verification of the hypotheses and the fourth sub section is the summary of the findings based on the research questions and research hypotheses.

**Presentation of Demographic Information**

The demographic information here was presented following the two groups of respondents involved in the study; the students and the teachers who constituted the sample and the population in general. Demographic information serve as

extraneous variables that in one way or the other may infringe into the variables and cause them to vary according to their variations. The selected aspects of the demographic information for both the students and the teachers could be considered in the better understanding of the analysis and the interpretation of the results. This demographic information could also be considered as the main variables in further research on this topic and the correlations made to confirm or contrast the research outcomes.

**Demographic information on Students**

The demographic aspects of the students included their sex and average age of the respondents; their classes and the type of schools they attended also were considered in the presentations that follow:

Table 17

*Summary of the Characteristics of students used in the study*

| **Variable** | **Item Description** | **Representation** |
|---|---|---|
| Sex | Sex | Male |
| | | Female |
| Age | Age range | 15-20 |
| | | 21-25 |
| | | 26-30 |
| | | 31-35 |

Table 17 presents the sex, age and age range of the students who took part in the study. The representations were fully given to avoid the use of codes since the representations are short and direct in meaning. The age range of the students was

from 15 years to 35 years. This was to ensure that the youngest students and the oldest students be taken into consideration to avoid bias if they were present. However, these extreme cases were rare in the study.

Table 18

*Student participants' information in frequency and percentages*

| **Student Characteristics** | **Category** | **Frequency** | **Percentage (%)** |
|---|---|---|---|
| Sex | Male | 753 | 46.97 |
| | Female | 850 | 53.03 |
| | Total | 1603 | 100 |
| Age range | 15-20 | 1490 | 92.95 |
| | 21-25 | 113 | 07.05 |
| | 26-30 | 00 | 00 |
| | 31 -35 | 00 | 00 |
| | Total | 1603 | 100 |
| Class/series | Form 5 Science | 993 | 61.94 |
| | Upper Sixth Science | 610 | 38.06 |
| | Total | 1603 | 100 |
| School | Government | 689 | 42.98 |
| | Mission/ Denominational | 512 | 31.95 |
| | Lay Private | 402 | 25.07 |
| | Total | 1603 | 100 |

Table 18 presents the frequencies and percentages of each of the characteristics of the student participants. The age range, the sex, the class/series and the school are the described characteristics. Considering the sex of the students, the table shows that there were more female students than the male students who responded to the

questionnaire. 850 female students made up the sample and gave a percentage of 53.03, while 753 of the student respondents were males making a percentage of 46.97. 1490 students were within the age range of 15-20, making a percentage of 92.95. The rest of the sample; 113 were within the age range of 21-25, making a percentage of 07.05. There were no student respondents in the other age ranges.

Out of the 1603 student respondents, 993 of them were in form five, making a percentage of 51.94 while the remaining 610 making 38.06% were of the upper sixth class.

42.98% of the students (689) attended government institutions, 512 (31.95%) were drawn from mission/denominational schools and 402 students (25.07%) were drawn from lay private colleges.

Table 19

*Response rate of sections of Questionnaires for students*

| **Description** | **Test Item Statistics (TIS)** | **Item Distractor Analysis (IDA)** | **Test Validity Analysis (TVA)** | **Test reliability Analysis (TRA)** | **Test Grading Analysis (TGA)** |
|---|---|---|---|---|---|
| Valid N | 1603 | 1603 | 1603 | 1603 | 1603 |
| Missing | 0 | 0 | 0 | 0 | 0 |

Table 19 presents the total number of questionnaires that were answered and returned and the sections that were given total valid responses and those which had missing values. N represents the total number of respondents or sample. ‘Valid’

represents the number of questionnaire sections rightly responded to, while 'missing' represents the number of items or sections not responded to. All together it can be seen that the respondents all answered all the sections of the questionnaire and all gave back their answered questionnaire.

**Demographic information on teachers of Biology**

The aspects considered in the demographic information for teachers are more than those of the students. Sex, age range, highest academic qualification, longevity in teaching and the marking of the G.C.E., the class taught and the type of school are all those aspects that may influence the responses of the teachers and then affect the results of the research.

Table 20

*Summary of the Characteristics of teachers used in the study*

| **Variable Name** | **Item Description** | **Representation** |
|---|---|---|
| Sex | Sex | Male |
| | | Female |
| Age | Age range | 26-30 |
| | | 31-35 |
| | | 36-40 |
| | | 41-45 |
| | | 46-50 |
| | | 51-55 |
| | | 56-60 |

Table 20 presents the sex, age and age range of the teachers who took part in the study. The representations were fully given to avoid the use of codes since the representations are short and direct in meaning. The researcher decided to choose the age range from 26 years to 60years so that the young ones could not be left out of the sample and to avoid bias in the results. Sixty years is the year of retirement for civil servants and state workers, so some teachers could still be hanging around at that age.

Table 21

*Teacher participants' information in frequency and percentages*

| **Teachers Characteristics** | **Category** | **Frequency** | **Percentage** |
|---|---|---|---|
| Sex | Male | 17 | 47.2 % |
| | Female | 19 | 52.8 % |
| | Total | 36 | 100 % |
| Age range | 26-30 | 00 | 00% |
| | 31-35 | 07 | 19.4% |
| | 36-40 | 06 | 16.7% |
| | 41-45 | 03 | 8.3% |
| | 46-50 | 09 | 25% |
| | 51-55 | 09 | 25% |
| | 56-60 | 02 | 5.6% |
| | Total | 36 | 100% |
| Highest Academic Qualification | A/L | 14 | 38.8% |
| | First degree | 20 | 55.5% |
| | Master's | 02 | 5.6% |
| | Others | 00 | 00% |
| | Total | 36 | 100% |
| Class taught | Form 5 | 24 | 66.7% |
| | Upper sixth | 12 | 33.3% |
| | Total | 36 | 100% |
| Years of Teaching | 1-5 | 07 | 19.4% |
| | 6-10 | 09 | 25% |
| | 11-15 | 08 | 22.2% |
| | 16-20 | 05 | 13.9% |
| | 21-25 | 05 | 13.9% |
| | 26-30 | 02 | 5.6% |
| | 30 + | 00 | 00% |
| | Total | 36 | 100% |
| Years of GCE marking | Zero | 12 | 33.3% |
| | 1-5 | 06 | 16.7% |
| | 6-10 | 10 | 27.8% |
| | 11 and above | 08 | 22.2% |
| | Total | 36 | 100% |
| School | Gov't | 21 | 58.3% |
| | Mission | 09 | 25% |
| | Lay Private | 06 | 16.7% |
| | Total | 36 | 100% |

Table 21 presents four columns under the headings; teachers' characteristics, category, the frequencies and percentages of each of the characteristics of the teacher participants. The age range, the sex, the class/series taught, the highest academic qualification, years of teaching and the number of years of marking the GCE are the described characteristics which we need to understand the teacher sample used in the study.

Table 22

*Response rate of sections of Questionnaires for teachers*

| **Description** | **Test Item Analysis (TIA)** | **Item Distractor Analysis (IDA)** | **Test Validity Analysis (TVA)** | **Test Reliability Analysis (TRA)** | **Test Grading Analysis (TGA)** |
|---|---|---|---|---|---|
| Valid | 36 | 36 | 36 | 36 | 36 |
| N | | | | | |
| Missing | 0 | 0 | 0 | 0 | 0 |

Table 22 presents the total number of questionnaires that were answered and returned and the sections that were given total valid responses and those which had missing values. N represents the total number of respondents or sample. Valid represents the number of questionnaire sections rightly responded to, while missing represents the number of items or sections not responded to.

## Presentation of Descriptive Statistics and Inferential Statistics based on Hypotheses

For the purpose of coherence of facts and figures the descriptive and inferential statistics have been treated according to the various hypotheses, as gotten from the students then as gotten from the teacher respondents. The presentations began with the descriptive statistics where the semi-structured interview has been treated one item after the other and then the observational checklist.

## Presentation of Descriptive statistics

The descriptive statistics were presented by grouping the responses of the respondents into two or three groups according to the closeness of them. The responses were presented in the form of tables and bar charts and the conclusions or findings drawn thereof. There were seven questions or items on the interview guide as seen below.

Table 23

*Interview items and responses*

| Question number | Question | Grouped Responses |
|---|---|---|
| 1 | What theories do you or your colleague consider as basis for test or item development? | A- GCE board prescriptions<br>B-None<br>C- No laid down theory known |
| 2 | Where do you or your colleagues get items for a test? | A-Past question<br>B-Some textbooks<br>C- set test at any time |
| 3 | When developing a multiple-choice test item, what considerations do you make in selecting the distractors? | A- I don't understand<br>B- just propose any wrong answer |
| 4 | What procedures do you follow in validating a particular item or a test in general? | A- Colleague review<br>B- HOD review<br>C- Similar to GCE type |
| 5 | How do you consider the reliability of a test as a whole? | A- it is past GCE item<br>B- Found in the text book<br>C- I am a marker of the GCE |
| 6 | Could you please enlighten me about the processes involved in test grading in your department? | A- Marking guide on paper<br>B- Experienced marker<br>C- fixed rubric |
| 7 | Do you think item statistics, distractor analysis, test validity and reliability together with test grading analysis could have any correlation with creative student development in Biology? | A- Pedagogically correlated<br>B- teachers need to be enlightened<br>C- Need new ways to impact learners better |

Table 23 presents the semi-structured interview items. Numbers 1 up to 7 are the questions numbers indicating seven leading questions presented against each number as in the second column on the table. The third column presents the possible responses grouped into two as in question number 3 and into three groups as in the other questions. There were twelve respondents to each interview session

at each school the researcher visited. The respondents were either the head of the biology department or the vice principal in charge of the forms five or upper sixth biology classes. The frequencies of the detailed responses are presented below:

Table 24

*Frequency on question one responses*

**A, B, C**

| Variables | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 8 | 66.0 | 66.0 | 66.0 |
| | B | 1 | 8.0 | 8.0 | 75.0 |
| | C | 3 | 25.0 | 25.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 24 is the presentation of the frequency responses on question 1 of the semi-interview. From table 57, the question 1 was 'What theories do you or your colleague consider as basis for test or item development?' The answers were grouped into three groups labeled A, B, and C as follows:

A- GCE board prescriptions, B-None, C- No laid down theory known. As seen in table 57, 8 out of the 12 respondents said that they follow the laid down prescriptions of the G.C.E. board. This indicates that they have no guiding theories like the classical test theory or the item response theory on which they base their arguments in test or item development. The other two groups are similar in that they ascertain that they knew no theories in that light. This shows that all the teachers of the schools in the areas in which the research was carried do not consider any existing theories in the development of test items and the test as a whole. We can obviously conclude that there is need to carry out seminars and workshops in the areas to put teachers abreast with the theories of test development.

Figure 4

*Bar chart on question one*

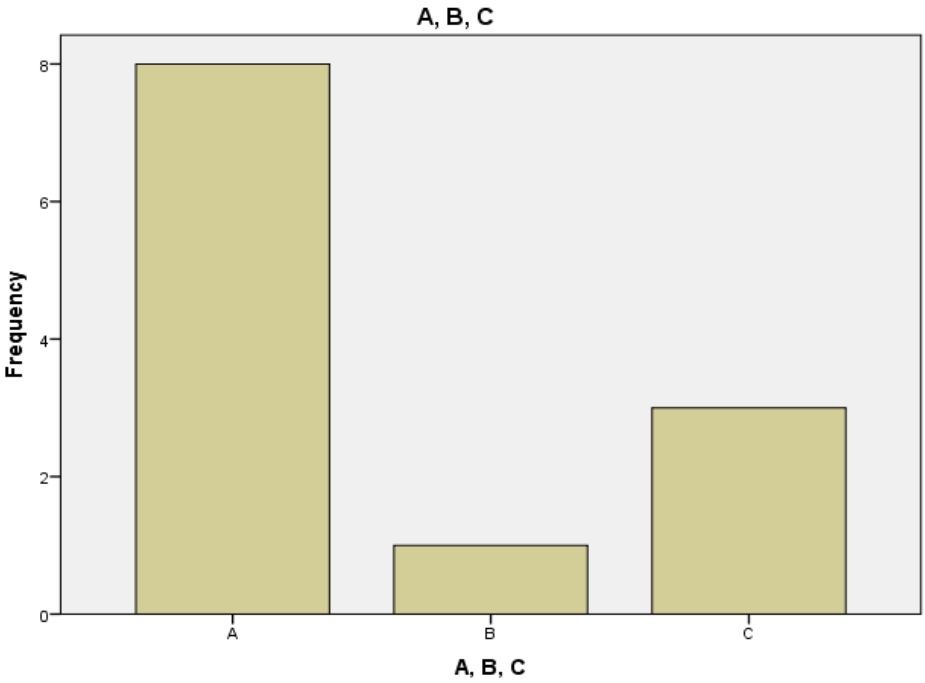



Figure 4 is a bar chart presenting the various responses from question 1. The highest bar 'A' is at 8, followed by 'C' with 3 and the lowest is 'B' with 1. These interpretations go in depth in the explanation of the results. As inferred from the table, most of the respondents' answers to the interview fell on the grouped response 'A'.

Table 25

*Frequency on question two responses*

**A, B, C**

| Variable | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 7 | 58.0 | 58.0 | 58.0 |
| | B | 3 | 25.0 | 25.0 | 83.0 |
| | C | 2 | 16.0 | 16.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 25 is the presentation of the frequency responses on question 2 of the semi-interview. From table 57, the question 2 was 'Where do you or your colleagues get items for a test?' The answers were grouped into three groups labeled A, B, and C as follows:

A-Past question, B-Some textbooks, C- set test at any time. As seen in table 59, 7 out of the 12 respondents said that they get test items from past question. This indicates that they have no guiding principles like those proposed by the classical test theory or the item response theory on which they base their arguments in test or item development. The other two groups are also different in that 3 out of the 12 respondents; i.e., 25% ascertain that they get questions for tests from the textbooks they own. Two respondents representing 16% of the sample declared that they set test at any time; this means that they are able to set tests at any time that a test is needed for a particular content and context. The responses show that all the teachers of the schools in the areas in which the research was carried do not own any test bank or item bank which they have built for a long time containing valid items or test for particular contents and or context. There is therefore every reason to conclude that there is the need for seminars and workshops in the areas of the research to put teachers abreast with the need to own, develop and maintain an item bank or test bank as we teach along the years.

Figure 5

*Bar chart on question two*

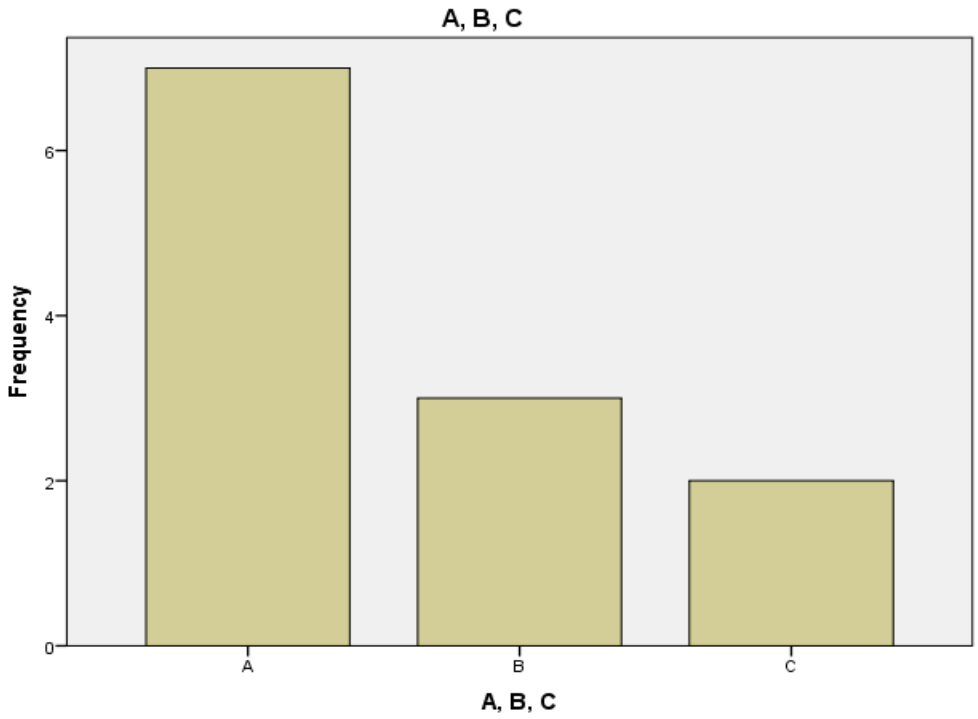



Figure 5 is a bar chart presenting the various responses from question 2. The highest bar 'A' is at 7, followed by bar 'B' with 3 and the lowest is bar 'C' with 2. That is A-Past question up to 7 represents 58%, B-Some textbooks with 3 responses represents 25% and C- set test at any time with 2 responses representing 16%. These interpretations go in depth in the explanation of the results obtained in the field.

Table 26

*Frequency on question three responses*

**A, B**

| Variables | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 7 | 59.0 | 59.0 | 59.0 |
| | B | 5 | 41.0 | 41.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 26 is the presentation of the frequency responses on question 3 of the semi-interview. From table 57, the question 3 was 'When developing a multiple-choice test item, what considerations do you make in selecting the distractors?' The answers were grouped into two groups labeled A and B as follows:

A- I don't understand, B- just proposes any wrong answer. As seen in table 60, 7 out of the 12 respondents, making 59%, said that they did not understand what the question required of them like answers. This indicates that they have no guiding principles like those proposed by the classical test theory or the item response theory on which they base their arguments in the selection of item distractors. The other group differed in that 5 out of the 12 respondents; i.e., 41% ascertained that they just proposed any wrong answer in biology that came across their minds. The responses show that the teachers' notions on distractor analysis need much to be desired so that the learners could get creative development in the domain. The teachers in the schools in the areas in which the research was carried therefore need to be schooled on how to select quality distractors in the midst of the right answers they propose such that those who know the right answer will find it obviously but those who do not know could not see it by chance. This is a way forward in the development of creative spirit in students of Biology.

Figure 6

*Bar chart on question three*

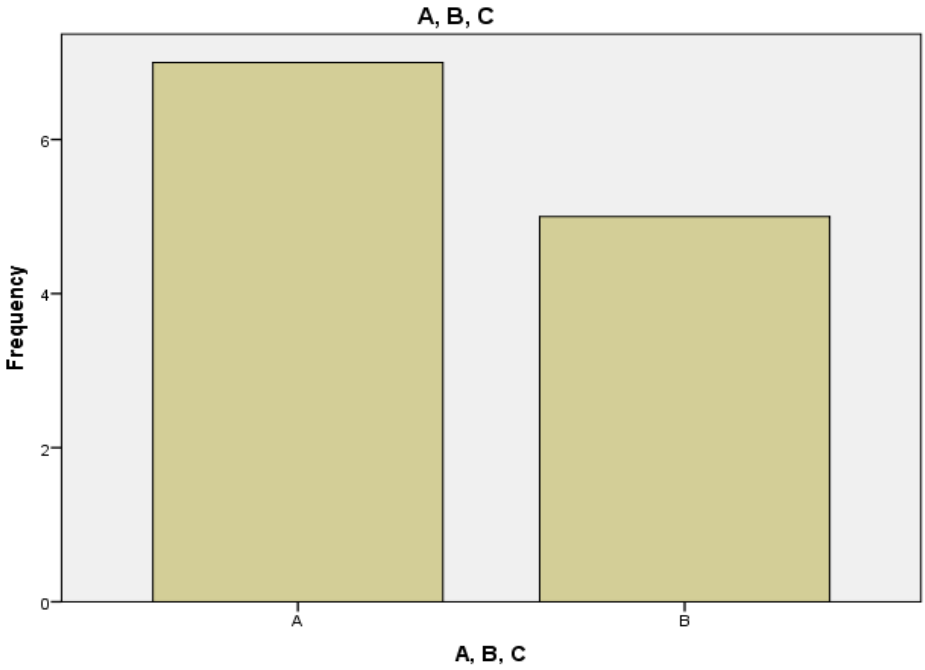



Figure 6 is a bar chart presenting the various responses from question 3. The highest bar 'A' is at 7, representing 59%. Bar 'B' is at 5, representing 41%. That is A- I don't understand represented by 7 respondents made up 59%, B- just proposes any wrong answer had 5 respondents representing 41%. These interpretations go in depth in the explanation of the results obtained in the field.

Table 27

*Frequency on question four responses*

**A, B, C**

| Variable | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 3 | 25.0 | 25.0 | 25.0 |
| | B | 5 | 42.0 | 42.0 | 67.0 |
| | C | 4 | 33.0 | 33.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 27 is the presentation of the frequency responses on question 4 of the semi- structured interview. From table 57, the question 4 was 'What procedures do you follow in validating a particular item or a test in general?' The answers were grouped into three groups labeled A, B, and C as follows:

A- Colleague review

B- HOD review

C- Similar to GCE type

As seen in table 23, 5 out of the 12 respondents (42%) said that their test items or test instruments are validated by the review of the head of department of Biology. This indicates that they have no guiding principles or procedures like those proposed by the classical test theory or the item response theory on which they base their arguments in test or item analysis or development. The other two groups are also different in that 4 out of the 12 respondents (C); i.e., 33% ascertain that they consider their tests validated because they are similar to the GCE type. Three respondents representing 25% of the sample (B) declared that they get their test items or instrument validated by the review of their colleagues; this means that test items or the instrument as a whole are treated for having validity without the proper follow up of the validation procedure by the teacher or test developer. The

responses show that all the teachers of the schools in the areas in which the research was carried do not follow the procedures in setting tests as prescribed by the renowned theories of classical test theory and item response theory. There is therefore every reason to conclude that there is the need for seminars and workshops in the areas of the research to put teachers abreast with the need to follow well laid down procedures of validating of items in a test and the test as a whole.

Figure 7

*Bar chart on question four*

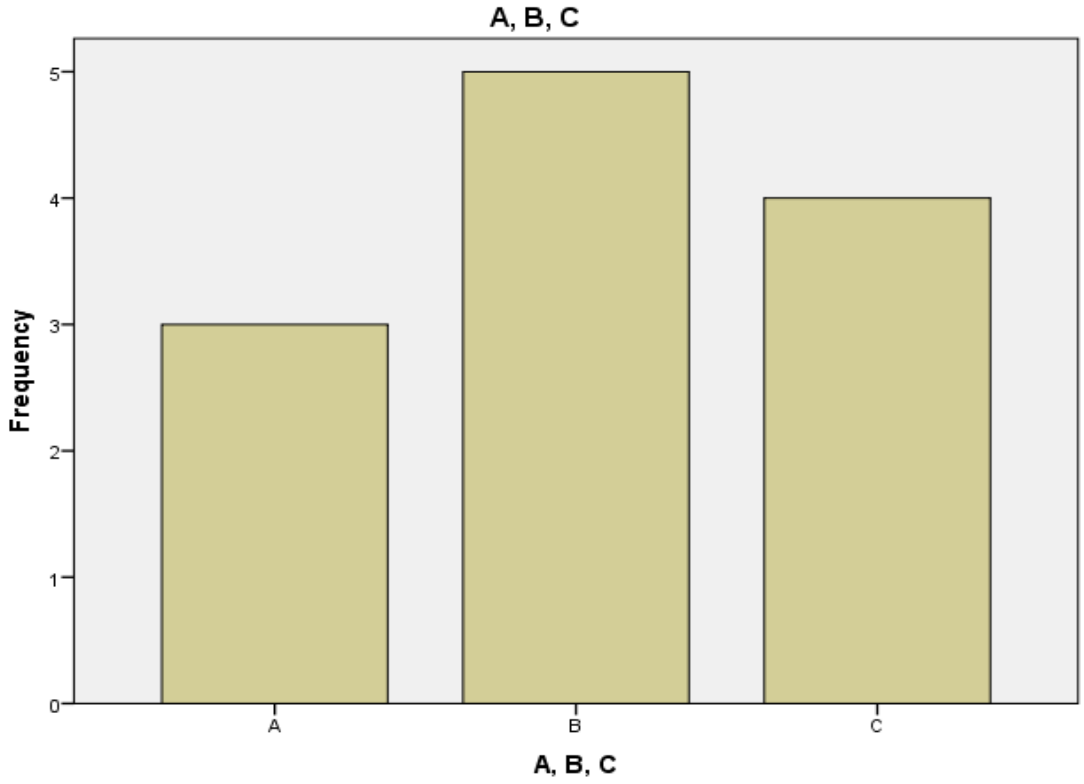



Figure 7 is a bar chart presenting the various responses from question 4. The highest bar 'B' is at 5, followed by bar 'C' with 4 and the lowest is bar 'A' with 3. That is A- Colleague review (25%), B- HOD review (42%), C- Similar to GCE type with 4 responses represents 33%. These interpretations go in depth in the

explanation of the results obtained in the field. While much trust is placed on the head of department and colleagues to assure the validity of the items and the instrument, every teacher should have a grasp of the procedure in validating a test item or instrument.

Table 28

*Frequency on question five responses*

**A, B, C**

| Variable | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 9 | 75.0 | 75.0 | 75.0 |
| | B | 1 | 9.0 | 9.0 | 84.0 |
| | C | 2 | 16.0 | 16.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 28 is the presentation of the frequency responses on question 5 of the semi-structured interview. From table 23, the question 5 was 'How do you consider the reliability of a test as a whole?' The answers were grouped into three groups labeled A, B, and C as follows:

A- It is past GCE item

B- Found in the text book

C- I am a marker of the GCE

As seen in table 28, 9 out of the 12 respondents (75%) said that their test instruments are considered to be reliable by the fact the test they use are or the test items involved are items of the past GCE Biology test. This indicates that they have no guiding principles or procedures like those proposed by the classical test theory or the item response theory on which they base their arguments in test or item analysis or development for a reliable test or test item. The other two groups are

also different in that 2 out of the 12 respondents (C); i.e., 16% ascertain that they consider their tests reliable because they are markers of the Biology tests in the GCE and master the indicators. One respondent representing 09% of the sample (B) declared that they get their test items or instrument assured in reliability by the fact that the textbooks are prescribed textbooks and so the items therein are thus reliable. This means that test items or the instrument as a whole are treated for having reliability without the proper follow up of the procedure by the teacher or test developer. The responses show that all the teachers of the schools in the areas in which the research was carried do not follow the procedures in setting tests as prescribed by the renowned theories of classical test theory and item response theory. There is therefore every reason to conclude that there is the need for seminars and workshops in the areas of the research to put teachers abreast with the need to follow well laid down procedures of assuring the reliability of items in a test and the test as a whole.

Figure 8

*Bar chart on interview question five*

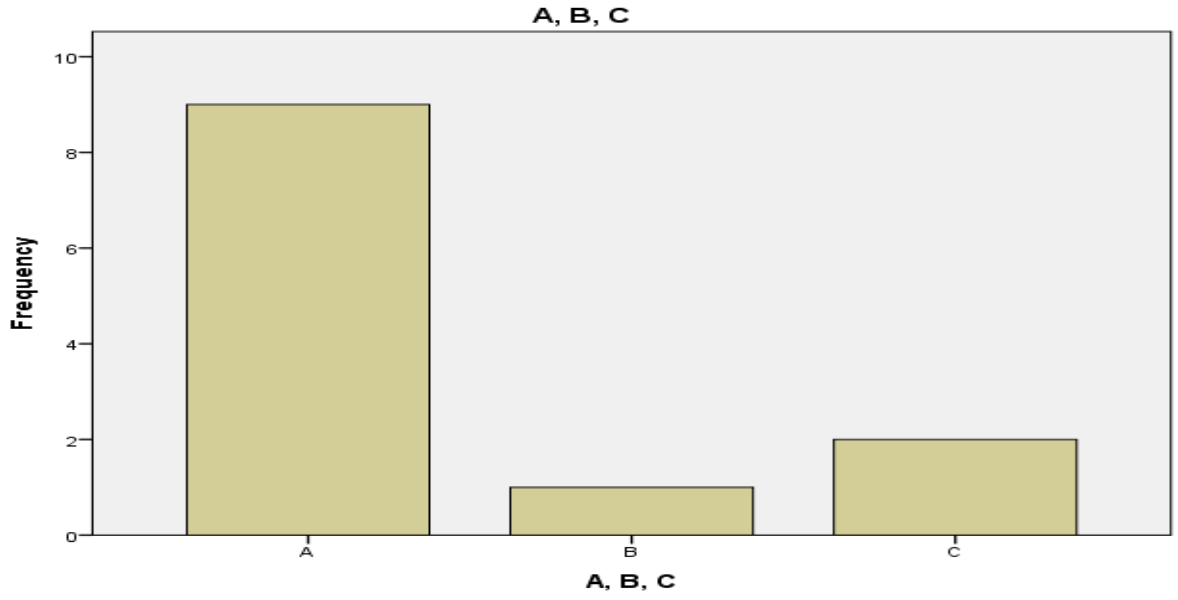


Figure 8 is a bar chart presenting the various responses from question 5. The highest bar 'A' is at 9, followed by bar 'C' with 2 and the lowest is bar 'B' with 1 respondent. That is A- It is past GCE item (75%), B- Found in the text book (9%) and C- I am a marker of the GCE (16%). These interpretations go in depth in the explanation of the results obtained in the field. While much trust is placed on past GCE questions and on textbooks to assure the reliability of the items and the instrument, every teacher should have a grasp of the procedure in assuring the validity of a test item or test instrument.

Table 29

*Frequency on question six responses*

**A, B, C**

| Variable | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 6 | 50.0 | 50.0 | 50.0 |
| | B | 4 | 34.0 | 34.0 | 84.0 |
| | C | 2 | 16.0 | 16.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 29 is the presentation of the frequency responses on question 6 of the semi-structured interview. From table 23, the question 6 was 'Could you please enlighten me about the processes involved in test grading in your department?' The answers were grouped into three groups labeled A, B, and C as follows:

A- Marking guide on paper

B- Experienced marker

C- Fixed rubric

As seen in table 29, 6 out of the 12 respondents (50%) said that the process involved in test grading involves the production of a marking guide on paper referring to it as they mark along. This indicates that they have the notions and

understand the guiding principles or procedures like those proposed by the classical test theory or the item response theory on which they base their arguments in test or item analysis or development for a good test grading. Just like the response on “C” (2 respondents with 16%) indicated for the need of a fixed rubric to follow in the grading of students in a test, the development of a marking guide if followed scrupulously would permit test grading to go smoothly and the students would be grateful and satisfied for the creative aspects of their development to be boosted. The remaining response “B” with 4 respondents making 34% presents that the respondents being experienced markers of the GCE need no marking guide or fixed rubric handy during the grading of students. This may be objective or not, but the students will have no base on which to verify the correctness of their responses and so the possibility of bias in the marking and grading processes. These will mare the development of creativity in learners as they may not be motivated to do more with such attitudes of the teacher. In this light there is need for some of the teachers to follow laid down grading processes for accurate scoring and the mitigation of serious errors.

Figure 9

*Bar chart on interview question six*

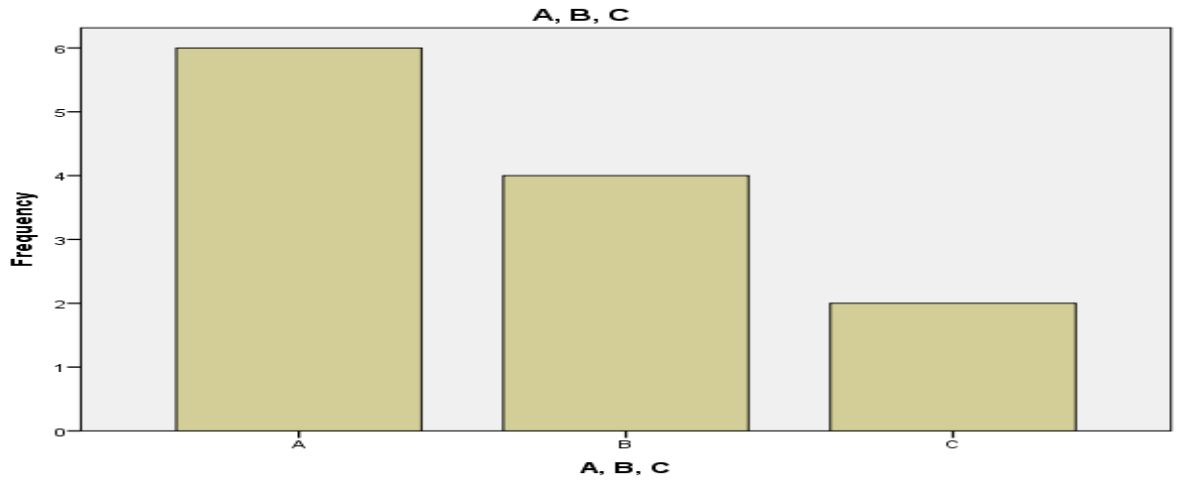


Figure 9 is a bar chart presenting the various responses from question 6. The highest bar ‘A’ is at 6, followed by bar ‘B’ at 4 and the lowest is bar ‘B’ at 2 as their frequency points. That is A- Marking guide on paper with 9 respondents (50%), B- Experienced marker with 4 respondents (34%) and C- Fixed rubric with 2 respondents (16%). These interpretations go in depth in the explanation of the results obtained in the field.

Table 30

*Frequency on question seven responses*

**A, B, C**

| Validity | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | A | 2 | 16.0 | 16.0 | 16.0 |
| | B | 7 | 58.0 | 58.0 | 74.0 |
| | C | 3 | 26.0 | 26.0 | 100.0 |
| | Total | 12 | 100.0 | 100.0 | |

Table 30 is the presentation of the frequency responses on question 7 of the semi-structured interview. From table 23, the question 7 was ‘Do you think item statistics, distractor analysis, test validity and reliability together with test grading analysis could have any correlation with creative student development in Biology?’ The answers were grouped into three groups labeled A, B, and C as follows:

A- Pedagogically correlated

B- Teachers need to be enlightened

C- Need new ways to impact learners better

As seen in table 30, 7 out of the 12 respondents in “B” (58%) said that teachers need to be enlightened about the independent variable mentioned in the interview item such that they could in turn apply the required exigencies of the domain. They

have the conviction that if well understood, the notion of test structure analysis with the various independent variables would well enhance creative student development if well impregnated and schooled on them. The other two groups are also different in that 3 out of the 12 respondents (C); i.e., 26% ascertain that they need new ways to better impact students. This falls in line with the response of 'B' above in their interpretations though stated in other words. Two respondents representing 16% of the sample (A) declared that the independent variables are pedagogically correlated with the dependent variable. This could be explained that with the implementation of the variables in pedagogy, there would be one causing the variance of the other and vice versa. The responses show that all the teachers of the schools in the areas in which the research was carried do acknowledged that the correlation between the independent variable and the dependent variable is a worthwhile correlation. There is therefore every reason to conclude that there is the need for seminars and workshops in the areas of the research to put teachers abreast with the well laid down procedures of assuring test structure analysis that will ensure creative student development in biology in secondary schools in Cameroon.

Figure 10

*Responses on interview question seven on a bar chart*

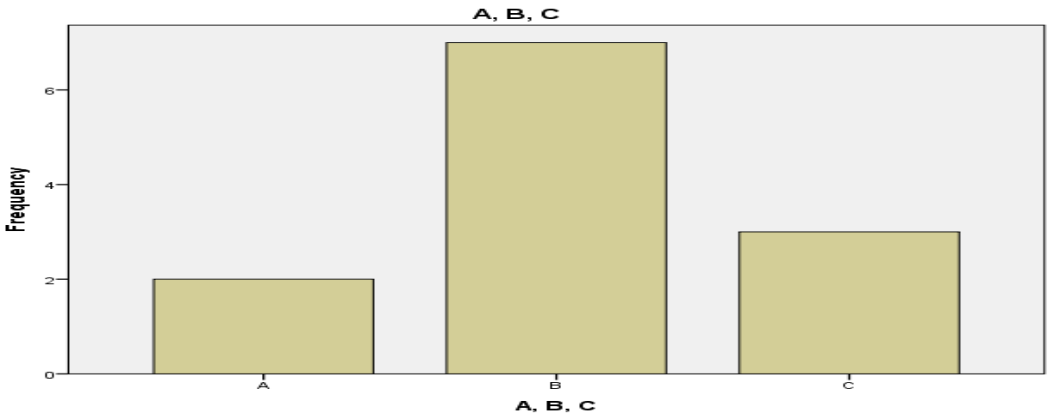


Figure 10 is a bar chart presenting the various responses from question 7. The highest bar 'B' is at 7, followed by bar 'C' at 3 and the lowest is bar 'A' at 2 as their frequency points. That is A- Pedagogically correlated 2 responses (16%), B- Teachers need to be enlightened 7 respondents (58%) and C- Need new ways to impact learners better; 3 respondents (26%). These interpretations go in depth in the explanation of the results obtained in the field.

Table 31

*Results from observations on the checklist*

| Researched elements | | Available | Not available |
|---|---|---|---|
| Item bank | -Personal | 3 | 9 |
| | -Departmental | 4 | 8 |
| | -All | 0 | 12 |
| Test bank | -Personal | 5 | 7 |
| | -Departmental | 7 | 5 |
| | - All | 3 | 9 |
| Test type | -Multiple choice | 12 | 0 |
| | -Essay | 12 | 0 |
| | - Structural | 9 | 3 |
| | - All | 9 | 3 |
| Test validity coefficient | -Face | 0 | 12 |
| | -Content | 0 | 12 |
| | -Criterion-related | 0 | 12 |
| | -Construct | 0 | 12 |
| | -Translation | 0 | 12 |
| | -Predictive | 0 | 12 |
| | -Concurrent | 0 | 12 |
| | -Convergent | 0 | 12 |
| | -Discriminant | 0 | 12 |
| Test reliability coefficient | -Test-retest | 0 | 12 |
| | -Parallel (equivalent form) | 0 | 12 |
| | -Inter-rater | 0 | 12 |
| | -Internal consistency | 0 | 12 |
| Test grading | -Marking guide | 12 | 0 |
| | -Score sheet | 12 | 0 |
| | -Marks per item | 12 | 0 |
| Practice site | -Laboratory | 9 | 3 |
| | -Farmland | 3 | 9 |
| | -Accessible land | 3 | 9 |
| School location | -Urban centre | 12 | 0 |
| | -Rural area | 0 | 12 |
| | -Fenced | 12 | 0 |
| | -Unfenced | 0 | 12 |

Table 31 is the checklist items used in collecting information about the availability of the items following the given rubrics as presented. The numbers indicate the number of schools that the various items were available or not available.

**Item bank**

Three individual classroom teachers presented some item banks in three different schools while four other schools indicated the presence of item banks belonging to the department. In closer check, the items bank for the departments were however still controlled by the actual head of department, indicating that in essence, he was the sole contributor of items to the item bank and could keep some items of other colleagues who could not do it or even had no idea of such. No individual teacher had an item bank in nine schools of the study. Eight of the schools did not show any departmental item bank. Generally, no school presented both the individual items bank and the departmental item bank. The notion of the keeping of item banks both by the teachers individually and collectively as a department should be spread out to biology teachers such that the setting of tests should be harmonious and with standards such as to promote creative student development.

**Test bank**

As concerns the test bank, five individual teachers from the twelve schools indicated that they had their personal test banks. Seven departments also presented some tests from their test banks but only three of the twelve schools presented both an individual item bank and the departmental item banks. It should therefore be the worry of the educational stakeholders to see how we could propagate the need of test banks in our secondary schools to enable easy access to valid and reliable test instruments and items that would promote creative student development in schools.

**Test types**

All of the twelve schools presented both essay type tests and the multiple choice test items. Nine of the schools also presented structural test items. Nine out of the twelve schools therefore, provided all the test types to the learners of biology. However, of those three schools which did not present the structural tests, there were structural test items in the essay tests. They thus, combined both the structural tests and the essay tests on the same paper. It is therefore clear that the schools in the study exploit all the types of tests offered in biology to enable the students to be creative in dealing with them since they are all part of the examination requirements in setting tests.

**Test validity and reliability coefficients**

Generally, no school presented any indicators of validity or reliability coefficients thus, 12 in all the column of unavailable. This indicates that of all the test papers checked, no coefficients of reliability or validity were tagged against them or found in a separate sheet. Indicating that the tests are not given the validity or reliability that they deserve for a fair score calculation for all students. This means that the results of a test at any time (t) may not give the same results after administration of the test to the same students and the same conditions. These results show the absence of validity and reliability of the instruments used in biology to test students in the schools involved in the study.

**Test Grading**

As indicated in the table 65 above, scores are allocated per item in the test in all the twelve schools. The teachers prepare marking guides that accompany the test such that any biology teacher could mark and score the candidates objectively and ensure similar grading of students in the scores sheets which they develop as they

mark the students' scripts. This practice in all the schools, ensure reliable scores void of bias. It is therefore a commendable effort for all the schools to ensure that test grading remains a vital and reliable part of building trust and creative thinking in learners of biology.

**Practice sites**

Nine out of the twelve schools of the study presented a reasonable laboratory for scientific experiments. Three other schools did not present any laboratory. The reverse holds for the farmland and accessible land. That is only three schools presented either a farmland or an accessible land on which experiments could be carried out. The availability of land for possible biological experiments is needed for schools to give access to students to carry out biological experiments at their zeal and convenience. This will give learning and testing the ecological validity that students must master for creative development both in and out of school.

**School location**

It was recorded that all the schools used in the study were located in urban centres and fenced. This serves the advantage of students learning in a serene confinement with little intruders into the schools. Nevertheless, some of the schools were so confined that no space of even growing a plant for experiment existed within the school, thus making the study of biology mostly of theory and classroom set-up that an environmental issue. These limit the development of creative students as their critical thinking could not be applied in real life within the school milieu.

**Presentation of Quantitative Data Collected from the Students**

Here, the data was presented following the number of hypotheses used in the study and the responses got from the items addressing the predictive or independent variable.

1- Ha1-Test item Analysis (TIA) to a significant extent relates with creative student development in Biology.

Table 32

*Responses on items addressing hypothesis one by the student respondents*

| STATEMENT | Strongly Disagree (SD) | | Somewhat Disagree (SWD) | | | Somewhat Agree (SWA) | | | Strongly Agree (SA) | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Test items are clear in the type of responses they need | 93 | 56 | 96 | 406 | 466 | 305 | 98 | 63 | 11 | 09 | 1603 |
| The language level and visual effect of items are without complaints in a test | 11 | 00 | 103 | 257 | 302 | 540 | 181 | 203 | 06 | 00 | 1603 |
| The items in the test are very different from the usual questions asked during class lessons | 17 | 09 | 93 | 400 | 472 | 299 | 63 | 98 | 96 | 56 | 1603 |
| The response time for each item is not indicated in the test paper | 03 | 08 | 100 | 260 | 300 | 542 | 172 | 212 | 06 | 00 | 1603 |
| The test items are easily comprehensible by all the test takers | 09 | 02 | 103 | 254 | 305 | 540 | 179 | 201 | 08 | 02 | 1603 |

From observation, table 32 reported that most of the respondents chose either SWD or SWA. This is seen from the scores of 4 and 6 where a huge chunk of the responses is found. It could therefore be read that test item analysis has not been fully carried out in most schools in Cameroon.

For statement number one, majority of the respondents SWD that test items were clear in the type of responses they needed. In statement two, most of the respondents ‘SWA’ that the language level and visual effects were without complaints in the tests. In statement three, majority responded that they SWD that items in the test were very different from the questions usually asked in class during lessons. Therefore, the test items set by most teachers were similar to those questions usually asked in class during lessons. As a reversed item, the reverse is true in that most test items were similar to those usually asked during lessons. Item four also is a reversed item. The response fell on the other side of the scale. Here, the majority SWA that the response time for each item is not indicated in the test paper. Thus, when reversed, they mean ‘they SWD’ that the response time is indicated in the test paper. Finally, in statement five, the majority of the respondents SWA that test items were easily comprehensible by all the test takers.

1- Ha2- Test Item Distractor Analysis (IDA)

Table 33

*The data on the response numbers per rubric on hypothesis two by student respondents.*

| STATEMENT | Strongly Disagree (SD) | | Somewhat Disagree (SWD) | | | Somewhat Agree (SWA) | | | Strongly Agree (SA) | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| The test items are easily understood by the test takers of the class in relation to the proposed responses/distractors | 93 | 156 | 196 | 406 | 366 | 205 | 98 | 72 | 11 | 00 | 1603 |
| The distractors are in the same range of understanding as the right response without outright divergence of concepts | 11 | 100 | 203 | 457 | 302 | 340 | 81 | 103 | 06 | 00 | 1603 |
| A fixed number of distractors is considered per item | 07 | 09 | 103 | 300 | 472 | 299 | 163 | 98 | 96 | 56 | 1603 |
| Item writing flaws are ignored when choosing an item for a test | 03 | 108 | 200 | 360 | 300 | 542 | 72 | 12 | 06 | 00 | 1603 |
| Functional and nonfunctional distractors are part of the responses | 09 | 102 | 103 | 154 | 405 | 440 | 279 | 101 | 08 | 02 | 1603 |

From observation, table 33 reported that most of the respondents chose SWD. This is seen from the scores of 4 and 5 where a huge chunk of the responses is found. It could therefore be read that test item distractor analysis has not been fully carried out in most schools in Cameroon.

For statement number one, majority of the respondents SWD that test items were easily understood in the type of distractors given in the responses. In statement two, most of the respondents SWD that the distractors were without outright divergence of concepts with the right answers in the tests. In statement

three, majority responded that they SWD that fixed number of distractors per item is considered in the test. Item four is a reversed item. The response fell on the other side of the scale. Here, the majority SWD that item writing flaws were ignored when choosing items for a test. Thus, when reversed, they mean that they ‘SWA’ that item writing flaws were not ignored when choosing items for a test. Finally, in statement five, the majority of the respondents ‘SWA’ indicated that the functional and nonfunctional responses were part of the responses in the test.

2- Ha3- Test Validity Analysis (TVA)

Table 34

*The data on the number of responses per item addressing Hypothesis three by student respondents*

| **STATEMENT** | **Strongly disagree (SD)** | | **Somewhat Disagree (SWD)** | | | **Somewhat agree (SWA)** | | | **Strongly agree (SA)** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Face validity: The presentation of the test items is appealing to the students' zeal to take the test. | 73 | 156 | 196 | 306 | 466 | 220 | 98 | 63 | 16 | 09 | 1603 |
| Content validity: The items address mainly what the students have been taught or are expected to be taught | 14 | 57 | 100 | 300 | 402 | 340 | 281 | 103 | 06 | 00 | 1603 |
| Criterion validity: The test follows laid down rules and criteria provided by the examination board or laid down format of a previously validated test | 10 | 16 | 83 | 410 | 473 | 199 | 163 | 97 | 97 | 55 | 1603 |
| Construct validity: The test items provoke logical reasoning and coherent outline of facts to be given | 53 | 58 | 200 | 260 | 100 | 542 | 172 | 110 | 08 | 100 | 1603 |
| Validity coefficient: A pre-test of the instrument is carried out from which the validity coefficient is calculated and tagged to the test pack. | 79 | 52 | 203 | 254 | 305 | 540 | 109 | 51 | 08 | 02 | 1603 |

Table 34 gives an indication that the responses centered on SWD in most of the responses. The individual items show similar trends that the respondents somewhat disagree to the items or statements posed.

3- Ha4- Test Reliability Analysis (TRA)

Table 35

*The number of responses per rubric addressing hypothesis four by student respondents*

| STATEMENT | Strongly Disagree (SD) | | Somewhat Disagree (SWD) | | | Somewhat Agree (SWA) | | | Strongly Agree (SA) | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Item revision: Test items are relooked upon before using them in the test | 94 | 60 | 91 | 406 | 470 | 301 | 100 | 61 | 20 | 00 | 1603 |
| Item discarding: rejected items are thrown away or not used in the test | 11 | 100 | 103 | 357 | 402 | 340 | 187 | 100 | 03 | 00 | 1603 |
| Item retention: Items are retained from a variety of options available or from an item bank | 20 | 09 | 190 | 470 | 402 | 199 | 53 | 108 | 93 | 59 | 1603 |
| Item ordering: The selected test items are arranged in a logical order as to ease comprehension and coherence in the flow of ideas | 53 | 58 | 142 | 260 | 450 | 350 | 172 | 112 | 06 | 00 | 1603 |
| Reliability coefficient: A calculated value of the reliability coefficient is tagged against each test pack | 09 | 102 | 203 | 254 | 405 | 340 | 179 | 101 | 08 | 02 | 1603 |

Table 35 gives an indication that the responses centered on SWD in most of the responses. The individual items show similar trends that the respondents somewhat disagree to the items or statements posed.

4- Ha5 – Test Grading Analysis (TGA)

*Table 36*

*The number of respondents per item on the hypothesis five by students*

| STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat SWA | | Agree | Strongly Agree SA | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Test type: Tests are graded according to the type of test involved | 93 | 56 | 96 | 406 | 366 | 105 | 198 | 103 | 71 | 109 | 1603 |
| Item scoring: Each item is assigned related score | 18 | 40 | 103 | 150 | 302 | 400 | 181 | 203 | 106 | 100 | 1603 |
| Marking guide: each item in a test is accompanied by a marking guide | 17 | 109 | 193 | 200 | 272 | 199 | 163 | 98 | 96 | 56 | 1603 |
| Harmonization: all test takers of the same level but different classes or teachers take the same test | 03 | 08 | 100 | 260 | 300 | 542 | 172 | 212 | 06 | 00 | 1603 |
| Deliberation: the results of a test are discussed in a board or committee and success/failure level fixed in common | 09 | 102 | 103 | 254 | 205 | 340 | 279 | 201 | 108 | 02 | 1603 |

Table 36 indicates that most of the student respondents chose that the responses that centered on SWD in most of their responses. The individual items show similar trends that the respondents somewhat disagree to the items or statements posed.

6 – The Criterion (dependent) variable: Creative Student development in Biology (CSD).

Table 37

*The responses of the students to each item on the dependent variable*

| STATEMENT | Strongly disagree (SD) | | Somewhat disagree (SWD) | | | Somewhat agree (SWA) | | | Strongly agree (SA) | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Test items arouse curiosity in the students who feel they are responding to practical issues | 93 | 56 | 96 | 406 | 466 | 305 | 98 | 63 | 11 | 09 | 1603 |
| Tests permit students to do experimentation of facts even after the test | 11 | 00 | 103 | 257 | 302 | 540 | 181 | 203 | 06 | 00 | 1603 |
| Tests encourage students to do collective work to solve real life problems | 17 | 09 | 93 | 400 | 472 | 299 | 63 | 98 | 96 | 56 | 1603 |
| Tests promote sustainable learning as they give room for verifiable/researchable responses | 03 | 08 | 100 | 260 | 300 | 542 | 172 | 212 | 06 | 00 | 1603 |
| Tests serve a means of communication between the student and teachers on what they have learnt and what they can do naturally | 09 | 02 | 103 | 254 | 305 | 540 | 179 | 201 | 08 | 02 | 1603 |

Table 37 gives an indication that the responses centered between SWD and SWA in most of the responses. The individual items show similar trends that the respondents somewhat disagree or somewhat agree to the items or statements posed. In this light, it indicates that not all the schools have similar practices. While one school practices the right thing, the other

practices may not be up to date as required by this study, reasons why we have the diverse responses on the table of creative student development.

**Presentation of quantitative data collected from teachers of Biology**

1 – Ha1- test item analysis for teacher respondents (TIA)

Table 38

*The number of teacher responses per item on the test item analysis variable*

| STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Item difficulty index is calculated before the item is validated | 00 | 00 | 02 | 03 | 11 | 05 | 09 | 03 | 03 | 00 | 36 |
| Item discrimination is ensured in the selection of test items | 00 | 00 | 04 | 07 | 09 | 06 | 04 | 03 | 01 | 02 | 36 |
| The mean score after item trial survey is considered in validating an item for a test | 00 | 01 | 03 | 09 | 12 | 05 | 03 | 02 | 01 | 00 | 36 |
| The response time for each item is not calculated before it is considered as part of a test | 03 | 07 | 05 | 05 | 07 | 08 | 01 | 00 | 00 | 00 | 36 |
| The comprehensibility of an item is ensured before using it in a test | 00 | 00 | 01 | 06 | 05 | 09 | 06 | 06 | 01 | 02 | 36 |

Table 38 gives an indication that the responses centered on SWD in most of the responses on items 1, 2 and 3, SWA on the last item five. The individual items show specific trends that the respondents somewhat disagree to the items or statements posed while the last statement indicates SWA. The

reversed item four indicated between disagree and somewhat disagree. Taken the other way, the respondents somewhat agree or agree to the item.

2 – Ha2 – Test item distractor analysis (IDA) for teacher respondents

Table 39

*The number of teachers per response on the item distractor analysis*

| **STATEMENT** | **Strongly Disagree (SD)** | | **Somewhat Disagree (SWD)** | | | **Somewhat Agree (SWA)** | | | **Strongly Agree (SA)** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| The cognitive level of the class is considered in choosing the distractors of responses | 00 | 00 | 00 | 05 | 09 | 07 | 06 | 06 | 03 | 00 | 36 |
| The distractor efficiency is considered in each item during its selection | 00 | 00 | 05 | 06 | 09 | 05 | 05 | 04 | 02 | 00 | 36 |
| A fixed number of distractors is considered per item | 00 | 00 | 02 | 09 | 12 | 07 | 03 | 03 | 00 | 00 | 36 |
| Item writing flaws are ignored when choosing an item for a test | 03 | 07 | 05 | 05 | 07 | 08 | 01 | 0 | 00 | 00 | 36 |
| Functional and nonfunctional distractors are part of the responses | 00 | 00 | 01 | 06 | 05 | 09 | 06 | 6 | 01 | 02 | 36 |

Table 39 indicated that SWD and SWA are the areas many teachers of the sample indicated that test items distractors are located. This indicates that many schools take into consideration the aspect of distractors in the development if test items in their tests. Thus, there is a correlation of the two variables.

3 – Ha3 – Test Validity Analysis (TVA) for teachers

Table 40

*Responses of teachers on the test validity analysis variable*

| **STATEMENT** | **Strongly Disagree SD** | | **Somewhat Disagree SWD** | | | **Somewhat Agree SWA** | | | **Strongly Agree SA** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Face validity: The presentation of the test items is appealing to the students' zeal to take the test. | 00 | 00 | 00 | 05 | 10 | 10 | 05 | 04 | 01 | 01 | 36 |
| Content validity: The items address mainly what the students have been taught or are expected to be taught | 00 | 00 | 01 | 07 | 09 | 09 | 04 | 03 | 01 | 02 | 36 |
| Criterion validity: The test follows laid down rules and criteria provided by the examination board or laid down format of a previously validated test | 00 | 00 | 03 | 06 | 12 | 05 | 04 | 05 | 01 | 00 | 36 |
| Construct validity: The test items provoke logical reasoning and coherent outline of facts to be given | 00 | 01 | 02 | 03 | 10 | 08 | 07 | 03 | 02 | 00 | 36 |
| Validity coefficient: A pre-test of the instrument is carried out from which the validity coefficient is calculated and tagged to the test pack. | 00 | 02 | 02 | 09 | 07 | 09 | 06 | 01 | 00 | 00 | 36 |

Table 40 gives an indication that the responses centered between SWD and SWA in most of the responses. The individual items show similar trends that the respondents somewhat disagree or somewhat agree to the items or statements posed. In this light, it indicates that not all the schools have similar practices. While one school practices the right thing, the other

practices may not be up to date as required by this study, reasons why the diverse responses on the table of test validity analysis.

4 – Ha4 – Test Reliability Analysis (TRA) for teachers

*Table 41*

*The responses of teacher on the test reliability analysis variable*

| **STATEMENT** | **Strongly Disagree (SD)** | | **Somewhat Disagree (SWD)** | | | **Somewhat Agree (SWA)** | | | **Strongly Agree (SA)** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Item revision: Test items are relooked upon before using them in the test | 00 | 00 | 04 | 06 | 11 | 05 | 09 | 01 | 00 | 00 | 36 |
| Item discarding: rejected items are thrown away or not used in the test | 00 | 00 | 04 | 07 | 09 | 09 | 04 | 03 | 00 | 00 | 36 |
| Item retention: Items are retained from a variety of options available or from an item bank | 00 | 02 | 05 | 09 | 12 | 05 | 03 | 00 | 00 | 00 | 36 |
| Item ordering: The selected test items are arranged in a logical order as to ease comprehension and coherence in the flow of ideas | 00 | 02 | 05 | 07 | 07 | 08 | 04 | 02 | 01 | 00 | 36 |
| Reliability coefficient: A calculated value of the reliability coefficient is tagged against each test pack | 00 | 00 | 03 | 08 | 08 | 09 | 06 | 02 | 0 | 00 | 36 |

Majority of the teachers ‘SWD’ that these elements addressing the reliability analysis are fully put in place by them. Though some accepted that these concerns are being looked into, majority disagrees. Test reliability analysis would therefore be needed in the test structure analysis for creative student development to be ensured.

5 – Ha5 – Test Grading Analysis (TGA) for teachers

Table 42

*The responses of teachers on the test grading analysis variable*

| STATEMENT | Strongly disagree (SD) | | Somewhat disagree (SWD) | | | Somewhat agree (SWA) | | | Strongly agree (SA) | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Test type: Tests are graded according to the type of test involved | 00 | 00 | 00 | 03 | 11 | 05 | 09 | 05 | 03 | 00 | 36 |
| Item scoring: Each item is assigned related score | 00 | 00 | 00 | 07 | 09 | 08 | 06 | 03 | 01 | 02 | 36 |
| Marking guide: each item in a test is accompanied by a marking guide | 00 | 01 | 03 | 09 | 12 | 05 | 03 | 02 | 01 | 00 | 36 |
| Harmonization: all test takers of the same level but different classes or teachers take the same test | 00 | 00 | 05 | 07 | 07 | 08 | 06 | 03 | 00 | 00 | 36 |
| Deliberation: the results of a test are discussed in a board or committee and success/failure level fixed in common | 00 | 00 | 00 | 07 | 05 | 09 | 06 | 07 | 01 | 01 | 36 |

From the table 42, majority of the teachers SWD that they include the grading analysis prescribed by the study into the analysis of their test structures. Others however SWA but they are not as many as those for SWD. Nevertheless, it was noted that many who set exams and mark them tend to practice the grading system more than those who do not.

6 – The criterion (dependent) variable: Creative Student Development (CSD) in Biology for teachers

Table 43

*The number of responses on the dependent variable from the teacher respondents*

| STATEMENT | Strongly disagree SD | | Somewhat Disagree SWD | | | Somewhat agree SWA | | | Strongly agree SA | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Test items arouse curiosity in the students who feel they are responding to practical issues | 00 | 00 | 02 | 03 | 11 | 05 | 09 | 03 | 03 | 00 | 36 |
| Tests permit students to do experimentation of facts even after the test | 00 | 00 | 04 | 07 | 09 | 06 | 04 | 03 | 01 | 02 | 36 |
| Tests encourage students to do collective work to solve real life problems | 00 | 01 | 03 | 09 | 12 | 05 | 03 | 02 | 01 | 00 | 36 |
| Tests do not promote sustainable learning as they do not give room for verifiable/researchable responses | 03 | 07 | 05 | 05 | 07 | 08 | 01 | 00 | 00 | 00 | 36 |
| Tests serve a means of communication between the student and teachers on what they have learnt and what they can do naturally | 00 | 00 | 01 | 06 | 05 | 09 | 06 | 06 | 01 | 02 | 36 |

Table 43 presents the responses got from the sample of teachers per item analysing the dependent variable. The responses could be presented in the form of a frequency table and the descriptions carried out to bring out the mode median, standard deviation and the percentages. This has not been done in all the tables because the study is typically a psychometric study and emphasis has been places

on the measurement aspects dealing directly with numbers of quantities that could be given a validity and a reliability accompanied by other parametric variables.

## Verification of Hypotheses

### Responses from Students

Table 44

*The means and standard deviations of TIA and CSD from student respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TIA | 1603 | 8 | 34 | 20.32 | 7.183 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 44 shows the descriptive statistics of the data on Test Items Analysis (TIA) and Creative Student Development (CSD) as collected from student respondents. For the TIS, the lowest score from the students was 08 and the highest was 34. The mean score was 20.32 indicating that most of the scores centered around it. The standard deviation was 7.18. The first standard deviation for the TIS would therefore be -27.5 and +27.5, while that of CSD would be -37.39 and +37.39. We can deduce that most of the respondents on TIS had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve.

That of the CSD also indicates a normal distribution but with outliers like 47 found within the scores, making the mean to be that high. The total number of

respondents is indicated as 1603 listwise to both variables. The mean falling around 20.32 for TIS indicates that from the questionnaire, most respondents took a stand and ‘somewhat disagree” about the items posed on Test Item Statistics. On the other hand, the mean of CSD falling on 26.66 also fall in the same line like ‘somewhat disagree’ with some outliers making it move to the first column of the ‘somewhat agree’. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students’ development.

Table 45

*The Correlation between TIA and CSD from the student respondents*

**Correlations**

| | | TIA | CSD |
|---|---|---|---|
| TIA | Pearson Correlation | 1 | .689** |
| | Sig. (2-tailed) | | .000 |
| | N | 1603 | 1603 |
| CSD | Pearson Correlation | .689** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 45 presents the results of the Pearson correlation between Test Item Statistics and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the

correlation gave the value 0.689 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.689 = r, to have $r^2$, we have $(0.689)^2 = 0.47$. Therefore, $r^2 = 0.47$. By multiplying this value by 100, that is 0.47 x 100 = 47. We have 47. The explanation given by this value is that 47% of the variances that occur on the independent variable are caused by the dependent variable, and 47% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 47% of the variances on Test Item Statistics are caused by the Creative Student Development and 47% of the variances on Creative Student Development are caused by the Test Item Statistics.

By this calculation of the coefficient of correlation we can conveniently conclude that Test Item Statistics to a significant extent relates with Creative Student Development in Biology in some secondary schools in Cameroon, thus, confirming the first hypothesis.

Table 46

*The means and standard deviations of IDA and CSD from student respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| IDA | 1603 | 8 | 45 | 21.64 | 8.275 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 46 shows the descriptive statistics of the data on Item Distractor Analysis (IDA) and Creative Student Development (CSD) as collected from student respondents. For the IDA, the lowest score from the students was 08 and

the highest was 45. The mean score was 21.64 indicating that most of the scores centered around it. The standard deviation was 8.28. The first standard deviation for the IDA would therefore be -29.92 and +29.92, while that of CSD would be -37.39 and +37.39. We can deduce that most of the respondents on IDA had their total scores centred around the mean and within the first standard deviations of +1 and -1 with outliers like 45 as the maximum score.

This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve. That of the CSD also indicates a normal distribution but with outliers like 47 found within the scores, making the mean to be that high. The total number of respondents is indicated as 1603 list wise to both variables. The mean falling around 21.64 for IDA indicates that from the questionnaire, most respondents took a stand and 'somewhat disagree" about the items posed on Item Distractor Analysis. On the other hand the mean of CSD falling on 26.66 also fall in the same line like 'somewhat disagree' with some outliers making it move to the first column of the 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students' development.

Table 47

*Correlation between IDA and CSD from student respondents*

**Correlations**

| Variable | | IDA | CSD |
|---|---|---|---|
| IDA | Pearson Correlation | 1 | .471** |
| | Sig. (2-tailed) | | .000 |
| | N | 1603 | 1603 |
| CSD | Pearson Correlation | .471** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 47 presents the results of the Pearson correlation between Item Distractor Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.471 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.471 = r, to have $r^2$, we have $(0.471)^2 = 0.22$. Therefore, $r^2 = 0.22$. By multiplying this value by 100, that is 0.22 x 100 = 22. We have 22. The explanation given by this value is that 22% of the variances that occur on the independent variable are caused by the dependent variable, and 22% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 22% of the variances on Item Distractor Analysis are caused by the Creative Student Development and 22% of the

variances on Creative Student Development are caused by the Item Distractor Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that there is a significant relationship between test item distractor analysis and Creative student development in Biology in some secondary schools in Cameroon, thus, confirming the second hypothesis.

Table 48

*The means and standard deviations of TVA and CSD from student respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TVA | 1603 | 9 | 47 | 23.92 | 11.610 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 48 shows the descriptive statistics of the data on Test Validity Analysis (TVA) and Creative Student Development (CSD) as collected from student respondents. For the TVA, the lowest score from the students was 09 and the highest was 47. The mean score was 23.92 indicating that most of the scores centered on it. The standard deviation was 11.6. The first standard deviation for the TVA would therefore be -35.52 and +35.52, while that of CSD would be -37.39 and +37.39. We can deduce that most of the respondents on TVA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve with great outliers like 47. That of the CSD also indicates a normal distribution with outliers like 47 found within the scores also, making the mean to be that high. The total number of respondents is indicated

as 1603 list wise to both variables. The mean falling around 23.92 for TVA indicates that from the questionnaire, most respondents took a stand and 'somewhat disagree" about the items posed on Test Validity Analysis. On the other hand, the mean of CSD falling on 26.66 also fall in the same line like 'somewhat disagree' with some outliers making it move to the first column of the 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students' development. This suggested that TVA was not done enough to enable creative student development in Biology in secondary schools in Cameroon.

Table 49

*Correlation between the TVA and the CSD from the student respondents*

**Correlations**

| Variables | | TVA | CSD |
|---|---|---|---|
| TVA | Pearson Correlation | 1 | .751** |
| | Sig. (2-tailed) | | .000 |
| | N | 1603 | 1603 |
| CSD | Pearson Correlation | .751** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 49 presents the results of the Pearson correlation between Test Validity Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the

correlation gave the value 0.751 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.751 = r, to have $r^2$, we have $(0.751)^2 = 0.56$. Therefore, $r^2 = 0.56$. By multiplying this value by 100, that is 0.56 x 100 = 56. We have 56. The explanation given by this value is that 56% of the variances that occur on the independent variable are caused by the dependent variable, and 56% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 56% of the variances on Test Validity Analysis are caused by the Creative Student Development and 56% of the variances on Creative Student Development are caused by the Test Validity Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that to a significant level, Test Validity Statistics relate with Creative Student Development in Biology in some secondary schools in Cameroon, thus, confirming the third hypothesis.

Table 50

*The means and standard deviations of TRA and CSD from student respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TRA | 1603 | 8 | 45 | 25.54 | 9.410 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 50 shows the descriptive statistics of the data on Test Reliability Analysis (TRA) and Creative Student Development (CSD) as collected from

student respondents. For the TRA, the lowest score from the students was 08 and the highest was 45. The mean score was 25.54 indicating that most of the scores centered on it. The standard deviation was 9.41. The first standard deviation for the TRA would therefore be -34.95 and +34.95, while that of CSD would be -37.39 and +37.39. We can deduce that most of the respondents on TRA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve with some few large scores or outliers out of the majority scores. That of the CSD also indicates a normal distribution with outliers like 47 found within the scores, making the mean to be that high. The total number of respondents is indicated as 1603 list wise to both variables. The mean falling around 25.54 for TRA indicates that from the questionnaire, most respondents took a stand and 'somewhat disagree" about the items posed on Test Validity Analysis. On the other hand, the mean of CSD falling on 26.66 also fall in the same line like 'somewhat disagree' with some outliers making it move to the first column of the 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students' development, meaning that much of TRA was not done in test structure analysis such that creative student development could not easily be enhanced.

Table 51

*Correlation between the TRA and CSD from the student respondents*

**Correlations**

| Variable | | TRA | CSD |
|---|---|---|---|
| TRA | Pearson Correlation | 1 | .646** |
| | Sig. (2-tailed) | | .000 |
| | N | 1603 | 1603 |
| CSD | Pearson Correlation | .646** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 51 presents the results of the Pearson correlation between Test Reliability Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.646 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.646 = r, to have $r^2$, we have $(0.646)^2 = 0.42$. Therefore, $r^2 = 0.42$. By multiplying this value by 100, that is 0.42 x 100 = 42. We have 42. The explanation given by this value is that 42% of the variances that occur on the independent variable are caused by the dependent variable, and 42% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 42% of the variances on Test Reliability Analysis are caused by the Creative Student Development and 42% of the

variances on Creative Student Development are caused by the Test Reliability Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that significantly, test reliability coefficient relates with creative student development in Biology in some secondary schools in Cameroon, thus, confirming the fourth hypothesis.

Table 52

*The means and standard deviations of TGA and CSD from student respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TGA | 1603 | 8 | 47 | 25.84 | 11.444 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 52 shows the descriptive statistics of the data on Test Grading Analysis (TGA) and Creative Student Development (CSD) as collected from student respondents. For the TGA, the lowest score from the students was 08 and the highest was 47. The mean score was 25.84 indicating that most of the scores centered on it. The standard deviation was 11.44. The first standard deviation for the TGA would therefore be -37.28 and +37.28, while that of CSD would be -37.39 and +37.39. We can deduce that most of the respondents on TGA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve with few outliers or large scores. That of the CSD also indicates a normal distribution with outliers like 47 found within the scores, making the mean to be that high. The total number of respondents is

indicated as 1603 list wise to both variables. The mean falling around 25.84 for TGA indicates that from the questionnaire, most respondents took a stand and 'somewhat disagree" about the items posed on Test Grading Analysis. On the other hand, the mean of CSD falling on 26.66 also fall in the same line like 'somewhat disagree' with some outliers making it move to the first column of the 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students' development.

Table 53

*Correlation between TGA and CSD from the student respondents*

**Correlations**

| Variable | | TGA | CSD |
|---|---|---|---|
| TGA | Pearson Correlation | 1 | .683** |
| | Sig. (2-tailed) | | .000 |
| | N | 1603 | 1603 |
| CSD | Pearson Correlation | .683** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 53 presents the results of the Pearson correlation between Test Grading Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.683 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.683 = r, to have $r^2$, we have $(0.683)^2 = 0.47$. Therefore, $r^2 = 0.47$. By multiplying this value by 100, that is 0.47 x 100 = 47. We have 47. The explanation given by this value is that 47% of the variances that occur on the independent variable are caused by the dependent variable, and 47% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 47% of the variances on Test Grading Analysis are caused by the Creative Student Development and 47% of the variances on Creative Student Development are caused by the Test Grading Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that there is a significant level of relationship between Test Grading Analysis and Creative Student Development in Biology in some secondary schools in Cameroon, thus, confirming the fifth hypothesis.

**Responses from teachers**

Table 54

*The means and standard deviations of TIA and CSD from teacher respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TIA | 36 | 10 | 47 | 32.47 | 11.290 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 54 shows the descriptive statistics of the data on Test Items Analysis (TIA) and Creative Student Development (CSD) as collected from teacher respondents. For the TIS, the lowest score from the teachers was 10 and the highest

was 47. The mean score was 32.47 indicating that most of the scores centered on it. The standard deviation was 11.29. The first standard deviation for the TIS would therefore be -43.76 and +43.76, while that of CSD would be -46.93 and +46.93 two tailed. We can deduce that most of the teacher respondents on TIS had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve. That of the CSD also indicates a normal distribution but with outliers like 47 found within the scores, making the mean to be that high. The total number of respondents is indicated as 36 listwise to both variables. The mean falling around 32.47 for TIS indicates that from the questionnaire, most respondents took a stand and 'somewhat agree" about the items posed on Test Item Statistics. On the other hand, the mean of CSD falling on 35.08 also fall in the same line like 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat agree about the item indicators in relation to creative students' development.

Table 55

*Correlation between TIA and CSD from the teacher respondent*

**Correlations**

| Variable | | TIA | CSD |
|---|---|---|---|
| TIA | Pearson Correlation | 1 | .824** |
| | Sig. (2-tailed) | | .000 |
| | N | 36 | 36 |
| CSD | Pearson Correlation | .824** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 55 presents the results of the Pearson correlation between Test Item Analysis (TIA) and Creative Student Development (CSD); the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.824 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.824 = r, to have $r^2$, we have $(0.824)^2 = 0.679$. Therefore, $r^2 = 0.68$. By multiplying this value by 100, that is 0.68 x 100 = 68. We have 68. The explanation given by this value is that 68% of the variances that occur on the independent variable are caused by the dependent variable, and 68% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 68% of the variances on Test Item Statistics are caused by the Creative Student Development and 68% of the variances on Creative Student Development are caused by the Test Item Statistics.

By this calculation of the coefficient of correlation we can conveniently conclude that Test Item Statistics to a significant extent relates with Creative Student Development in Biology in secondary schools in Cameroon, thus, confirming the first hypothesis.

Table 56

*The means and standard deviations of IDA and CSD from teacher respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| IDA | 36 | 10 | 38 | 27.58 | 8.623 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 56 shows the descriptive statistics of the data on Item Distractor Analysis (IDA) and Creative Student Development (CSD) as collected from teacher respondents. For the IDA, the lowest score from the teachers was 10 and the highest was 38. The mean score was 27.58 indicating that most of the scores centered on it. The standard deviation was 8.623. The first standard deviation for the IDA would therefore be -36.20 and +36.20, while that of CSD would be -46.26 and +46.26. We can deduce that most of the respondents on IDA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve. That of the CSD also indicates a normal distribution. The total number of respondents is indicated as 36 list wise to both variables. The mean falling around 27.58 for IDA indicates that from the questionnaire, most respondents took a stand and ‘somewhat agree” about the items posed on Item Distractor Analysis. On the other hand, the mean of CSD falling on 35.08 also fall in the same line like ‘somewhat agree’. This indicates that the respondents somewhat agree about the item indicators in relation to creative students’ development.

Table 57

*Correlation between IDA and CSD from the teacher respondents*

**Correlations**

| Variable | | IDA | CSD |
|---|---|---|---|
| IDA | Pearson Correlation | 1 | .727** |
| | Sig. (2-tailed) | | .000 |
| | N | 36 | 36 |
| CSD | Pearson Correlation | .727** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 57 presents the results of the Pearson correlation between Item Distractor Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.727 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.727 = r, to have $r^2$, we have $(0.727)^2 = 0.53$. Therefore, $r^2 = 0.53$. By multiplying this value by 100, that is 0.53 x 100 = 53. We have 53. The explanation given by this value is that 53% of the variances that occur on the independent variable are caused by the dependent variable, and 53% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 53% of the variances on Item Distractor Analysis are caused by the Creative Student Development and 53% of the

variances on Creative Student Development are caused by the Item Distractor Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that there is a significant relationship between test item distractor analysis and Creative student development in Biology in secondary schools in Cameroon, thus, confirming the second hypothesis.

Table 58

*The means and standard deviations of TVA and CSD from teacher respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TVA | 36 | 11 | 43 | 29.81 | 9.130 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 58 shows the descriptive statistics of the data on Test Validity Analysis (TVA) and Creative Student Development (CSD) as collected from teacher respondents. For the TVA, the lowest score from the teachers was 11 and the highest was 43. The mean score was 29.81 indicating that most of the scores centered on it. The standard deviation was 9.13. The first standard deviation for the TVA would therefore be -38.93 and +38.93, while that of CSD would be -46.27 and +46.27. We can deduce that most of the respondents on TVA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve. That of the CSD also indicates a normal distribution. The total number of respondents is indicated as 36 listwise to both variables. The mean falling around 29.81 for TVA indicates that from the

questionnaire, most respondents took a stand and ‘somewhat agree” about the items posed on Test Validity Analysis. On the other hand, the mean of CSD falling on 35.08 also fall in the same line like ‘somewhat agree’. This indicates that the respondents somewhat agree about the item indicators in relation to creative students’ development.

Table 59

*Correlation between TVA and CSD from the teacher respondents*

**Correlations**

| | | TVA | CSD |
|---|---|---|---|
| TVA | Pearson Correlation | 1 | .949** |
| | Sig. (2-tailed) | | .000 |
| | N | 36 | 36 |
| CSD | Pearson Correlation | .949** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 59 presents the results of the Pearson correlation between Test Validity Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.949 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.949 = r, to have $r^2$, we have $(0.949)^2 = 0.90$. Therefore, $r^2 = 0.90$. By multiplying this value by 100, that is 0.90 x 100 = 90. We have 90. The explanation given by this value is that 90%

of the variances that occur on the independent variable are caused by the dependent variable, and 90% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 90% of the variances on Test Validity Analysis are caused by the Creative Student Development and 90% of the variances on Creative Student Development are caused by the Test Validity Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that to a significant level, Test Validity Statistics relate with Creative Student Development in Biology in secondary schools in Cameroon, thus, confirming the third hypothesis.

Table 60

*The means and standard deviations of TRA and CSD from teacher respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TRA | 36 | 10 | 44 | 25.92 | 13.334 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 60 shows the descriptive statistics of the data on Test Reliability Analysis (TRA) and Creative Student Development (CSD) as collected from teacher respondents. For the TRA, the lowest score from the teachers was 10 and the highest was 44. The mean score was 25.92 indicating that most of the scores centered on it. The standard deviation was 13.33. The first standard deviation for the TRA would therefore be -39.25 and +39.25, while that of CSD would be -46.27 and +46.27. We can deduce that most of the respondents on TRA had their total

scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve. That of the CSD also indicates a normal distribution. The total number of respondents is indicated as 36 teachers listwise to both variables. The mean falling around 25.92 for TRA indicates that from the questionnaire, most respondents took a stand and 'somewhat disagree" about the items posed on Test Reliability Analysis. On the other hand, the mean of CSD falling on 35.08 falls in the same line like 'somewhat agree' with some outliers making it move to the first column of the 'somewhat agree'. Nevertheless, this indicates that the respondents somewhat disagree about the item indicators in relation to creative students' development.

Table 61

*Correlation between TRA and CSD from the teacher respondents*

**Correlations**

| Variable | | TRA | CSD |
|---|---|---|---|
| TRA | Pearson Correlation | 1 | .760** |
| | Sig. (2-tailed) | | .000 |
| | N | 36 | 36 |
| CSD | Pearson Correlation | .760** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 61 presents the results of the Pearson correlation between Test Reliability Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 36 teacher

respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.760 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.760 = r, to have $r^2$, we have $(0.760)^2 = 0.58$. Therefore, $r^2 = 0.58$. By multiplying this value by 100, that is 0.58 x 100 = 58. We have 58. The explanation given by this value is that 58% of the variances that occur on the independent variable are caused by the dependent variable, and 58% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 58% of the variances on Test Reliability Analysis are caused by the Creative Student Development and 58% of the variances on Creative Student Development are caused by the Test Reliability Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that significantly, test reliability analysis relates with creative student development in Biology in secondary schools in Cameroon, thus, confirming the fourth hypothesis.

Table 62

*The means and standard deviations of TGA and CSD from teacher respondents*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TGA | 36 | 10 | 50 | 33.61 | 13.970 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 62 shows the descriptive statistics of the data on Test Grading Analysis (TGA) and Creative Student Development (CSD) as collected from teacher

respondents. For the TGA, the lowest score from the teachers was 10 and the highest was 50. The mean score was 33.61 indicating that most of the scores centered on it. The standard deviation was 13.97. The first standard deviation for the TGA would therefore be -47.58 and +47.58, while that of CSD would be -46.27 and +46.27. We can deduce that most of the respondents on TGA had their total scores centred around the mean and within the first standard deviations of +1 and -1. This interpretation indicates that the scores were distributed normally or in a normal bell-shaped distribution curve with few outliers or large scores. That of the CSD also indicates a normal distribution with outliers like 46 found within the scores, making the mean to be that high. The total number of respondents is indicated as 36 teachers list wise to both variables. The mean falling around 33.61 for TGA indicates that from the questionnaire, most respondents took a stand and 'somewhat agree" about the items posed on Test Grading Analysis. On the other hand the mean of CSD falling on 35.08 also fall in the same line like 'somewhat agree'. This indicates that the respondents somewhat agree about the item indicators in relation to creative students' development.

Table 63

*Correlation between TGA and CSD from the teacher respondents*

**Correlations**

| Variable | | TGA | CSD |
|---|---|---|---|
| TGA | Pearson Correlation | 1 | .913** |
| | Sig. (2-tailed) | | .000 |
| | N | 36 | 36 |
| CSD | Pearson Correlation | .913** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 63 presents the results of the Pearson correlation between Test Grading Analysis and Creative Student Development; the independent variable and the dependent variable respectively. This was carried out using the Statistical Package for Social Sciences (SPSS) version 20. For a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.913 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed).

Using the Coefficient of Determination denoted $r^2$, correlation can better be explained. In this case, given the correlation coefficient of 0.913 = r, to have $r^2$, we have $(0.913)^2 = 0.83$. Therefore, $r^2 = 0.83$. By multiplying this value by 100, that is 0.83 x 100 = 83. We have 83. The explanation given by this value is that 83% of the variances that occur on the independent variable are caused by the dependent variable, and 83% of the variances that occur at the dependent variable are caused by the independent variable. In this case, 83% of the variances on Test Grading Analysis are caused by the Creative Student Development and 83% of the variances on Creative Student Development are caused by the Test Grading Analysis.

By this calculation of the coefficient of correlation we can conveniently conclude that there is a significant level of relationship between Test Grading Analysis and Creative Student Development in Biology in secondary schools in Cameroon, thus, confirming the fifth hypothesis.

**Summary of findings got from students**

This section on Summary of findings presents the basic elements of the finding got from the respondents. On the basis of these summaries of the findings, the main conclusion can be drawn. This section highlights summary tables presenting the various correlations as presented per variable and hypothesis. I

began with the summary presentation as gotten from the student respondents and followed by the summary of presentation got from teachers.

Generally, the concept of motivation plays a very great role in the way learners go about creative student development with regards to the analysis of the test. Naturally, in familiarity with the fauna and flora, and considering the different validities and reliabilities, test instruments could be so motivated that the student either intrinsically or extrinsically gets motivated to creative student development. In this light the following aspects of motivation could boost creative student development.

Motivation is a concept commonly applied in the classroom. The teacher uses motivation to easily put the learner to work either on his own by his willingness or to work for a purpose that is attached to an external factor. In this light we talk of intrinsic and extrinsic motivation. In this study, we expected that with the considerations placed on the motivational factors, tests developed should be analyzed so as to motivate the learner to be creative at all level of it. The tests should not be robotic or routine, but analyzed and structured in such a way that as the learners take the test they are motivated by the results they expect after the test or that they find themselves solving particular problems plaguing their society or themselves as they take the test.

Different authors define motivation differently. Koontz and O'Donnell, (1980, p.334) for example define motivation as "to act in a desired manner". Pinder (1984 p. 24) on his part considers motivation as "the force within a person that affects his or her direction, intensity and persistence of voluntary behavior which could be covert or overt". The main ambiguity with motivation is the way to influence a class of students, each one with a distinctive set of needs and a unique personality to work together with the aim of achieving creative skills. The ability to understand

why students behave the way they do and the ability to motivate them to behave in a specified desired manner through a test are two inter related qualities which are very important for creativity development.

Luthans, (1998 p. 8) positions that motivation is the process which directs, energizes, arouses and sustains behaviours and performances of learners. That is, it is the process of stimulating students to action, being committed towards the achievement of a desired objective; one way to go about this is to implement appropriate technique of motivation that will make students more satisfied with and committed to their studies.

The priority concern of test developers and teachers is the making of the students to take part in the achievement of testing objectives, but they should be aware that if the students are unhappy with the teachers and test developers; there would be a very timid rate of creative student development in the classroom or school. For this reason, Hekina & Jones, (1967) viewed that; students should be considered and valued as assets for the development of creative and innovative resources.

The beginning of the theory of Maslow is the question: 'What motivates behaviour?' Maslow wrote that, our actions as human beings are motivated in order to achieve certain needs. Maslow first introduced his concept of a hierarchy of needs and motivation in his 1943 article "A Theory of Human Motivation" and his subsequent book in 1954, "Motivation and Personality."

Maslow brought to light a new section of interest in the study of psychology. When schools of thought and theories of psychology were overwhelmed by psychoanalysis and behaviourism, psychology focused deeply on problematic behaviours. Maslow on the other hand was interested more in learning and

understanding what motivates students or people as a whole. He also had much interest in the comprehension of what individuals do in order to acquire that which gave them happiness.

In the light of a humanist psychologist, Maslow had in mind that individuals have an internal zeal to be self-actualized: that is, the zeal to be everything they can possibly become. For Maslow however, in a bid to acquire these ultimate interest of what individuals really want to become, a number of more basic needs must be met such as the need for food, safety, love, and self-esteem, among others as applied to this study.

While this theory is generally known as the hierarchy of needs, Maslow never displayed his theory as a pyramid Eaton, (2012). Presented in the form of a pyramid however, the lowest levels are made up of the most 'basic needs', while the more 'complex needs' are located at the top of the pyramid.

Table 64

*The means and standard deviations of all the IVs and CSD from student*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TIA | 1603 | 8 | 34 | 20.32 | 7.183 |
| IDA | 1603 | 8 | 45 | 21.64 | 8.275 |
| TVA | 1603 | 9 | 47 | 23.92 | 11.610 |
| TRA | 1603 | 8 | 45 | 25.54 | 9.410 |
| TGA | 1603 | 8 | 47 | 25.84 | 11.444 |
| CSD | 1603 | 9 | 47 | 26.66 | 10.727 |
| Valid N (listwise) | 1603 | | | | |

Table 64 shows all the minimum and maximum scores, means and standard deviations got from the student respondents following all the variables of the study.

It is a recapitulation of all the descriptive statistics presented on one table. The total number of the student respondents is indicated in the column as N.

Table 65

*All the correlations between IVs and the dependent variable from the students*

**Correlations**

| Variable | | TIA | IDA | TVA | TRA | TGA | CSD |
|---|---|---|---|---|---|---|---|
| TIA | Pearson Correlation | 1 | .559** | .655** | .633** | .650** | .689** |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .000 | .000 |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |
| IDA | Pearson Correlation | .559** | 1 | .485** | .541** | .562** | .471** |
| | Sig. (2-tailed) | .000 | | .000 | .000 | .000 | .000 |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |
| TVA | Pearson Correlation | .655** | .485** | 1 | .755** | .702** | .751** |
| | Sig. (2-tailed) | .000 | .000 | | .000 | .000 | .000 |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |
| TRA | Pearson Correlation | .633** | .541** | .755** | 1 | .757** | .646** |
| | Sig. (2-tailed) | .000 | .000 | .000 | | .000 | .000 |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |
| TGA | Pearson Correlation | .650** | .562** | .702** | .757** | 1 | .683** |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | | .000 |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |
| CSD | Pearson Correlation | .689** | .471** | .751** | .646** | .683** | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | .000 | |
| | N | 1603 | 1603 | 1603 | 1603 | 1603 | 1603 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 65 shows the correlation table of all the five independent variables correlated with themselves and with the other variables including the dependent

variable as collected from the one thousand six hundred and three (1603) student respondents in the study.

The correlation of each variable with itself gives the value of 1 as seen on the table. For example, the correlation of Test Item Analysis (TIA) on the first column and the TIS on the first row, the Pearson correlation is 1, the sig. (2-tailed) signifies that the level of significance is two-tailed; non directional. All of the significances of the correlations occur at the 0.000 point and the all are therefore significant at the 0.01level of the two-tailed.

The correlation of one variable with the other variables of the study indicates that all the correlations are highly correlated with one another with possibly very high inter correlation values. This depicts that the combination of these variables in a test structure would highly vary the creativity development of learners of Biology in secondary schools in Cameroon as concerns the responses from the students considered in the study.

**Summary of findings got from teachers**

Motivation involves a means of creating great level of enthusiasm to attain testing purposes and this situation is accommodated by satisfying some individual need. Basically, motivation refers to achieving the main goals of assessment by satisfying individual student's needs, demands or expectations.

Holding on to Pinder (1998) cited in Ambrose & Kulik, (1999), teacher motivation may be viewed as a collection of internal and external forces which trigger work-related behaviour and determine its duration, direction, form and intensity. The concept is in relation to the teaching context particularly and includes the influence on work behaviour of both environmental forces, and those inherent in the person. In the school, motivation of work is seen as a personal, invisible and

hypothetical structure which show off itself in the way of observable (overt) and therefore measurable behaviours.

The improvement of productivity is a great hurdle to surmount that considers the engagement of the attention of proprietors or employers in ither private or government by strategizing appropriate mechanism for motivating their teachers. The seriousness of this challenge can be understood from the management's perception of the strong functional correlation between teacher's motivation and school performance. For years, proprietors of schools have been making experiment with diverse techniques in a bid to effectively relate the motivation of teachers with their teaching/learning inputs.

Ouchi, (1987) has given an explanation on the importance of motivation in relation to performance. He studied the reasons why the national productivity of Japan is doing well. He moved ahead to say that the reason behind the Japanese success is the much interest the Japanese managers present for the well-being and development of workers. It is not a new discovery; the fact that motivation is the most important factor for productivity and quality. The studies carried out by Elton Mayo within the 1930 presented that it has been earmarked for the first time. From that time, several studies have been confirmed that the same findings were gotten in several industries including the one of software development. Ouchi, (1987).

Motivation is very necessary for the principals and teachers to get hold of, and understand why teachers have different behaviours at school and the way to go about their behaviour such that they put in their utmost efforts to acquire good academic performance. It is the aim of principals at each stage to have teachers motivated in order that work can improve at desired time, rate and pace.

When teachers have the feeling that what they input like commitment, loyalty, efforts, enthusiasm and trust to the school are satisfactorily and correctly given reward by the outputs like financial, material and other benefits or incentives, they remain motivated and keep on providing inputs towards higher productivity. Teachers feel de-motivated when they foresee the absence of that equity. Ahmad, Mukaddes, Rashed, & Samad, (2010).

Three early motivation theories have this far served as foundations to the contemporary motivation theory: The hierarchy of needs of Maslow, the motivation & Hygiene theory of Herzberg and expectancy theory of Victor Vroom. Although the development of more valid explanations of motivation have been made, these early theories are important due to the fact that they represent the foundation from which contemporary motivation theories were developed and because many practicing school heads still use them.

The hierarchy of needs by Maslow should be among the best-known motivation theories worldwide. Maslow is a good beginning to commence the examination of the different motivation theories. Maslow started with the notion that individuals always intend to expect something and what they expect depends on what they have already. Maslow propounded that there exist five divergent levels of needs that individuals have to seek for the satisfaction of their basic needs. Mullins, (2007, p. 258). The first and lowest level is Physiological needs. Before these needs that comprise for example water, food, clothing and shelter are acquired, an individual does not have any other needs. In the case where persons do not have the feeling of cold, thirst or hunger all their needs go to the second level. The next level after the lowest level is Security needs. In that level an individual requires to feel secured in their family and in the society, and also feel protected against violence. The requirement for safety is presented with job security, savings

and for insurances of health, mental health, old age and disability. The third level of Maslow's hierarchy is love and belonging needs. After the feeling of security, individuals need to feel that they receive and give love, they are appreciated and they have good friendships. Esteem needs is the fourth level. It consists of the need to be uniquely special with personal-respect and to enjoy esteem from other persons. Many Persons expect to do self-evaluation highly based on their achievement and to receive appreciation from other people. The absence of these needs may lead to the feeling of weakness, helplessness and inferiority. Self-actualization is the highest level of Maslow's hierarchy of needs. The development of this need depends on the satisfaction at the other four lower levels. It refers to the need of self-fulfillment and to the tendency to become actualized in what an individual has as potential.

The centre of this theory dwells in the fact that when one need is fulfilled its strength diminishes and the strength of the next level increases. Latham, (2007, p. 31). Maslow however noted that one level of needs does not have to be fulfilled totally before an individual would move to the next higher level. The person can be satisfied partly with one level and still seek for satisfaction at the next level. Salanova & Kirmanen, (2010).

Table 66

*The means and standard deviations of all the IVs and CSD from teachers*

**Descriptive Statistics**

| Variable | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| TIA | 36 | 10 | 47 | 32.47 | 11.290 |
| IDA | 36 | 10 | 38 | 27.58 | 8.623 |
| TVA | 36 | 11 | 43 | 29.81 | 9.130 |
| TRA | 36 | 10 | 44 | 25.92 | 13.334 |
| TGA | 36 | 10 | 50 | 33.61 | 13.970 |
| CSD | 36 | 11 | 46 | 35.08 | 11.185 |
| Valid N (listwise) | 36 | | | | |

Table 66 shows a recapitulation of all the descriptive statistics as per variable; both independent and the dependent variables. N indicates the sample or the number of respondents. The minimum and maximum values of all the variables are presented including the means and the standard deviations.

Table 67

*All the correlations between IVs and the DV from the teacher respondents*

**Correlations**

| Variable | | TIA | IDA | TVA | TRA | TGA | CSD |
|---|---|---|---|---|---|---|---|
| TIA | Pearson Correlation | 1 | .579** | .768** | .530** | .849** | .824** |
| | Sig. (2-tailed) | | .000 | .000 | .001 | .000 | .000 |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |
| IDA | Pearson Correlation | .579** | 1 | .610** | .503** | .695** | .727** |
| | Sig. (2-tailed) | .000 | | .000 | .002 | .000 | .000 |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |
| TVA | Pearson Correlation | .768** | .610** | 1 | .896** | .920** | .949** |
| | Sig. (2-tailed) | .000 | .000 | | .000 | .000 | .000 |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |
| TRA | Pearson Correlation | .530** | .503** | .896** | 1 | .856** | .760** |
| | Sig. (2-tailed) | .001 | .002 | .000 | | .000 | .000 |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |
| TGA | Pearson Correlation | .849** | .695** | .920** | .856** | 1 | .913** |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | | .000 |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |
| CSD | Pearson Correlation | .824** | .727** | .949** | .760** | .913** | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | .000 | |
| | N | 36 | 36 | 36 | 36 | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 67 shows the correlation table of all the five independent variables correlated with themselves and with the other variables including the dependent variable as gathered from the teacher respondents.

The correlation of each variable with itself gives the value of 1 as seen on the table. For example, the correlation of Test Item Analysis (TIA) on the first column and the TIS on the first row, the Pearson correlation is 1, the sig. (2-tailed) signifies that the level of significance is two-tailed; non directional. All of the significances of the correlations occur at the 0.000 point and all are therefore significant at the 0.01level of the two-tailed.

The correlation of one variable with the other variables of the study indicates that all the correlations are highly correlated with one another with possibly very high inter correlation values. This depicts that the combination of these variables in a test structure would highly vary the creativity development of learners of Biology in secondary schools in Cameroon as concerns the responses from the thirty-six (36) teachers considered in the study.

**Presentation of the findings according to the research objectives using the scattered plots**

Scattered plots present dots which are scattered all over the grid. These dots could be given orientation by a line which indicates positive gradient, negative gradient or neutral depending on the plots. It is in this light that the five objectives are treated in this study.

The five objectives of this study seek to find out the relationship between the independent variables and dependent variable. In correlation, this relationship is seen by a straight line with positive or negative gradient. It is in this light that the scattered plots were conducted and presented as follows:

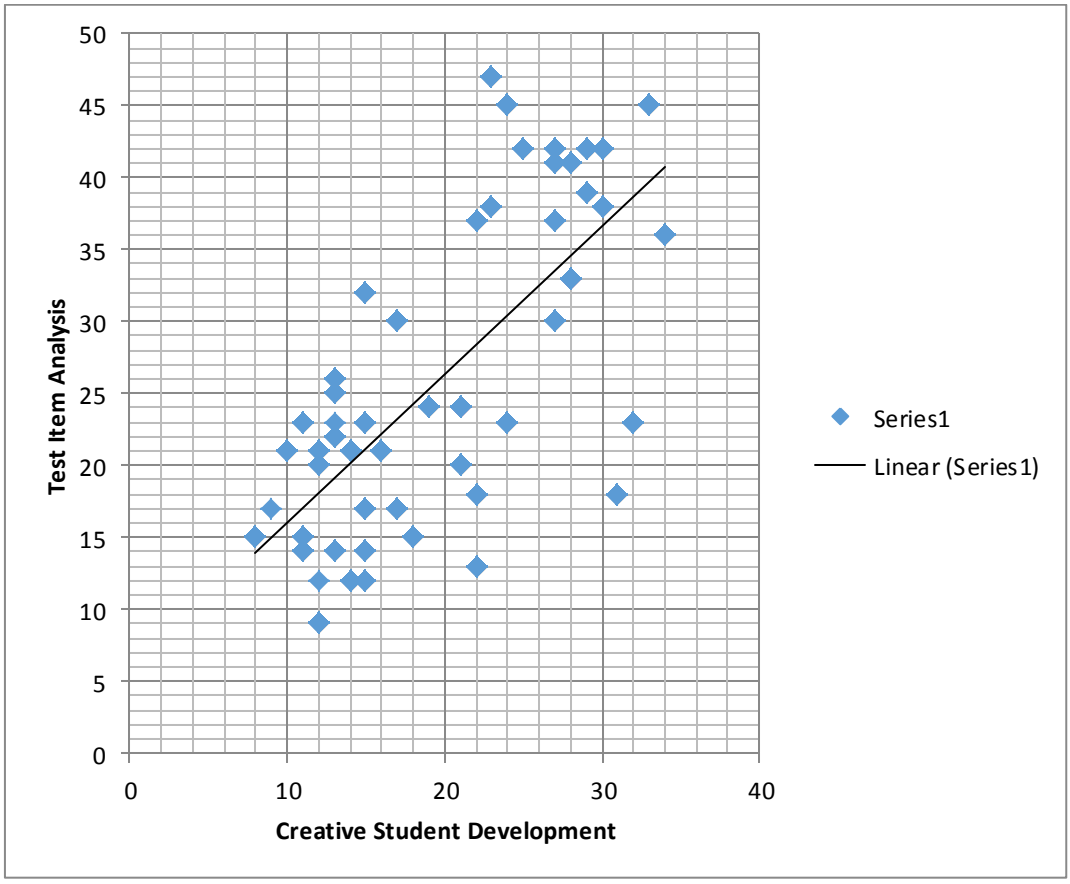
**1- To find out if Test Item Analysis significantly relates with Creative Student Development in Biology in Secondary Schools in Cameroon.**

This is the first hypothesis of the study that links the predictive variable 'test item analysis' to the criterion variable 'creative student development'.

A scattered plot was conducted to find out the possible relationship among the two variables. The best fit line was used and the results as seen in figure 6.

Figure 11

*Scattered plot of TIA versus CSD from student respondents*




From figure 11 we can see that there is a positive correlation between the TIA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of

the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Test Item Statistics significantly relates with Creative Student Development in Biology in secondary schools in Cameroon.

Figure 12

*Scattered plot of TIA versus CSD from teacher respondents*

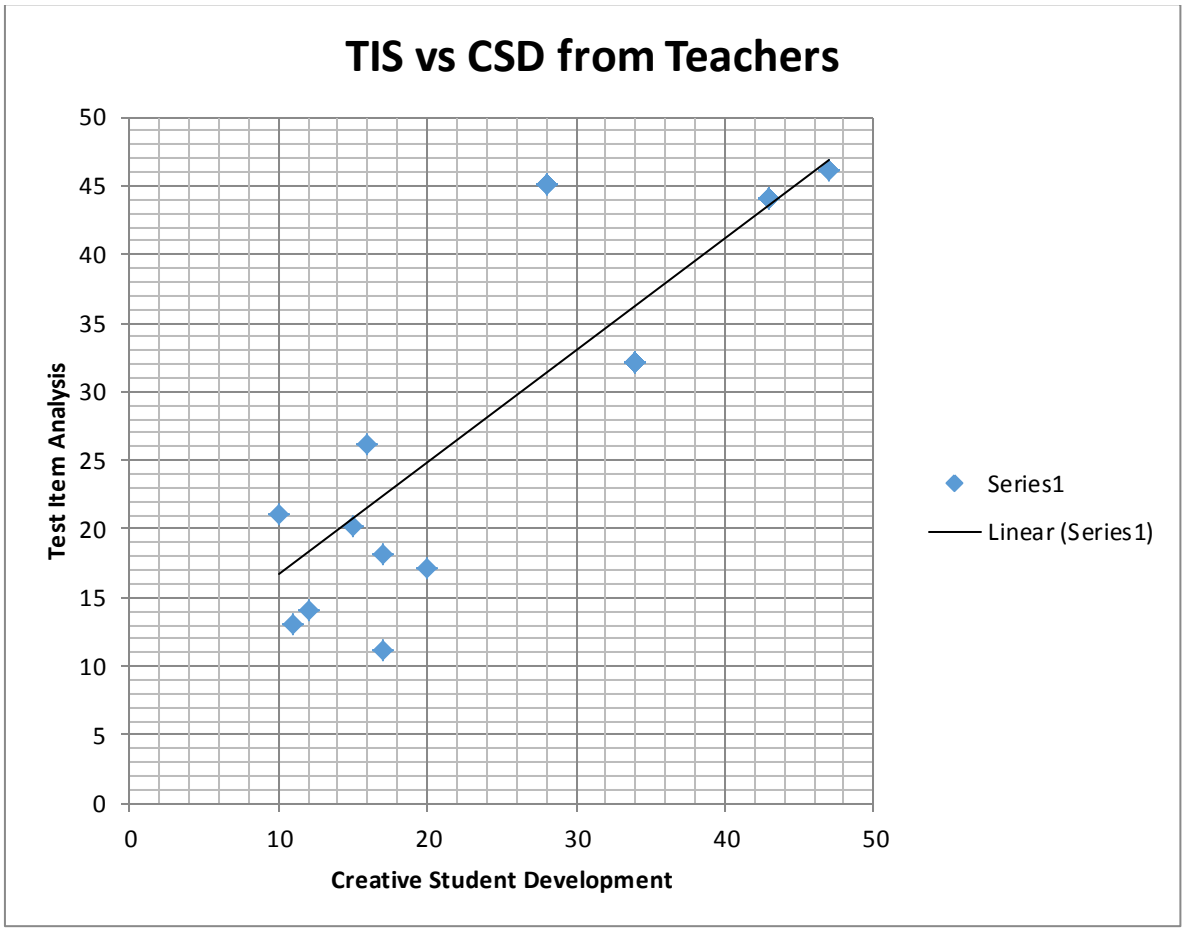


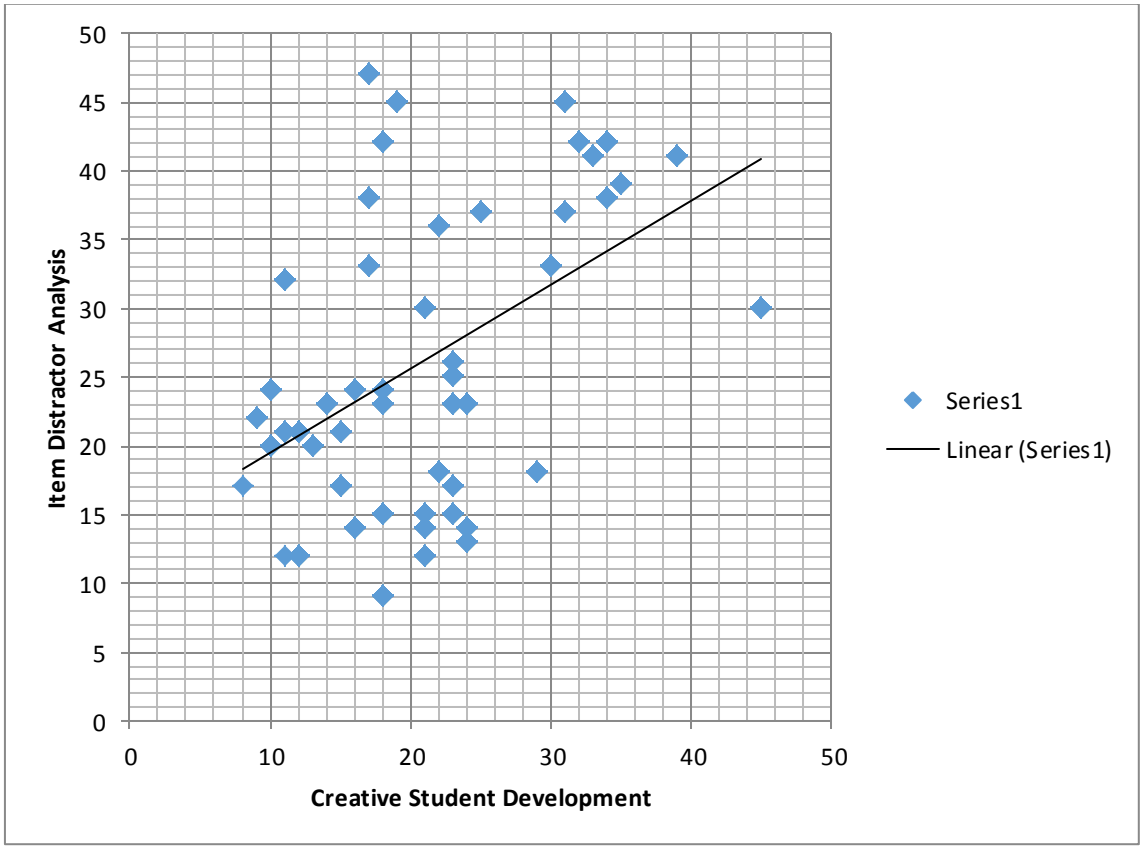

Figure 12 shows a linear correlation of the two variables. The trend line moves from bottom left to the top right, indicating that the correlation is also positive. The dots show the values used for the plot and the concentration of each

of the plots indicates the concentration of certain values to a particular series. The trend line is the best fit of the relationship, indicating the perfect linear relationship for the series. Thus, Test Item Analysis significantly relates with Creative Student Development in Biology in secondary schools in Cameroon.

**2-To bring out the possible level of relationship between Test Item Distractor Analysis with Creative Student Development in Biology in secondary schools in Cameroon**

Figure 13

*Scattered plot of Test IDA versus CSD from student respondents*

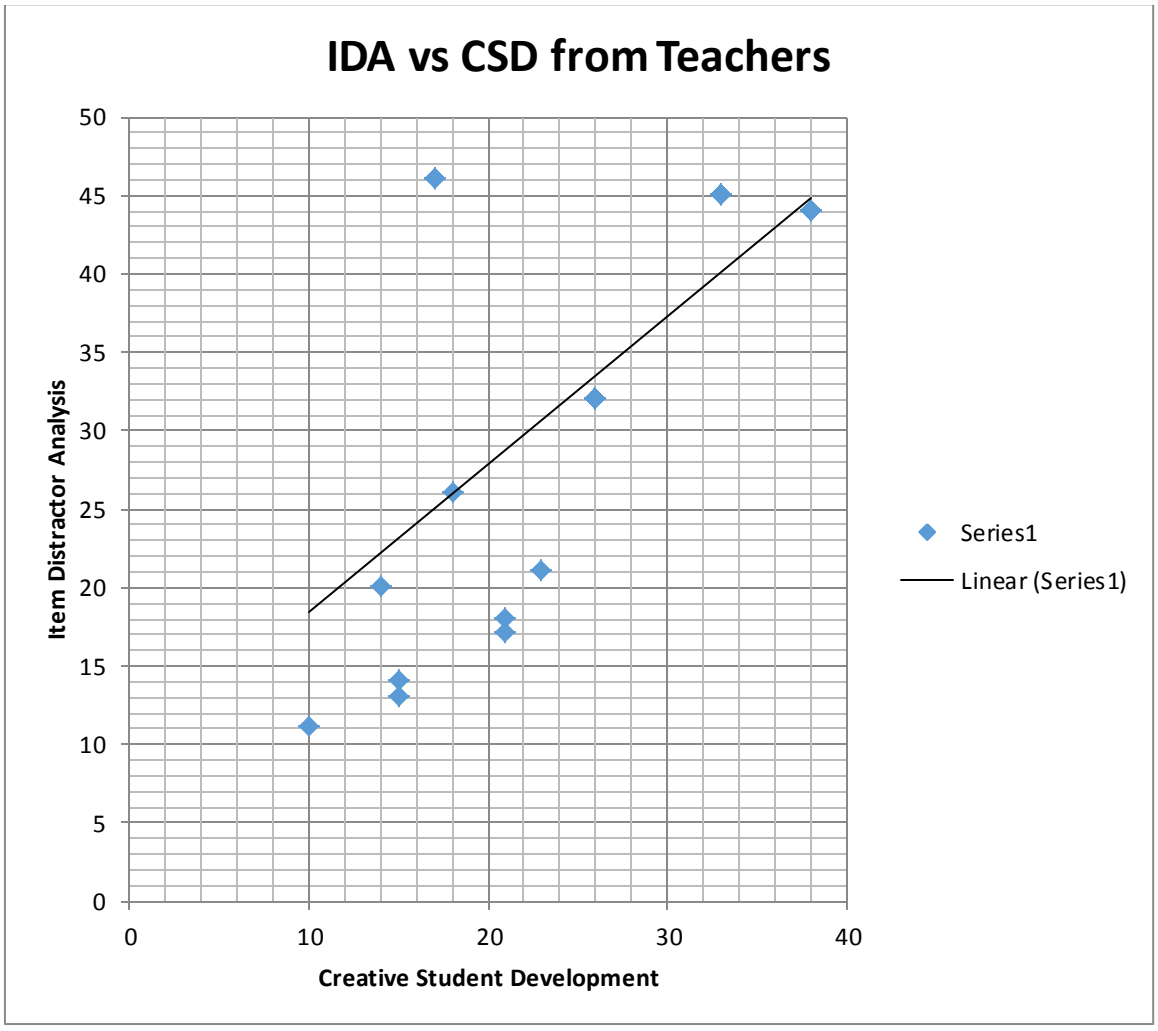



From figure 13 we can see that there is a positive correlation between the IDA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of

the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Item Distractor Analysis to a significant level relates with Creative Student Development in Biology in secondary schools in Cameroon.

Figure 14

*Scattered plot of Test IDA versus CSD from teacher respondents*



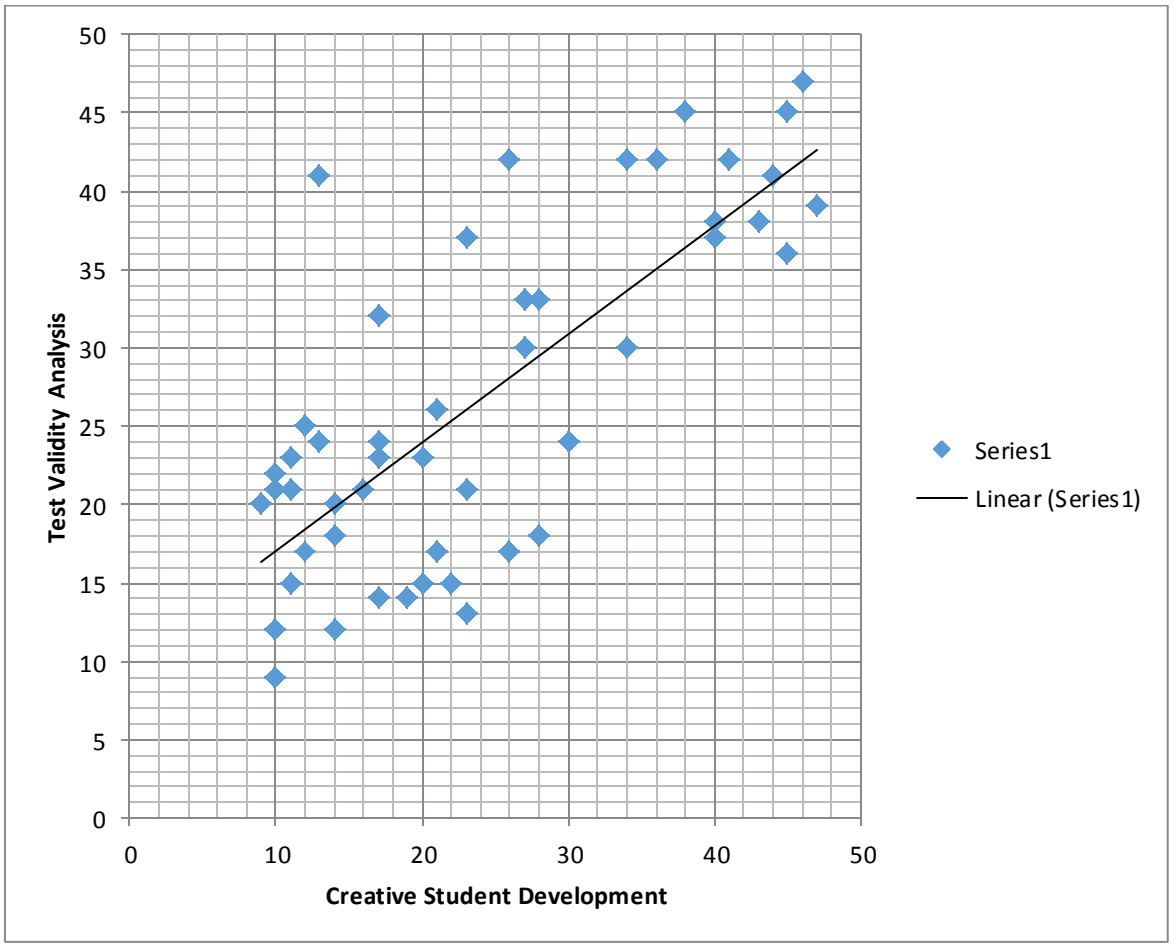

From figure 14 we can interpret that there is a positive correlation between the IDA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of

the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Item Distractor Analysis to a significant level relates with Creative Student Development in Biology in secondary schools in Cameroon.

**3- To determine the degree of significance in the relationship between Test Validity Analysis and Creative Student Development in Biology in secondary schools in Cameroon**

Figure 15

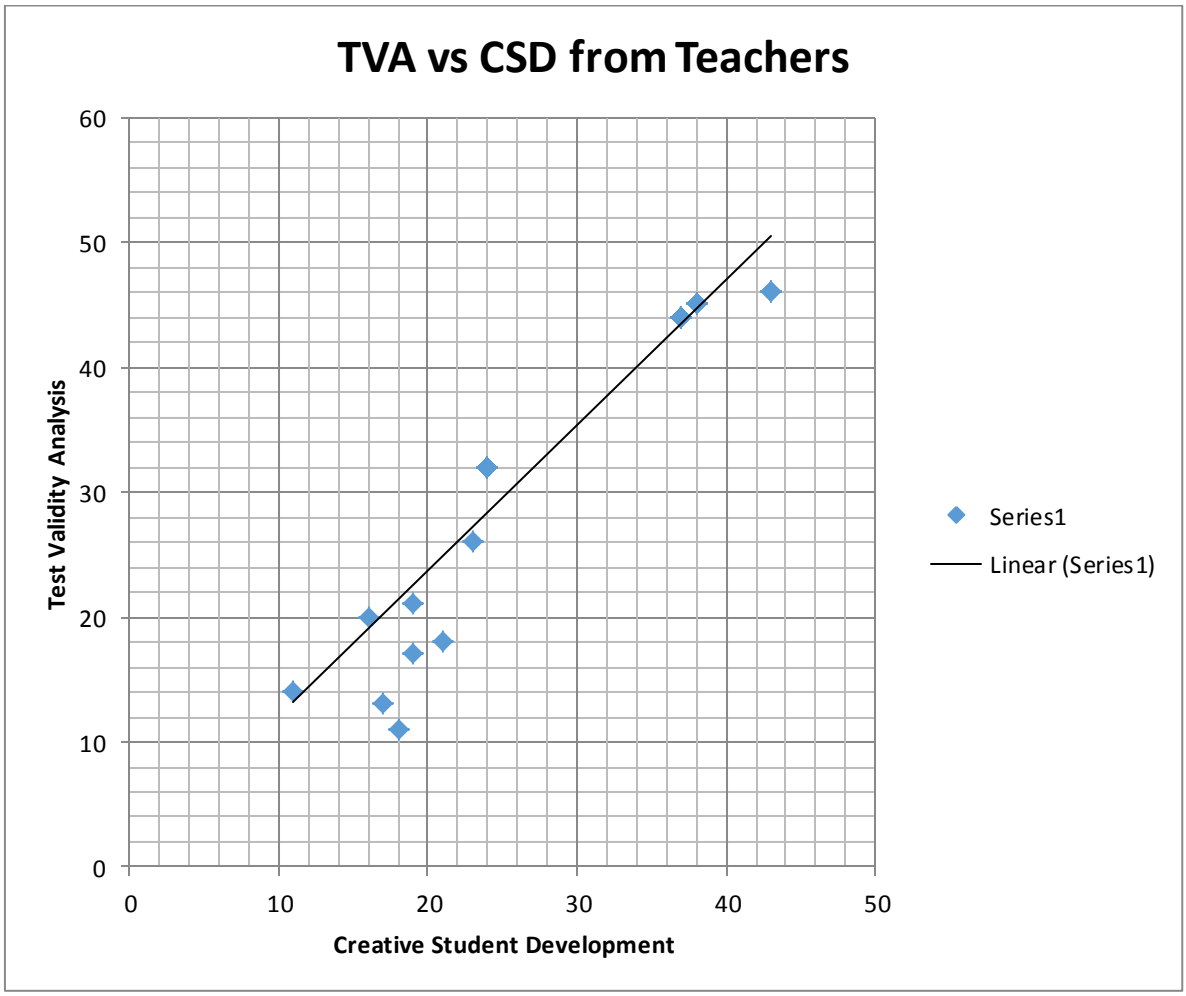
*Scattered plot of TVA versus CSD from student respondents*



From figure 15 we can deduce that there is a positive correlation between the TVA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Test Validity Analysis to a certain degree significantly relates with Creative Student Development in Biology in secondary schools in Cameroon.

Figure 16

*Scattered plot of TVA versus CSD from teacher respondents*

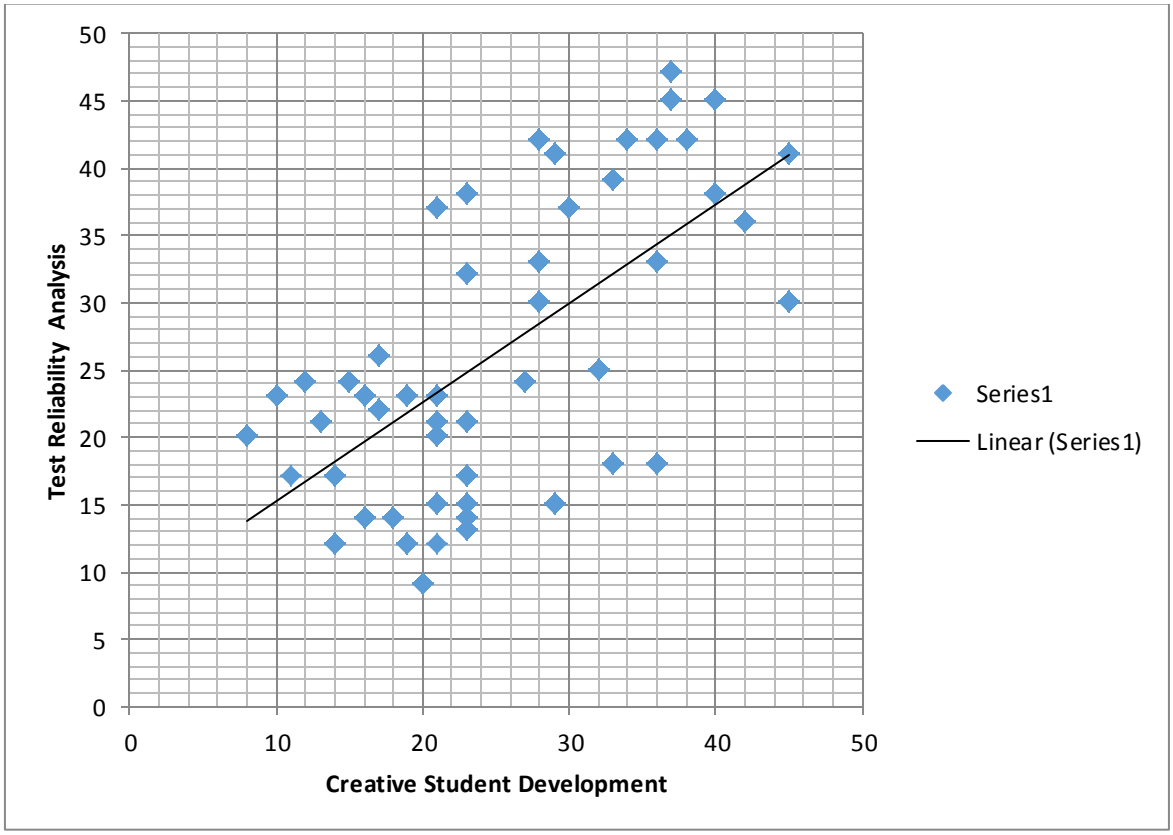


From figure 16 we can deduce that there is a positive correlation between the TVA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Test Validity Analysis to a certain degree significantly relates with Creative Student Development in Biology in secondary schools in Cameroon.

**4- To find out at what level of significance the Test Reliability Analysis relates with Creative Student Development in Biology in secondary schools in Cameroon.**

Figure 17

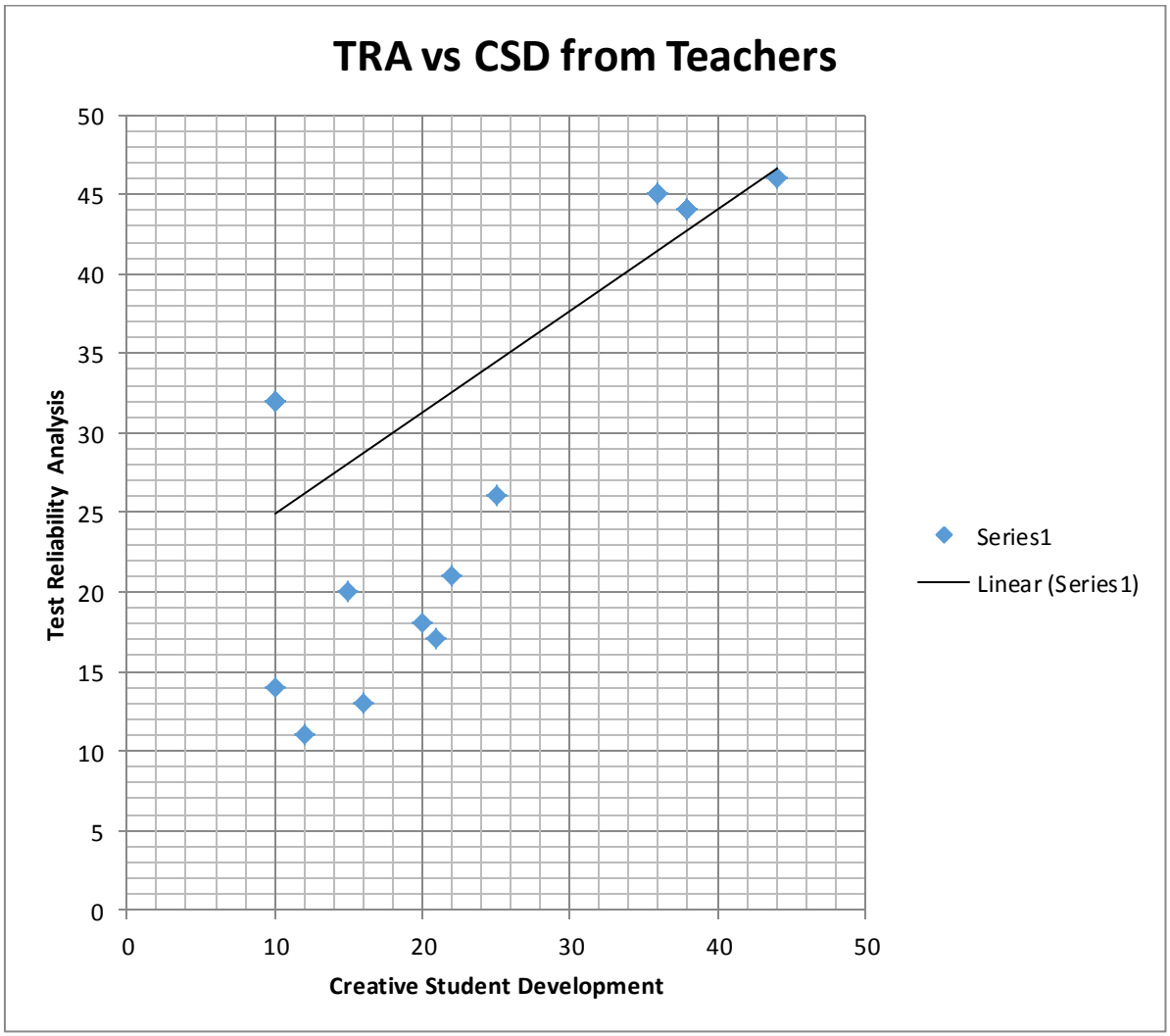
*Scattered plot of TRA versus CSD from student respondents*



From figure 17 we can see that there is a positive correlation between the TRA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Test Reliability Analysis to a significant level relates with Creative Student Development in Biology in secondary schools in Cameroon

Figure 18

*Scattered plot of TRA versus CSD from teacher respondents*

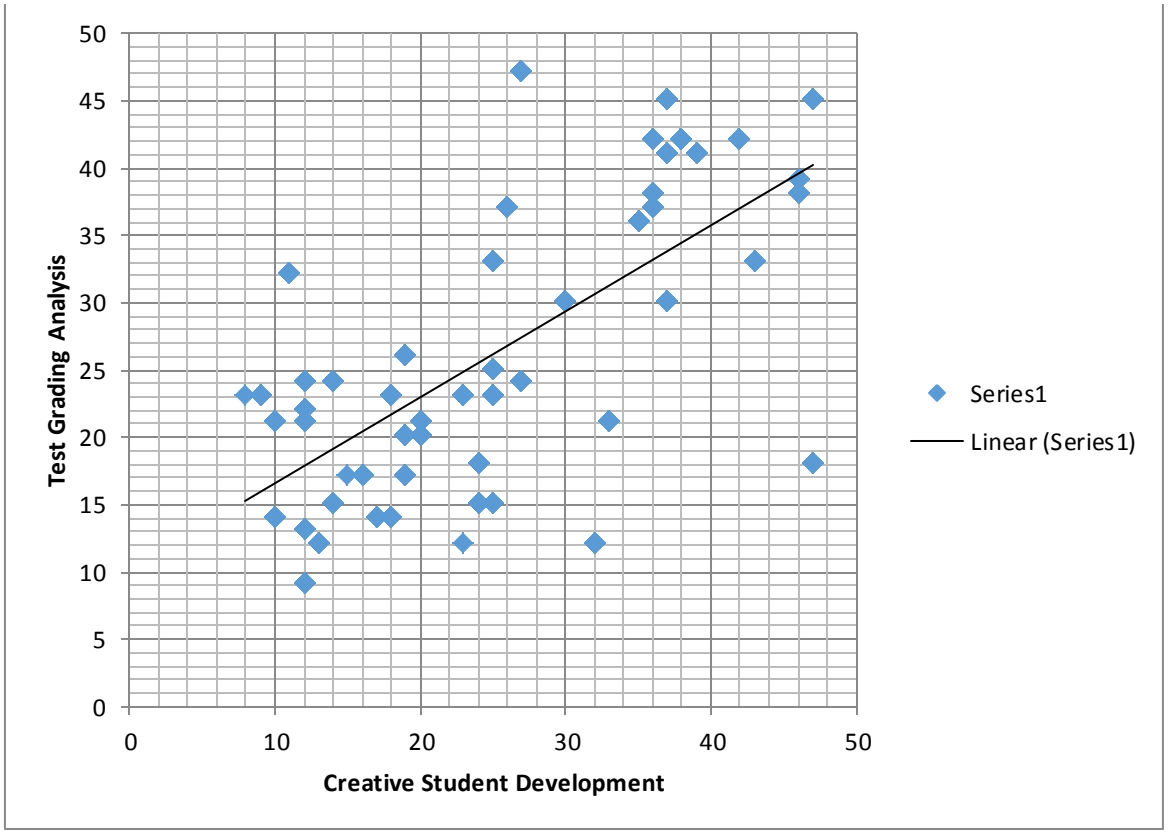


From figure 18 we can read that there is a positive correlation between the TRA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, Test Reliability Analysis to a significant level relates with Creative Student Development in Biology in secondary schools in Cameroon.

**5-To find out the existence of a significant relationship between Test Grading Analysis and Creative Student Development in Biology in secondary schools in Cameroon**

Figure 19

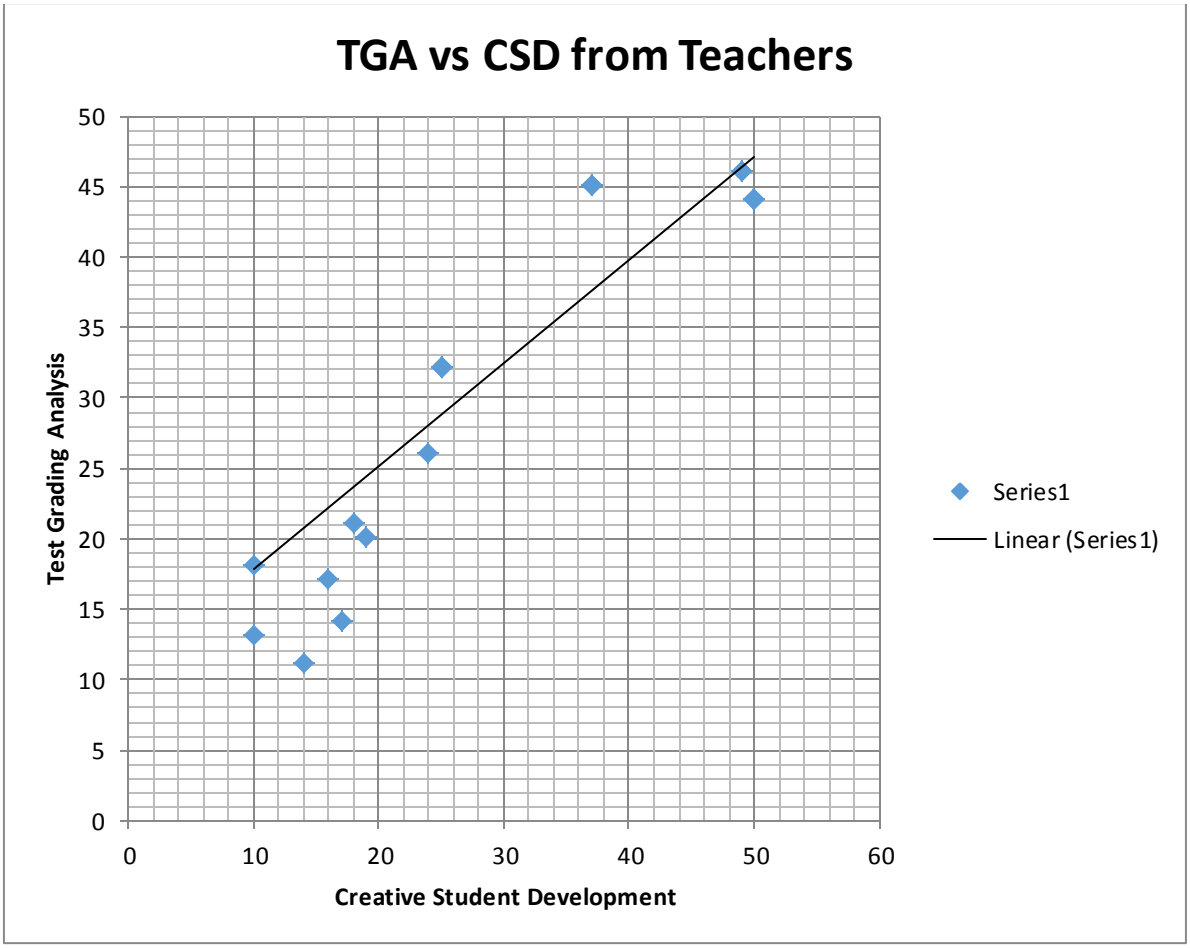
*Scattered plot of TGA versus CSD from student respondents*



From figure 19 we can see that there is a positive correlation between the TGA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, there exists a significant relationship between Test Grading Analysis and Creative Student Development in Biology in secondary schools in Cameroon.

Figure 20

*Scattered plot of TGA versus CSD from Teacher respondents*



From figure 20 shows that there is a positive correlation between the TGA and the CSD. The trend line passing through the dots indicates the gradient to which the relationship can be calculated. The various dots indicate the values of the responses. The values of both scores ranged from 1 to 50. Since correlation is either positive or negative, we can see that the above correlation is positive. Thus, there exists a significant relationship between Test Grading Analysis and Creative Student Development in Biology in secondary schools in Cameroon.

Table 68

*Summary table of results*

| **Variable** | **Correlation with CSD** | **P-value (significance)** | **Sample students** | **Sample teachers** | **Conclusion** |
|---|---|---|---|---|---|
| TIA | Positive | 0.01(2-tailed) | 1603 | 36 | Ha retained |
| IDA | positive | 0.01(2-tailed) | 1603 | 36 | Ha retained |
| TVA | positive | 0.01(2-tailed) | 1603 | 36 | Ha retained |
| TRA | positive | 0.01(2-tailed) | 1603 | 36 | Ha retained |
| TGA | positive | 0.01(2-tailed) | 1603 | 36 | Ha retained |

The level of statistical significance is often expressed as a p-value between 0 and one (1): The smaller the p-value, the stronger the evidence that, 'we reject the null hypothesis and retain the alternative hypothesis'. A p-value typically less than, or equal to 0.05 is statistically significant, thus, all the p-values of the study having 0.01(2-tailed) are strongly significant. These can be seen from the gradients of the scattered plots presented above. There is therefore a strong positive significant

relationship between the test structure analysis and creative student development in biology in secondary schools in Cameroon.

This study has shown the correlation that exists between test structure analysis (TSA) and creative student development (CSD) in Biology in secondary schools in Cameroon. The respondents drawn from teachers and students of all cultural zones and varied contexts of education have proven that great correlation exists between these two variables. It is therefore left for the powers-that-be to see how our children could develop creativity in solving real life problems such that they bring in innovations or new ways of doing the old things that would move our country, Africa and the world to a better position in the empowering of the learners and the youths.

Psychometric practices may be familiar to most teachers but the proper implementation of the required processes as prescribed by renowned theories and protocols are still to be verified. This research can therefore be carried further into bringing out best practices and best results for creative student development for sustainability.

With sponsorship, the researcher could go further to elaborate a sample of standard contextualized test in Biology using either the Classical Test theory or the Item Response Theory or both. I could only develop a test structure for the ordinary level, but promise to develop a sample test beyond this study.

# CHAPTER FIVE

# DISCUSSION, CONCLUSIONS, RECOMMENDATIONS AND PERSPECTIVES

The discussions about the findings of the research work were done in relations to the theories and previous works of others related in the same domain. These discussions were related to what was reviewed in the literature, the theories used, the research questions, and the research hypotheses and the research objectives. The implications of the study, limitations to the study, recommendations and suggestions for further study were also presented in this chapter.

The discussion took into consideration the all the variables in the research objectives, research hypotheses and the research question linking the predictive variables to the criterion variable. The five variables operationalized from the main independent variable are thus discussed and the conclusions drawn thereof.

The conclusion considered the need for the learners and teachers to be motivated intrinsically or extrinsically in order to see tests as a major tool to the development of creativity and critical thinking virtues, and for teachers to be zealous in making the test instrument as valuable as possible in terms of validity and reliability.

The recommendations are valuable elements emanating from the findings of the study which could be useful to the mentioned stakeholders in the section of the work. The consideration of the proposed recommendations by the stakeholders would be of great impact in the assessment and evaluation of studies in the schools.

The perspectives probe into the researcher's view of the impact of the work to the future and the possible opportunities available if the outcomes of this work

were to be exploited to a large extend. These components of this chapter and others like contribution to knowledge would have a great exposition of the highlights of this study to the reader.

**Interpretation and discussion on demographic information**

The demographic information concerned the student respondents and the teacher respondents. The presentation was done in that order; first for the students and later for the teachers. This demographic information is interpreted and discussed because a great influence of one may cause the variation in the results of the study. Though the population of the study was heterogeneous, this common demographic information would guide the readers in the understanding of the research work. Many aspects of the demographic information are influenced by so many aspects too. Among them is the concept of motivation to both the students and the teachers. This is discussed to a limited extent in both the students' and the teachers' findings.

**Interpretation and discussion of demographic information from students**

The demographic information about the student respondents concerned their sex, age range, class and their school. Generally the age range was typically between 15 years and 20 years of age. Majority of the respondents were female students with 53 percent and majority of the students were drawn from the government schools.

**Sex of the students**

Out of the 1603 Biology students who made up the sample of the study, 753 of them were male students, making a percentage of 46.97%. 850 of the students were female, giving a 53.03%. Generally, this reflects the actual population distribution of the Cameroonian society where there are more females than males.

This percentage representation is therefore a replica of the society in which we live, making the sample the right one for the study. Though in some regions more boys are interested in Biology than girls, others show more girls than boys, leading to the given proportion.

**Age range**

A total of 1490 students fell in the age range of 15-20, giving a 92.95% of the total population. The remaining 113 students were in the age range of 21-25; this gave a percentage of 07.05% of the sample. No student of the sample fell in any other age range further than 25 years. This revealing of the sample of students presents the normal age of students worthy to be in the classes chosen for the research following the chronological school ages. The motivation of this group could be easy but their sociological and cultural backgrounds present the need to reconsider individual differences in the setting and analysis of test items for creative student development in biology.

**Class**

There were 993 form five students who offered Biology at the 'O' level GCE sampled out of the 1603 students. This gave a percentage of 61.94% of the total. 610 students were of the upper sixth science class who offered Biology at the 'A' level GCE. This gave a percentage of 38.06% of the sample. Biology is a compulsory subject at the ordinary level G.C.E in most schools but in the high school it is a choice to specialize in the study of science with biology. This reason explains why the number of students in form five was greater than that of the upper sixth class. Their motivation to study biology at these two levels therefore differs as individuals and as per class.

**School**

689 out of the 1603 students sampled were of the government schools. This gave 42.98% of the total. 512 students were of the denominational or mission schools. This gave a 31.95% of the sample. 402 of the students were drawn from lay private schools or colleges. This gave a 25.07% of the sample.

The school administrators, teachers and I, ensured that every questionnaire given out was collected already filled by the respondents. A 100% return rate was therefore ensured and achieved from the respondents in all the schools.

**Interpretation and discussion of demographic information from teachers**

The information on teachers was based on the sex, age range, highest academic qualification the class taught and the number of years of teaching. These in one way or the other could be extraneous variables that could infringe into the study and cause a great effect. This is why the study has considered discussing them for the interest of the readers and other researchers.

**Sex**

Out of 36 respondents, 17 of them were male teachers, making 47.2% of the sample. 19 female teachers responded to the questionnaire, making 52.8% of the sample.

**Age range**

The ages of the teachers in the sample ranged from 31-60 years with more than 56% of them being between 46-60 years. This indicated that most of the biology teachers are experienced in the teaching of Biology, development of tests and the marking of the GCE examinations, thus, appropriate for the study.

**Highest academic qualification**

About 55.5% of the teachers, that is 20 out of 36 of them were holders of the bachelor's degree. This also gave the indication of a good sample for the study to ensure viable and reliable responses for data analysis.

**Class taught**

Generally, as a teacher of Biology each one could teach any level, depending on the need. However, 24 teachers indicated as teachers of form five, giving a percentage of 66.7% of the sample, while 12 teachers, that is 33.3% indicated that they were teachers of the upper sixth science.

**Years of teaching**

More than 80% of the teachers, that is, 29 of them had taught for more than six years, making them the right sample for the study. However, the needs of this group of teachers may not be the same in relation to their motivations in the development of tests or the analysis of test structure. Their cultural as well as social backgrounds make them different. The understanding of these differences and the way of handling them make the study unique in the handling of test structure development and creative student development.

**Years of marking the GCE examination**

24 out of the 36 teachers that made up the sample that is 66.7% had marked the GCE examination at their various levels at least for one year. This percentage appropriately fits the study sample for reliable data. If the policy of setting of test items for the G.C.E. is that of bottom to top, then most of the teachers sampled in this study set test items for the examination and so may do it the right way or follow routine or do it the wrong way. Knowing however that pilot testing of test

instruments exists, this indicates that most teachers have the knowledge of test structure and the analysis of tests. They may also have idea on item banks and test banks but the practice may be very rare or in its most primitive stage. Despite the experience of the teachers in marking the public final examinations, the issue of motivation still stands as the various teachers could respond to motivation in their various ways.

**School**

21 teachers were sampled from the government schools, nine from mission or denominational schools and six from lay private schools. This gave 58.3%, 25 % and 16.7% respectively. Being a good representation since the proportion of the schools differs in that same light. A bulk of trained teachers is in the government schools and most of them take part time teaching in the lay private and denominational schools. The sampling of most of them in the public schools only gives more credibility to the study. It should be noted that the ratio of the samples of the various schools were not determined until after the study. This proportion also presents that the more teachers are of the public sector than the private sector. Since there is interchange of teachers or part time teaching of government teachers in the other schools, this study gains validity in the selection of the teacher respondents. Their motivation could not be too different as they maybe of the same employment conditions:

Herzberg's motivation/hygiene theory, also known as the two-factor theory helps to explain the motivation of teachers in the carrying out of test structure analysis. Herzberg started the study of job satisfaction in the 1950's in Pittsburg. The basis of Herzberg's work is in the Maslow's Hierarchy of Needs. He started with the idea that what causes the job satisfaction are the opposite of those things that cause job dissatisfaction. However, after studying thousands of books he

couldn't draw any guidelines. He conducted a survey where he asked participants to identify those things that made them feel positive with their job and those that made them feel negative. As a result, Herzberg found out that what makes people happy is what they do or the way they're utilized and what makes people unhappy is the way they're treated. Things that make people satisfied at work are different from those that cause dissatisfaction so those two feelings can't be opposite. Based on these findings Herzberg created his theory of Motivators and Hygiene factors. Both factors can motivate workers but they work for different reasons. Hygiene factors tend to cause only short-term satisfaction to the workers while motivators most probably cause longer-term job satisfaction.

Motivators, or satisfiers, are those factors that cause feelings of satisfaction at work. These factors motivate by changing the nature of the work. They challenge a person to develop their talents and fulfill their potential. For example, adding responsibility to work and providing learning opportunities to a person to work at a higher level can lead to a positive performance growth in every task a person is expected to do if the possible poor results are related to boredom of the task they are supposed to accomplish. Motivators are those that come from intrinsic feelings. In addition to responsibility and learning opportunities also recognition, achievement, advancement and growth are motivation factors. These factors don't dissatisfy if they are not present but by giving value to these, satisfaction level of the employees is most probably going to grow. (Bogardus 2007, 34). When hygiene factors are maintained, dissatisfaction can be avoided. When, on the other hand, dissatisfaction is most probable to occur, motivation can't take place.

Hygiene factors, or dissatisfiers, are those that the teachers expect to be in good condition. As motivators are those that at present cause satisfaction, hygiene factors are those that don't cause satisfaction but if they are lacking, it causes job

dissatisfaction. Salaries or wages must be equivalent to those salaries that other teachers in the same or similar school or geographical area get. The status of the person must be recognized and maintained. Teachers must feel that their job is as secured as it is possible in the current economic situation. The working conditions should be clean, sufficiently lit and safe in other way. Sufficient amount of fringe benefits like health, pension and child care must be provided and compensation in general equivalent to the amount of work done. Policies and administrative practices such as flex hours, dress code, holiday schedules and other scheduling affect workers and should be run efficiently. These factors relate to the content of the work and if they are in proper form, it tends to eliminate job dissatisfaction. Ellis, (2005, p.83). The main finding of Herzberg is that the opposite of satisfaction is not dissatisfaction but no satisfaction. Salanova & Kirmanen, (2010)

In the light of Hertzberg, teachers are expected to have some intrinsic or extrinsic motivation in order to do good test structure analysis for creative student development. Since all the teachers are not of the same age range and experiences, they need to be motivated according to their needs for the required output which is test structure analysis in biology in secondary schools in Cameroon.

**Discussion of main findings**

Four main theories explain the basis of the independent variable; 'test structure analysis as used in the study. These include; the Classical Test Theory, the Item Response Theory, the Motivation Theory and the True Score Theory.

The study revealed that the strict use of the Classical Test Theory in the test item statistics will ensure the outcomes of psychological measurement, measuring the ability of test takers such that the creativity of the students could be enforced. The consideration of the item difficulty, item-test correlation, reliability coefficient

and the standard error of measurement as indicators of the CTT would go a long way to ensure that students understand tests and the requirements of tests to explore their creativity in relation to the environment in which they live.

**Interpretation of findings per research question**

Five research questions guided this study. Their discussion and interpretations are linked with the results got from the research work. Research questions have been considered here because the hypotheses have been used above and the objectives further in the chapter. As dynamic as it is, research provides diverse options in the understanding of its processes. Many schools of thought prfer to understand research outcomes through the answering of the research questions and their interpretations, reason why the researcher has decided to treat the research questions in the presentation of the findings of the study.

**1-To what significant extent does test item analysis relate with creative student development in Biology?**

Quantitatively, from the data collected from students and teachers, it can be summarized that test item statistics to a great extent relate with creative student development in Biology in secondary schools in Cameroon. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.689 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.47$, where r =0.689, it is interpreted that 47% of the variances in creative student development can be explained by the test item statistics and vice versa.

So according to the statistics got from the student respondents, test item statistics relate with creative student development to the 0.689 significant extent

such that 47% of the variation of one variable can be explained by the other and vice versa.

In the same light the data collected from the teachers, shows that for a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.824 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.68$, where r=0.824, this indicates that 68% of the variances on one variable are explained by the other variable and vice versa.

Though the coefficients of significance and that of determination are different, both however show a degree of significance. Considering an average of both coefficients of determination however we had 47+68 = 115. This number divided by two gives the average coefficient of determination.

115/2 = 57.5. the value 57.5 thus indicated that more than an average percentage of variances that occur in one variable could be explained by the other variable and vice versa. Thus, test item analysis to a significant coefficient of determination (57.5%) relates with creative student development in Biology in secondary schools in Cameroon.

**2-What significant relationship does the test item distractor analysis have with creative student development in Biology?**

Quantitatively, from the data collected from students and teachers, it can be summarized that test item distractor analysis significantly relates with creative student development in Biology in secondary schools in Cameroon. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.471 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.22$, where r =0.22, it is interpreted that 22%

of the variances in creative student development can be explained by the test item distractor analysis and vice versa.

So according to the statistics got from the student respondents, test item distractor analysis significantly relates with creative student development to the 0.471 significance such that 22% of the variation of one variable can be explained by the other and vice versa.

In the same light the data collected from the teachers, shows that for a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.727 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.53$, where r=0.727, this indicates that 53% of the variances on one variable are explained by the other variable and vice versa.

Though the coefficients of significance and that of determination are different, both however show a degree of significance. Considering an average of both coefficients of determination however we had 22+53 = 75. This number divided by two gives the average coefficient of determination. 75/2 = 37.5. The value 37.5 thus indicated that less than an average percentage of variances that occur in one variable could be explained by the other variable and vice versa. Thus, test item distractor analysis to a significant coefficient of determination (37.5%) is significantly related to creative student development in Biology in secondary schools in Cameroon.

**3-At what level of significance does test validity Analysis tend to relate with creative student development in Biology?**

Quantitatively, from the data collected from students and teachers, it can be summarized that test validity analysis to a great level of significance, tend to relate with creative student development in Biology in secondary schools in Cameroon.

For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.751 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.56$, where r =0.751, it is interpreted that 56% of the variances in creative student development could be explained by the test validity analysis and vice versa.

So according to the statistics got from the student respondents, test validity analysis relate with creative student development to the 0.751 significant extent such that 56% of the variation of one variable can be explained by the other and vice versa.

In the same light the data collected from the teachers, showed that for a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.949 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.90$, where r=0.949, this indicates that 90% of the variances on one variable are explained by the other variable and vice versa.

Though the coefficients of significance and that of determination are different, both however show a degree of significance. Considering an average of both coefficients of determination however we had 56+90 = 146. This number divided by two gives the average coefficient of determination.

146/2 = 73. The value 73 thus indicated that more than an average percentage of variances that occur in one variable could be explained by the other variable and vice versa. Thus, test validity analysis to a very high level of significance of coefficient of determination (73%) relates with creative student development in Biology in secondary schools in Cameroon.

**4-To what extent of significance does the test reliability analysis relate with creative student development in Biology?**

Quantitatively, from the data collected from students and teachers, it can be summarized that test reliability analysis to a great extent relates with creative student development in Biology in secondary schools in Cameroon. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.646 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.42$, where r =0.646, it is interpreted that 42% of the variances in creative student development can be explained by the test item statistics and vice versa.

So according to the statistics got from the student respondents, test reliability analysis to a great extent relates with creative student development to the 0.646 significant extent such that 42% of the variation of one variable can be explained by the other and vice versa.

In the same light the data collected from the teachers, shows that for a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.760 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2 = 0.58$, where r=0.760, this indicates that 58% of the variances on one variable are explained by the other variable and vice versa.

Though the coefficients of significance and that of determination are different, both however show a degree of significance. Considering an average of both coefficients of determination however we had 42+58 = 100. This number divided by two gives the average coefficient of determination.

100/2 = 50. The value 50 thus indicated that an average percentage of variances that occur in one variable could be explained by the other variable and vice versa. Thus, test reliability analysis to a significant coefficient of determination (50%) relates with creative student development in Biology in secondary schools in Cameroon.

**5-What level of significance can test grading analysis provide in relationship with creative student development in Biology?**

Quantitatively, from the data collected from students and teachers, it can be summarized that test item grading analysis to a great extent relates with creative student development in Biology in secondary schools in Cameroon. For a sample size of 1603 student respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.683 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2$ = 0.47, where r =0.683, it is interpreted that 47% of the variances in creative student development can be explained by the test grading analysis and vice versa.

So according to the statistics got from the student respondents, test item statistics relate with creative student development to the 0.689 significant extent such that 47% of the variation of one variable can be explained by the other and vice versa.

In the same light the data collected from the teachers, shows that for a sample size of 36 teacher respondents drawn from the twelve schools in six regions in Cameroon, the correlation gave the value 0.913 significant at the value 0.000, indicating that the correlation had a significance level at 0.01 (two-tailed). Using the coefficient of determination; $r^2$ = 0.83, where r=0.913, this indicates that 83% of the variances on one variable are explained by the other variable and vice versa.

Though the coefficients of significance and that of determination are different, both however show a degree of significance. Considering an average of both coefficients of determination however we had 47+83 = 130. This number divided by two gives the average coefficient of determination of both students and teachers.

130/2 = 65. The value 65 thus indicated that more than an average percentage of variances that occur in one variable could be explained by the other variable and vice versa. Thus, test grading analysis to a great level of significant coefficient of determination (65%) relates with creative student development in Biology in secondary schools in Cameroon.

**Interpretation of findings per Research Hypothesis**

Five hypotheses guided this study. They are all stated in the alternative way to keep a positive view in the outcome of the research and to avoid stating and restating the same thing.

**Ha1-Test Item Analysis to a significant extent relates with Creative Student Development in Biology.**

From the items in the questionnaires given to the students and the teachers in order to get the quantitative measurement of their responses as samples, it is very clear that there is a significant relationship between Test Item Statistics and Creative Student Development.

Taking their common coefficient of determination, we found out that 57.5% of the variances on the Test Item Statistics could be explained by the Creative Student Development and vice versa. In this light the study shows that if we could redress or take into consideration these possible variances in the test items we develop in Biology, the level of creativity development would be highly improved.

The theories proposed in the study like the Classical Test Theory (CTT) and the Item Response Theory (IRT) go to give a wide range of possibilities on the ways we can handle the test items such that the creative student development could be ensured.

On the other hand, for the creative student development, motivation on both sides of the teachers and students themselves is very vital. With intrinsic motivation, students take upon themselves and move on to be creative as could be explained in the theory of self-determination. Motivation from without or extrinsic motivation could come from many aspects introduced into the test by the teacher or test developer such that on taking the test the learners develop the zeal to apply acquired knowledge or skills in the domain or environment in context.

This aspect of extrinsic motivation in the test could be explained from the theory of Flow explained in chapter two; in case of any obstacle students should see an opportunity to devise better or alternative means of solving problems in the context in which they are found.

The issues of test item statistics addressed in the questionnaire included the clarity in terms of the responses needed per item in a test, the language level and visual effect of the test items, the familiarity of the test items in line with the teaching/learning activities, the response time if items and the test as a whole, and the ease in understanding the test items or questions in the tests. These parameters considered in the questionnaire all gave a clear existence of a relationship between test item statistics and creative student development.

**Ha2- There is a significant relationship between Test Item Distractor Analysis and Creative Student Development in Biology.**

Multiple choice questions (MCQs) have been and still remain part of the evaluation process in Biology in secondary schools in Cameroon. While the practice goes on, the vocabulary or technical jargons associated to it are still far to be acquired by the respondents. Hardly could the respondents associate the word 'item' to questions. The word distractors baffled everyone. Nevertheless, for the sake of the responses needed from the respondents these terms were explained and the responses gotten.

It is high time these jargons associated with educational measurement and evaluation be included in the teaching/learning process such that the domain be well understood to foster creative student development.

Notwithstanding, the results obtained from the hypothesis number two on test distractor analysis prove that there is a significant relationship between the two variables; independent and the dependent. Considering the common average of their coefficients of determination (37.5%), there is a relationship in that 37.5% of the variations that occur in the test distractor analysis could be explained by the creative student development and vice versa.

This low coefficient of determination could be traced from the non-mastery of some basic concepts or terms. The items measured under this hypothesis included; understanding of items in relation to proposed distractors/responses, the distractor and right response divergence or convergence, the coherence in the number of distractors per item, the consideration given to item writing flaws in selecting items for a test, and the inclusion of functional and non-functional

distractors in the range of responses. The unfamiliar form of the items posed gave room for any choice of the responses.

Nevertheless, many Biology teachers and some students declared having learned or added new testing jargons into their vocabulary which would help them to improve in test item distractor selection on the part of the teachers and to easily get the right answer in multiple choice questions on the part of the students. Despite the pitfalls, the respondents by their responses established a relationship between test item distractor analysis and creative student development in biology in secondary schools in Cameroon.

**Ha3- To a significant level, Test Validity Analysis relates with Creative Student Development in Biology.**

The indicators in the measurement of this hypothesis in the questionnaire included the following: Face validity in terms of the presentation of items in a test, content validity in terms of what the students have been taught or have learned, criterion validity in terms of laid down rules or criteria or special format of a test as proposed by theories or other tests, construct validity as regards to the logical reasoning, presentation and coherence of facts, and the validity coefficient as gotten from a pretest and associated to the test pack. These, though not also familiar to the respondents, especially the students, clarifications thereof permitted them to respond to the items presented.

From the results got in calculating the common coefficient of determination of both the teacher and student respondents it shows that 73% of the variations that would occur in the test validity analysis could be explained by the creative student development and vice versa. This huge indicator of correlation signifies that the role of validity in an evaluation instrument could not be underestimated. Therefore,

a comprehensive application of the validity parameters in a test or evaluation instrument in Biology would be of great impetus in bringing about creative student development.

The responses given by the respondents indicate clearly that there is a great significant relationship between test validity analysis and creative student development in Biology in secondary schools in Cameroon.

**Ha4-Significantly, Test Reliability Analysis relates with Creative Student Development in Biology.**

The indicators of the test reliability analysis used in the questionnaire were the following: Item revision before inclusion in a test, the discarding of items not fit for the test, the retention of items from a wide range of them in the item bank, item ordering in relation to coherence in the logical flow of ideas, and the reliability coefficient calculated and tagged to a test. These were the indicators of reliability measures which though not quite explicit to the respondents, but for the purpose of the needed responses, they were explained and the responses gotten.

From the results portrayed by the responses, there is a significant relationship between test reliability analysis and creative student development in Biology in secondary schools in Cameroon. A calculation of the coefficient of determination of a correlation among the two groups of respondents each shows that there is a relationship between the two variables; independent and the dependent variable. The average of the two coefficients of determination gave 50%. The indication of this value stems as a result of the fact that very few of the respondents had a little understanding of the reliability concept in the domain of testing or evaluation. By this the responses were given just because they were needed especially after some explanations on the concept.

However, after the explanations, those who understood did what they did and even those who did not understand still had to respond. The final result was that there exists a significant relationship between test reliability analysis and creative student development in Biology in secondary schools in Cameroon.

**Ha5- There is a significant level of relationship between Test Grading Analysis and Creative Student Development in Biology**

This hypothesis may seem to have great toll on the teacher respondents than the student respondents as whatever is concerned about test grading is most likely be introduced or initiated by the teacher or test developer. Nevertheless, with the knowledge of what test grading is all about, it would be of the students' advantage to understand how the grading of their tests is done to equip them on what to present as answers and be sure of the awaited results or outcomes. In this light it would be very much easier for students to plan their work and develop some creative attitudes for sustainable development.

The aspects of test grading exploited in the questionnaire were the following: the type of test involved, item scoring, marking guide, harmonization of the test per level of students, and deliberation board or committee to decide the criteria for grading. Those for the creative student development included the arousal of curiosity in the students in taking the test, ability or possibility of students to carry out experiments on their own and at their will, student willingness to work in collaboration with others in order to solve real life practical and contextual problems, the promotion of sustainable learning and the use of test as means of communication on their acquired knowledge and abilities thereof.

From the responses given by the respondents, it showed that there is a significant relationship between test grading analysis and creative student

development. By the average of their coefficients of determination we got as 65%, it showed that that quantity of variances in one variable would be explained by the other and vice versa. By this the respondents confirmed that if a great deal is adjusted in terms of the grading of tests, much would also be achieved in terms of creative student development. Thus, there is a significant relationship between test grading analysis and creative student development in Biology in secondary schools in Cameroon.

### Chapter Appraisal

The findings show that all the predictive variables are all having positively significant relationship with the criterion variable. This indicates that test structure analysis would have considerable impact on the creative student development in Biology in secondary schools in Cameroon. This study took into consideration the classical test theory, the item response theory and the true score theories which have in them some models and sub theories. This study proposes that with the use of these theories of test development in relation with the learning theories of motivation, flow, self-determination and social constructivism learners would easily become creative in learning and pursuit of relevant knowledge that would enable them sell in the job market or create their own jobs and employ others. The study therefore puts into work all the educational stake holders such that they apply all these theories in test development to assure the holistic development of learners who would become creative and ensure sustainability in learning.

## Implications of the study

Many implications to various stakeholders abound but the ones precised in this study really have great roles to play in order for this study to be exploited. They include:

**To test developers**

Test developers will have to align with the theories of test development; the Classical Test Theory (CTT), the Item Response Theory (IRT), or the True Score Theory. That is they will have to follow the prescription of either one of the theories to guide them develop tests for examinations or particular purposes, or they will have to follow the prescriptions of all the theories.

They would have to develop tests in relation to the content and contexts in which the learning took place and in which the learners could carry out practical studies. The environment must be taken into consideration in developing the test items because they have a huge impact in the students' learning during the time of studies. They would have to follow the prescribed stages in item development or test development and ensure the validity and reliability of the test items to enhance creativity in the learners.

They would have to indicate the test parameters on the test packages and provide clear conditions on the administration of the tests.

The methods and conditions of test grading should form part of the prescriptions for the markers to follow during marking and compilation of the results.

The three domains of bloom's taxonomy should be given relative weighting and items developed and represented accordingly such that the test should not only evaluate one domain but the three; cognitive, psychomotor and socio-affective domains.

**To test administrators**

They would have to ensure that the right test is administered to the right population at the right time in the right conditions that consider the context in which the content was exposed to the learners such that improvising of contextual ideas would not come in to reduce the creativity of the learners.

**To teachers**

Teachers would have to create a test or item Bank such that in relation to the context in which a particular content was exposed to the learners, they would have to be familiar and be reminded on what took place and how they solved a problem and to be creative in problem solving even after school.

They would have to set test items per lesson as they teach such that the learners could recall the scenes and apply the necessary set of conditions necessary in solving the problems.

They may see that the link between items and the whole test be strengthened such that what an item addresses must align with what the test gears or aims to achieve. Testing will no longer be a tedious aspect as many teachers feel. They will only have to select items from the item bank and administer to the right persons at the right environment and context.

**To the learners /test takers**

The blending of one's learning context with the content especially in Biology will put the learners in a privileged position in which reality becomes the backbone of the study. Learners would see, touch and feel whatever they study and tend to look for solutions easily for any problem that arise in biological settings.

This study which had as background the addressing of issues related to quality education and the empowerment of the learner has opened doors to many projects and innovations in its light.

During the conception of this topic and its evolution, guided by my supervisors, a jury of evaluation on the advancement on it, my co-supervisees and some experts of other domains, many insights came up to the inspiration of solving the learners' problems. In this light, innovative projects like "My Class Monitor" in the ministry of secondary education, among others saw their conception and birth, to the light of the day. The results of the findings of this research work may not talk for themselves but their application in the light of solving the learners' problems may go a long way to create innovative projects for the education system or the nation as a whole.

The results of the study show that test structure analysis significantly correlates with creative student development in Biology in secondary schools in Cameroon. Thus, scrupulous implication and application of the following:

Test Item Analysis, Test Item Distractor Analysis, Test Validity Analysis, Test Reliability Analysis and Test Grading statistics in test development and administration practices, would go a long way to impact and improve on students' creativity in Biology.

Following the specific objectives of the study, the following conclusions have been drawn:

1- Test Item Analysis has a positive correlation with creative student development. This indicates that carrying out the process of test item analysis in the classroom by the teachers of Biology; students will be able to become creative following the item response theory and other theories used in the study.

2- Test Item Distractor Analysis in the same light proves that when the right distractors are used in the multiple choice responses, the students will be able to think rationally and critically, thus developing creative thinking skills which serve as a base for creativity development.

3- Test Validity Analysis also presented positive correlation with the creative student development. In this light it is expected that teachers do the validation of the test in most if not all of the validity faces and ensure that the validity coefficients are high enough for a classroom test before giving it to the learners in a test session. A valid test can predict the performance of learners in particular contexts and this will make learners also predict their performances and work harder or put more effort in some areas of the content by further research or group discussions , thus developing creative spirit in the solving of problems.

4- Test Reliability Analysis gave a positive correlation with creative student development in biology in secondary schools in Cameroon. Like its validity counterpart, students should be sure of what they have written in the test and be ready to reproduce it even after the test has been written. A student should be able to have similar score in the same test given some time later. This assures the students of their capabilities and coherence in facts and knowledge and thus they will be able to apply the acquired knowledge in the solving of real life problems, making them creative during and after school.

5- Test Grading Analysis is related to the entire above variable. It showed a positive correlation with creative student development. Analyzing students' grades will depend on the type of tests and performances of the learners including the purpose of the test. If the above variables are given due consideration in the development of tests and items, the scores obtained by the students will be

considered valid and reliable at all times. In the same light their creativity skills will be given due consideration in and out of the school.

Recommendations and suggestion have been addressed to the major actors of education that may be involved in the implementation of the outcomes of this research work.

**Limitations of the study**

Due to the prevailing war situation in the North West and South West regions of the Country, the researcher could not go to certain places to enlarge the scope of the study, reason why I ended at the central town of each region of the country chosen in the study.

Considering that this programme of psychometrics seemed new in the study programmes of the schools in the country, books on the topic were very hard to find, few were bought from some authors around but many could only be commanded online or from abroad.

Internet services in Cameroon are still very expensive and the lines are relatively slow.

The concept of measurement and evaluation put many people in an uncomfortable state as they thought that the researcher had been sent to spy or evaluate their work or output, so the teachers gave some information with much remorse, while others remain skeptical and would not want to fully cooperate. However, the timely intervention of the school hierarchy in some cases and my explanations and presentation of official documents, made most of them reasons to fully participate in the study.

- Anglophone crisis led many schools closed in the NW/SW regions.

- Educational Assessment and Research Network in Africa (EARNiA) meeting from November 1-6, 2019, engaged in preparations then postponed indefinitely due to COVID-19
- COVID-19 lockdown from March 2019 to October 2019 closing schools and movements

- Lack of sponsorship to further produce sample test instrument in Biology for the various regions or cultural areas.
- The writing of articles in line with the work
- Professional work and studies too tedious to carry along

**Conclusion**

This study has not been commonly carried out in our country and in our universities. The search for books articles and journals on the analysis of the test structure did not prove very favourable so the researcher considered this study in the light of being a basic study in the bridging of this gap so that many further studies could consider the results and carry out other studies in the same domain, same subject or different subjects. The projection of the results to the whole of Cameroon could give a clear image of what could be tackled without further basic studies. The exploitation of the results of this research will provide a roadmap for the analysis and development of test items worth the name for our schools and examination boards.

**Recommendations**

Generally, the stake holders concerned should:

1- Introduce the teaching of measurement and evaluation in the teacher training colleges as a practical subject evaluated during the teaching practice, where

students must set tests following laid down rules or principles in relation with given theories.

2- Have aspects of measurement and evaluation introduced to students of secondary schools in particular subjects or on its own, to enable them understand the needs of particular tests and how to go about contextualizing responses to questions in different contexts.

3- Send old teachers to schools for recycling on short courses to study their arts and science of measurement and evaluation such that they can easily embrace and follow the new paradigm in measurement and evaluation in our schools.

4- Organise seminars and workshops on the test structure analysis and its implications in the teaching/ learning process in our schools today on regular bases in the various regions or cultural spheres to introduce and follow up the way forward of the test structure analysis.

5- Ensure that learners are taught using the approved curriculum from the government policy and not the examination syllabuses from the examination board. This narrows the scope of the learners such that they would lack creativity due the inability to contextualize problems and seek their solution correspondingly.

6- Programmed the topics taught in Biology following the times and seasons in which students can easily apply the acquired skills in contexts where they learned and where they would be found.

7- Consider tests in Biology in the local realities of what are available and possible hand-on practical activities by all.

8- Draw up the syllabus of Biology considering the diverse regions or cultural spheres of the country.

9- Decentralize the assessment of Biology in the end of course examinations to reflect the diverse flora and fauna, so as to assess following the realities on the ground.

10- Teach Biology despite the prescriptions of the national syllabus of Biology, using the locally available samples for demonstrations other than importing samples or slides. This will make them be familiarized with the organisms and can manipulate them to solve local problems.

**To teachers / test developers**

Need to:

1- Train other teachers in the selection of distractors to put with the right answers such that the right answer should not be obvious even to the eyes of those who don't master the domain.

**2-** Organise seminars and workshops in the areas of the research to put teachers abreast with the need to own, develop and maintain an item bank or test bank as we teach along the years.

3- Be guided by test theories that guide the construction of test items which must be followed if there must be creative development in learners.

4- Make use of the test development guides and procedures proposed by the various test theories and get to develop good tests.

5- Use test items that come out of the discussions during lessons; asked by students and the teacher, and not from past questions or text books.

6- Consider test as a means of communication as to what the learners have learnt or have mastered and what they find difficult to consolidate.

7- Stop using test as a means of getting to sanction lazy students or to reward hardworking students.

8- Set a test based on the experiences of the learners and on what they have been exposed to, in relation to content and context.

9- Understand that a test is not to have a grade or score for the students; it is a means of enabling the learners prove their worth in solving practical live problems which they encounter on daily basis in their societies or communities.

10- Note that test items should be tested for validity and reliability, according to the test takers' characteristics. A valid test in Maroua in the month of April may not be valid in Buea: It is the peak of heat and dryness in Maroua while rain had been falling for about one month in Buea.

11- Stop teaching contents on Knowledge (remembering) only to test on their application. They should teach them the application of the knowledge if they must test or evaluate them on that.

12- Recognize the fact that learners are not the same all the time, the calculation of true score is needed as many psychological aspects may vary a test score.

**To test administrators**

They should:

**1-** Consider test accommodation in the test administration. A student who was fine yesterday may have a problem today and the testing environment may not be the best.

**2-** Scrupulously follow guidelines to testing as the conditions may vary from place to place.

**3-** Ensure the learners see the unveiling of the test scripts and the assessor contents for proper psychological preparedness.

**To test takers**

They should:

**1-** Endeavor to read and understand test guides in order to be sure of the nature and exigencies of the test.

**2-** Contextualize their responses to questions following their area of familiarity than generalize in contexts they are not familiar with.

**3-** See test as a means of solving real life problems and responding to it accordingly.

**4-** Be able to use the test given to them as a challenge in addressing live problems and not as a sanction or reward, but a worry which must be given a solution and so go ahead to find solution even after the test.

**To policy makers**

They should:

1- Ensure tests are set according to the requirements of a given region or community and not as the requirements for certification only.

2- Provide opportunities for test developers to be trained so as to ensure reliability and validity of the tests.

3- Create test verification services or departments in schools so as to ensure quality tests in the classrooms and operate item banks and test banks for the school.

**To examination boards**

They should:

1- Recycle workers for modern test item development practices

2- Accompany schools in the contextualizing of test items for creative student development.

3- Create a directorate for test development following the norms for creativity in students

4- Ensure tests and examinations are marked following diverse cultural and environmental differences.

5- Calculate the true scores of candidates following their diverse contexts and contents.

6- Carry out seminars and workshops in the areas to put teachers abreast with the theories of test development.

**Perspectives**

- The possible development of test instruments in Biology that will address the pertinent needs of each studied area for creative student development.
- Possibilities of Biology practical in the laboratory to be accompanied by practical field works to enable hands-on reality solutions to field problems.
- The creation of content verification departments in the various education boards and/or test content verification committees in the various departments in schools.
- The creation of test item review boards in charge of various examinations and/or test item review departments in the various examination boards and/or test item review committees in the various departments of the school.
- The inclusion of educational measurement and evaluation as a course or subject in secondary schools to enable students grasp the mastery of what

they are expected to be tested on and thus engage in personal creativity development.

## Contribution to Knowledge

Test Structure Analysis provides the teachers and learners with valid and objective information on which each one should base their actions. It gives a vivid guide to the examiner as to the domains and competencies involved while at the same time it provides a clue to the responses of the learners; whether the responses fall in the right domain or in other different domains of the same study or out of the study.

This study has come up with samples of test structure table, test specification table (or Test Blueprint) and a sample classroom test presented as follows:

Table 69

*Test Structure Table on Biology Ordinary Level*

| **Content Specifications Domain** | **Number of Competencies** | **Approximate Weighting** | **Approximate number of Multiple-Choice Items** | **Number and type of constructed-response items** | |
|---|---|---|---|---|---|
| | | | | **Focused educational problem and instructional tasks** | **Case study** |
| **Domain 1** Nature and variety of living organisms | 5 | 20% | 15 | 1(illustrative and diagrammatic presentation of responses) | 0 |
| **Domain 2** Functioning of living organisms | 10 | 60% | 35 | 2(experimental presentations, balanced equations and arrow diagrams) | |
| **Domain 3** Environmental Biology and human activities | 3 | 20% | 10 | 1(Problem solving approach to natural and man-made environment) | |
| **Total** | 18 | 100% | 60 | 4 | 0 |

The three major domains of the subjects and their contents are presented on the first column, the number of identified competencies on the second column, the weighting per content domain on the third column, the approximate number of multiple choice items on the fourth column. Possible essay type questions on the fifth column and sixth columns.

Table 70

*Test Specification Table (or Test Blueprint)*

| | | | **Cognitive Levels** | | | |
|---|---|---|---|---|---|---|
| **Domain** | **Competencies** | **Weight** | **Remember 45%** | **Comprehension 30%** | **Application/Analysis/ create 25%** | **Total Items** |
| 1 | 1.1 | 8% | 1 | 0 | 2 | 3 |
| | 1.2 | 2% | 1 | 0 | 1 | 2 |
| | 1.3 | 4% | 1 | 1 | 0 | 2 |
| | 1.4 | 2% | 1 | 2 | 3 | 6 |
| | 1.5 | 4% | 1 | 1 | 0 | 2 |
| 2 | 2.1 | 8% | 1 | 1 | 0 | 2 |
| | 2.2 | 6% | 1 | 1 | 0 | 2 |
| | 2.3 | 4% | 3 | 1 | 1 | 5 |
| | 2.4 | 10% | 1 | 1 | 0 | 2 |
| | 2.5 | 4% | 1 | 1 | 0 | 2 |
| | 2.6 | 12% | 0 | 1 | 1 | 2 |
| | 2.7 | 6% | 5 | 2 | 1 | 8 |
| | 2.8 | 3% | 4 | 1 | 1 | 6 |
| | 2.9 | 4% | 2 | 1 | 0 | 3 |
| | 2.10 | 3% | 2 | 1 | 0 | 3 |
| 3 | 3.1 | 5% | 1 | 1 | 1 | 3 |
| | 3.2 | 12% | 0 | 1 | 4 | 5 |
| | 3.3 | 3% | 1 | 1 | 0 | 2 |
| **Total** | 18 | 100% | 27 | 18 | 15 | 60 |

As a continuation of the test structure analysis, after the analysis, the test items were developed, the test constituted and administered, then carry out analysis to determine which of the domains/ competencies were covered by the candidates'

responses, whether they were representative or not. Below is the sample of a classroom test in Biology:

**Sample Biology Test from the test structure analysis in this study**

*Place a tick (√) against the letter (A or B or C) corresponding to the right answer to the given description or Question from 1 to 20. (1x20Marks)*

*C1.1 (3 items)*

1. An example of an amphibian
    A. gecko
    B. fish
    C. frog
2. Non pigmented plants-like growing organism
    A. fern
    B. fungus
    C. spirogyra
3. A class of organism related to human by the manner of birth
    A. mammalia
    B. hummalia
    C. humminidae

*C1.2 (2 items)*

4. Small pores in the epidermis of the leaf used for gaseous exchange
    A. flaccids
    B. porous
    C. stomata
5. Components of air send out by plants during the day
    A. $CO_2$
    B. $O_2$

C. CHO

*C1.3*

6. A vital organ for gaseous exchange in man

A. lung

B. bronchus

C. nostrils

*C1.4*

7. Organism that feed on dead organic matter

A. parasites

B. holozoits

C. saprophytes

*C1.5*

8. One reason for classifying rat and human in the same taxon

A. they give birth to live young ones

B. they use their teeth to eat

C. their female care for the young

*C2.1 (3 items); Questions 9, 10 and 11are based on experiments on photosynthesis*

9. To prove the presence of starch in a leaf the colour changes to

A. brown

B. pure white

C. dark blue

10. A requirement to test if sunlight is necessary for photosynthesis

A. aluminum foil

B. cardboard paper

C. polythene bag

11. A substance not use in the experiment to show that carbohydrate is necessary for photosynthesis

A. iodine

B. caustic soda

C. ethanol

*C2.2 (2 items)*

12. An organ specialized in the removal of waste products of metabolism from the human body

A. Skin

B. Kidney

C. Anus

13. The movement of plant food made In the leaves to other parts of the plats is known as

A. Translocation

B. Transportation

C. Transpiration

*C2.4*

14. This organism carries out respiration through the gills.

A. Tadpole

B. Frog

C. Toad

*C2.5*

15. This organism has an elaborate dentition for shearing of its prey for food.

A. snake

B. crocodile

C. monkey

*C2.7*

16. A reptile adapted to camouflage in its environment.

A. lizard

B. earthworm

C. chameleon

*C2.8*

17. Movement by this organism is done on eight legs.

A. cockroach

B. spider

C. bee

*C2.10*

18. A pair of tactile organs for sensory perception in insects.

A. nerves

B. compound eyes

C. antennae

*C3.3*

19. To naturally enrich the soil to boost plants growth, farmers are advised to apply

A. fertilizer

B. wood ash

C. manure

*C3.2*

20. In a food chain relationship, the spider feeds on the housefly. In this case the spider is called the

A. predator

B. victor

C. prey

Generally, as a study involving test items development, analysis, and selection, discarding, modifying and reinserting into the item and test banks, the procedure cannot just end in the setting of test items for the students. It goes further to test all the validities per item before validating its use in a test. This is not a procedure for a day or for a single person, so the curriculum experts, teachers and all the test developers and psychometrics experts need to put all hands on deck to see into it that the tests set for the classroom or end of course examinations be valid and reliable following laid down norms of international acceptance.

# REFERENCES


Abraham, J. A., & Brown, C. J. (2011). *The role of the counsellor in Nigeria secondary schools*. Calabar: Rapid publishers.

Adegoke, B. A. (2013). *Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response* Theory Frameworks. Journal of Education and Practice, 4(22), 87-96.

Adeleke, J. O. & Mitee, J. L. (2017), *Psychological Testing and Innovative Teaching Strategies*. ELYON 4ME publishers. Nigeria.

Adeleke, J. O. (2010), *The Basics of Research and Evaluation Tools*. Somerest ventures, Ikeja- Lagos.

Adewale, J. G. (2016), *Emerging Trends in Educational Measurement, Assessment and Evaluation in Africa*. Esthom Graphic Prints, Nigeria.

Akinade, E. A. (2009), *Psychological Testing in counselling*. *CASSON Journal 25* (1), 86-95

Akume, G.T. (2010), *Personality Development and Vocational Guidance*. Makurdi Nigeria: density venture

Alao, I. (2009) *Counselling and Nigerian National Policy on Education: The Question of Relevance and Competence*. The Nigerian Journal of Guidance and Counselling 14, 134 – 141

Allen, I. E., Seaman, J., Poulin, R., &Straut, T. T. (2016). o*nline report card: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group. Retrieved from Education Advisory Board Website on 15 Dec 2016 from http://onlinelearningsurvey.com/reports/onlinereportcard.pdf

Allen, M. J., & Yen, W. M. (1979), *Introduction to Measurement Theory.*

Monterey, CA: Brooks/Cole Publishing Company.

Allen, M.J. *(2004), Assessing Academic Programs in Higher Education. San Francisco:* Jossey-Bass.

Aluede, O., Mcearchern, A. G. & Kenny, M.C. (2005), *Counselling in Nigerian and the United States of America: Contrasts and Similarities.* International Journal for the Advancement of Counselling 27, 271-381

Alutu, A. N. G. (2005). *The status of guidance services in Secondary schools in Edo State*. The Nigerian journal of guidance and counselling 10, 114-125

Amin, M. E. (2004), *Foundations of Statistical Inference for Social Science Research*. Makerere University Printery, Kampala Uganda.

Amin, M. E. (2005). *Social Science Research: Conception, Methodology & Analysis*. Uganda: Makerere University Printery

Amod, ZA 2003: *A problem solving psycho educational assessment model for implementation for educators by South African schools.* Unpublished PhD thesis: University of the Witwatersrand, Johannesburg.

Anagbogu, M. A. (2008), *Problems of Guidance and Counselling in primary schools*. Lagos

Anagbogu, M. A., & Nwokolo, N. (2010), *Guidance and Counselling in primary schools*, Lagos: Mark New World.

Anastasi, A. (1976). *Psychological Testing* (4th ed.). New York: Macmillan publishing Co.Inc

Andeson,G. and Arsenaut, N. (2001). *Fundamentals of educational research U*K.

Asagwara, C.G. (1994) *Readings in Applied psychology in Education Volume one for_Universities Colleges of Education and polytechnics*: seasprint (Nig.) Co 20 Fillerstreet, Calabar

Assessent tools for Teaching and learning (asTTle)

ASTD, (1998), "*National HRD executive survey*: *measurement and Evaluation* (1997)", Fourth quarter survey Report, available at ww.astd.org/virtual-community/research/nhrd- executive survey-97me.htm.

Atumasa, B. A. (2009). *Diversification of evaluation practices: An objectivity in placement in professional schools in Cameroon. A survey of some professional schools in Yaounde Mfoundi division*, Post graduate Dissertation ENS Yaounde.

Atumasa, B. A. (2021). *Evaluation Practices that enhance objectivity in placement in professional schools in Cameroon*. www.academia.edu

Atumasa, B. A. (2018). *Test development practices for sustainable learning competencies in Biology in secondary schools in Meme division*. Masters dissertation; UBa Library.

Atumasa, B. A., Tanyi, M. E., Kongnyuy, P., Egbe, M. (2022). Psychometrics in the Classroom: The How, the why and the way forward. *Global Scientific Journal*. (unpublished)

Bassey, M. (1999). *Case study research in educational setting*. Philadelphia

Bates, R. (2004). A critical analysis of evaluation practice. The Kirk Patrick Model and the principle of beneficence. *Evaluation and program planning 27*(3), 341-347.

Bejar, I. I. (1993). A Generative Approach to Psychological and Educational Measurement. In N. Frederiksen, R.J. Mislevy,& I.I. Bejar (Eds.), Test Theory for a New Generation of Tests (pp. 323-357). Hillsdale, NJ: Erlbaum.

Best, JW & Kahn, JV 2006: Research in education. 10th edition. Boston: Pearson Education Inc.

Bhattacherjee, A. (2012). *Social science research: principles, methods and practices*.

Black, C. (2002). *Amending assessment for a dyslexic student.* http://www.nottingham.ac.uk/pesl/resources/assessment/amending100/ http://evaluationtoolbox.net.au/index.php?option=com_content&view =article&id =15&Itemid =19

Bloom, B.S. (Ed.). Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. David McKay Co Inc.

Borsboom, D. (2006). *The attack of the Psychometricians Psychometrika*,71(13), 425-144.

Candy and Crebert (2006), Sustainable Learning, Ontario, Ciea.

Castling, A. (1996). *Competence-based Teaching and Training,* London: MACMILLAN PRESS LTD.

Coady, J. & Huckin, T. (Eds.). *Second language vocabulary acquisition: A rationale for pedagogy.* Cambridge: Cambridge university press.

Carole L. K., & Winterstein, A. G. (2008). Validity and Reliability of Measurement Instruments used in Research. *American Journal Health-System Pharmacy*, Vol. 65 Dec 1, 2008

CGCE (2009): Ordinary Level Biology Syllabus for the Cameroon General Certificate of Education

Clark, R., Chopeta, L. (2004). *Graphics for Learning : Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials.* Jossey- Bass/Pfeiffer

Courville, T. G. (2004). An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Statistics. Unpublished Ph.D Dissertation, Texas A & M University.

Creese, A; Norwich, B & Daniels, H 2000: *Evaluating teacher support teams in secondary schools: supporting teachers for special educational needs*. Research papers in education. 15(3): 307-324.

Crocker, L., & Algina, J. (1986). Introduction to classical & modern test theory. Orlando, FL: Holt, Rinehart and Winston

Department of Education, 2008: National Strategy on Screening, Identification, Assessment and Support School Pack. Pretoria: Government Printers.

David, M., & Sutton, C. D. (2004). *Social Research*: *the basics*. London: Sage Publications.

Devies, S. (2010). Effective Assessment in a Digital Age. Retrieved on 28 August 2018 from www.jisc.ac.uk/elearningprogramme*Effective Assessment in a Digital Age: A guide to technology-enhanced assessment and feedback.*Retrieved on 28 August 2018 http://www.jisc.ac.uk/media/documents/programmes/elearning/digiassass_eada.pd f

EARNA, (2015), African Journal of Theory and Practice of Educational Research. Vol. 1, vol, 2, 2015

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of Educational Measurement (5th Ed). Engelwood Cliffs, N.J: Prentice Hall.

Edgar F. (2006). *Sustainable Learning*, Ontario, Ciea.

Erguven, M. (2014). Two Approaches to Psychometric Process: Classical Test Theory and Item Response Theory. Journal of Education, 2(2), 23-30.

Eric Grodsky, John Robert Warren and Erika Felts Annual Review of Sociology, Vol. 34 (2008), pp. 385-404 Annual reviews

Fonkeng, E. G. (nd), The History of Education in Cameroon, 1844-2004. University of Yaounde 1.

Fonkoua, P. & Marmoz, L. (2010). La décentralisation de l'éducation en Afrique subsaharienne: avancée et hésitations.

Fozing. I. (2009). Marche du travail en période d'ajustement structurel : Quelques résultats d'analyse du comportement des demandeurs d'emplois au Cameroun. *International symposium on Research in Education in the practices and policies of Africa*. Bamako (pp. 15-17).

Fozing. I. (2014). L'Education au Cameroun, entre crises et adjustment economique.

Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students learning. *Learning and Teaching Support in Higher Education.*1, 3-31.

Glenn K Baggott and Richard C Rayne (2007). The use of computer-based assessments in a field biology module, *Bioscience Education E-journal*, 9- 5 available at http://www.bioscience.heacademy.ac.uk/journal/vol9/beej-9- 5.aspx

Guskey, T 2002: *Professional development and teacher change. Teachers and Teaching: Theory and Practice. Vol.*8 (3/4). pp. 381–91.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. Educational Measurement: Issues and Practice, 12(3), 38-47.

Harvill, L. M. (1991). Standard error of measurement. Vol 10. Issue 2

Henning, E; Van Rensburg, M & Smit, B 2004: *Finding your way in qualitative research*. Pretoria: Van Schaik.

Henning, G. (1987). A Guide to Language Testing: Development, Evaluation, Research. Cambridge Mass: Newberry House Publisher.

https://doi.org/10.1111/j.1745-3992.1991.tb00195.x

Henning, E. Gravett,S. J.& Van Rensburg, W. (2002). *Finding your way in academic writing*. Pretoria: Vanshaik.

JISC (2009a). Effective Practice in a Digital Age.

Ji Zeng and Adam Wyse (2009). Introduction to classical test theory. Michigan Department of Education 608 W. Allegan St. Lansing, MI 48909

Kaplan, R. M. & Saccuzo, D. P. (1997). Psychological Testing: Principles, Applications and Issues. Pacific Grove: Brooks Cole Pub. Company

Kempa, R, (1986) *Assessment in Science.* Cambridge: Cambridge University Press.

Koutrouba, K., Vamvakari, M. & Steliou, M. 2006: Factors correlated with teachers'

Kongnyuy, P. (2015). Motivation and changing fortunes in teacher's output: Empirical evidence from selected secondary schools in the North West region of Cameroon. *Journal of Education and Practice*, 6(5), 111-118.

Kongnyuy, P. (2021). Teacher Motivation for quality output. *Current Research in Language, Literature and Education* Vol. 2, 66-72.

Kongnyuy, P. & Valery, W (nd). Job satisfaction as a determinant of work performance. An empirical survey of the Academic staff of the Higher Teacher Training College Bambili- Cameroon.

Krathwohl, D.R., Bloom, B.S., and Masia, B.B. (1964). Taxonomy of educational objectives: Handbook II: Affective domain. New York: David McKay Co.

Lankshear, C. & Knobel, M. (2004). *A hand book for teachers' research: From design to implementation.* London: Open University Press.

Linda, S. (2004). *Assessing Student Learning. Bolton, MA: Anker.*

Lord, F. (1952). A Theory of Test Scores (Psychometric Monograph No. 7). Richmond, VA: Psychometric Corporation http://www.psychometrika.org/journal/online/MN07.pdf

Lord, F. (1952). A theory of test scores. Psychometric Monographs, 7, x, 84.

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,* Vol.1, Issue 1. Pp. 1-11

Marcoulides, G. A. (1999). Generalizability theory: Picking up where the Rasch IRT model leaves off? In S. E. Embretson and S. L. Hershberger (Eds.), the new rules of measurement: What every psychologist and educator should know. Mahwah, NJ: Erlbaum, pp. 129-152

Marshall, C. and Rossman, G. B. (1989). *Designing qualitative research. 3rd edition*. London: Sage publishers.

Master of Sciences Thesis submitted to the Faculty of Virginia Polytechnic Institute and State University

Matlock-Hetzel, S. (1997). Basic Concepts in Item and Test Analysis. Texas A & M University, USA. files.eric.ed.gov/fulltext/ED406441.pdf. Accessed on 24 June, 2014.

Mbua, N. (2003). *Educational Administration Theory and Practice: The management of organisations and individuals*. Limbe: Design House.

McBride, N. L. (2001). An Item Response Theory Analysis of the Scales from The International Personality Item Pool and the Neo Personality Inventory-Revised.

McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum Associates International Journal for Social Studies Available at https://edupediapublications.org/journals ISSN: 2455-3220 Volume 02 Issue 09 September 2016

Melka, F. (1997). *Receptive vs. productive aspects of vocabulary.* Cambridge: Cambridge University Press.

Merriam, S.B. 1998: Case study research in education: a qualitative approach. San

Merriam, S. B. (1998). *Qualitative research and case study application in education.* Sanfrancisco: Jossey-Bass publishers.

Mertens, DM 2005: Research and Evaluation in Education and Psychology: integrating Diversity with qualitative, quantitative and Mixed Methods. Thousand Oaks CA: Sage Publication.
Garfield, L. 2016: Companies that use robots instead of humans. Business insider.com

Nenty, H. J. & Fotogang, E. B. (2014), African Journal of Theory and Practice of Educational Assesssment. Vol 1 2004 P.O. Box 30570 Biyemassi, Yaounde Cameroon.

Ngoh,N, & Mbangwana,M (2001) *Understanding Examinations: A Practical Guide to Measurements & Evaluation for Teachers.* Yaoundé: Presses Universitaires de Yaoundé.

Nkwenti Ndongfack, M.(2016b). Baseline Study on the Current State of Open and Distance Learning in Cameroon, *Commonwealth of Learning*

Nyam, N.N. & Atezah,M.M. ( 2001). *Understanding Examinations: A Practical Guide to Measurements and Evaluation for Teachers,* Yaoundé: Press Universitaires.

Rohlen, T. P. & Le Tendre, G. K. (1998). *Teaching and Learning in Japan*, Cambridge: Cambridge University Press.

Republic of Cameroon(2013a). Cameroon Education and Training Sector Strategy Paper. Published by the Ministry of Economy, Planning and Regional Development.

Republic of Cameroon, (2013a), Cameroon Education and training Sector Strategic Paper. Published by the Ministry of Economy, Planning and Regional Development.

Richard J. Torraco Encyclopedia of Life Support Systems (EOLSS) Carole J. Gallagher Educational Psychology Review, Vol. 15, No. 1 (March 2003), pp. 83-99 Springer Testing and Social Stratification in American Education

Ricketts, C. &Wilks, S. J. (2002). Improving student performance through computer- based assessment: insights from recent research. *Assessment and Evaluation in Higher Education*, Vol. 27, No. 5, pp.475-479.

Schmitt, N. & McCarthy,M. (Eds.). *Vocabulary: Description, acquisition and pedagogy.* Cambridge: university press.

Schumacker,R.E.(2010). Classical Test Analysis http://appliedmeasurementassociates.com/ama/assets/File/ CLASSICAL_TEST_ANALYSIS.pdf. Retrieved on 13 August, 2014.

Seels and Glasgow (1990). Exercises in instructional design. Columbus OH: Merrill Publishing Company University of Reading. https://www.reading.ac.uk /engage in assessment/ using-technology/eia-pros-and-cons-of-using-technology.aspx.

Shey, P. F., & Fangwi, M.L., (2020). Culturally Responsive Pedagogy: A pathway for the inclusion of diverse learners in today's contemporary classroom.

Shey, P. F., & Monju, P. M., (2019).Problem solving skills as predictor to success in Hawking Ventures of child workers in Fako Division of the South West Region of Cameroon.

Tambo, I. L. (2010). Educational Assessment in a multicultural society. In Keynote address presented at the 27th Annual conference of the Association for Educational Assessment in Africa (AEAA) in Yaounde-Cameroon.

Tambo, L. I. & Ndongko (2000), Educational Development in Cameroon 1961-1999: Issues and Perspectives.

Tambo, L. I. (2003), Principles and Methods of Teaching. Limbe. Anucam.

Tanyi, E. M. & Wirngo, T. E. (2015), Assessment Irregularities and Examination Malpractices in Tertiary Education. *African Journal of Theory and Practice of Educational Assessment* (AJTPEA), Vol. 2, 2015.

Tanyi, E. M. Theories of Learning: The process of Why, How and When. Cameroon. Africana Publication.

Tanyi, E. M. (2019), Lecture notes on Measurement and Evaluation. Faculty of Education. The University of Yaounde 1, Cameroon.

Tanyi, M. E. (2016). Pedagogic barriers in Cameroon inclusive classroom: The impact of Curriculum, Teachers' attitudes and classroom infrastructures. *Journal of education and practice*, 7(18), 210-221.

Tanyi, M. E. (2016). Psychological evaluation of attitudes of both primary teachers and special needs children towards each other in a regular school in Yaounde Cameroon. *Journal of education and practice*, 7(6), 63-73.

Tanyi, M. E. (2002).Quality Assessment in Cameroon Higher Education: Challenges and prospects. *La Revue de la Faculté des Sciences de l'éducation*, 1(1), 111-131.

Tchombe, T.M.S. (2019), Psychological Parameters in Teaching: An African perspective to learning as a process for cognitive enrichment. Design House, Limbe.

Teaching and Supporting Learning. In *National Student Forum Annual Report 2009*. pp. 12-31. Retrieved on 28 August 2018 from http://www.bis.gov.uk/assets/biscore/higher- education/docs/n/09-p83-national-student-forum-annual-report-09

The National Council of Teachers of English (NCTE)(2014). USA.

Traub, R.E., & Fisher, C.W. (1997). On The Equivalence of Constructed Responses and Multiple-Choice Tests. Applied Psychological Measurement, 1, 355370.

Zimmerman,C.B.(1997). *Historical trends in second language vocabulary inst ruction*.Cambridge: Cambridge University Press.

Zubairi, A. M., & Kassim, N. L. A. (2006). Classical and Rasch Analysis of Dichotomously Scored Reading Comprehension Test Items. Malaysian Journal of ELT Research, 2, 1-20.

# Appendices

# Appendix 1

## QUESTIONNAIRE TO BIOLOGY STUDENTS

**Dear Respondents,**

I am a doctorate student of the University of Yaounde 1, in the Faculty of Education, department of Curriculum and evaluation. I am a holder of a Bachelor Degree in education (B.Ed.) in Biology, Masters in Educational Measurement and Evaluation (M.Ed.), now researching for a Doctorate/PhD degree in Psychometrics. Here to gather data for the research.

This questionnaire serves to gather information on "***Test Structure Analysis for Creative Student Development in Biology in Secondary schools in Cameroon.***". This exercise is purely for academic purpose and a requirement in partial fulfilment of the requirements for the award of a Doctor of Philosophy (PhD) Degree in Curriculum and evaluation with a concentration on Psychometrics. Your objective responses will provide useful information on the best test Structure Analysis that will help us improve in our classrooms test practices. Try to be as frank as possible as the information you give will be treated with utmost confidentiality. I will use the data collected for research purpose only.

### Section A: Demographic Information

*Kindly fill in the blank or mark an "X" in the appropriate box:*

1.Class/Series……

2. School: □ Government □Mission □Lay private

3. Sex: □Male □Female

4. Age range: □15-20 □21-25 □26-30 □31 and above

**Instructions for sections B to F**

Please rate how strongly you agree or disagree with each of the following statements by placing a tick corresponding to one of the numbers selected from the table below.

| ***Strongly Disagree*** ***SD*** | | ***Somewhat disagree*** ***SWD*** | | | ***Somewhat agree*** ***SWA*** | | | ***Strongly agree*** ***SA*** | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

**Section B:** ***Test Item Analysis***

| **SN** | **STATEMENT** | **Strongly Disagree SD** | | **Somewhat Disagree SWD** | | | **Somewhat Agree SWA** | | | **Strongly Agree SA** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Test items are clear in the type of responses they need | | | | | | | | | | |
| 2 | The language level and visual effect of items are without complaints in a test | | | | | | | | | | |
| 3 | The items in the test are very different from the usual questions asked during class lessons | | | | | | | | | | |
| 4 | The response time for each item is not indicated in the test paper | | | | | | | | | | |
| 5 | The test items are easily comprehensible by all the test takers | | | | | | | | | | |

## Section C: *Item Distractor Analysis*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | The test items are easily understood by the test takers of the class in relation to the proposed responses/distractors | | | | | | | | | | |
| 2 | The distractors are in the same range of understanding as the right response without outright divergence of concepts | | | | | | | | | | |
| 3 | A fixed number of distractors is considered per item | | | | | | | | | | |
| 4 | Item writing flaws are ignored when choosing an item for a test | | | | | | | | | | |
| 5 | Functional and nonfunctional distractors are part of the responses | | | | | | | | | | |

## Section D: *Test validity analysis*

| SN | STATEMENT | Strongly disagree SD | | Somewhat Disagree SWD | | | Somewhat agree SWA | | | Strongly agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | The presentation of the test items is appealing to the students' zeal to take the test. | | | | | | | | | | |
| 2 | The items address mainly what the students have been taught or are expected to be taught | | | | | | | | | | |
| 3 | The test follows laid down rules and criteria provided by the examination board or laid down format of a previously validated test | | | | | | | | | | |
| 4 | The test items provoke logical reasoning and coherent outline of facts to be given | | | | | | | | | | |
| 5 | A pre-test of the instrument is carried out from which the validity coefficient is calculated and tagged to the test pack. | | | | | | | | | | |

## Section E: *Test reliability analysis*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Test items are relooked upon before using them in the test | | | | | | | | | | |
| 2 | Rejected items are thrown away or not used in the test | | | | | | | | | | |
| 3 | Items are retained from a variety of options available or from an item bank | | | | | | | | | | |
| 4 | The selected test items are arranged in an logical order as to ease comprehension and coherence in the flow of ideas | | | | | | | | | | |
| 5 | A calculated value of the reliability coefficient is tagged against each test pack | | | | | | | | | | |

## Section F: *Test grading analysis*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Tests are graded according to the type of test involved | | | | | | | | | | |
| 2 | Each item is assigned related score | | | | | | | | | | |
| 3 | Each item in a test test is accompanied by a marking guide | | | | | | | | | | |
| 4 | All test takers of the same level but different classes or teachers take the same test | | | | | | | | | | |
| 5 | The results off a test are discussed in a board or committee and success/failure level fixed in common | | | | | | | | | | |

## Section G: *Creative student development*

| SN | STATEMENT | Strongly disagree SD | | Somewhat disagree SWD | | | Somewhat agree SWA | | | Strongly agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Test items arouse curiosity in the students who feel they are responding to practical issues | | | | | | | | | | |
| 2 | Tests permit students to o experimentation of their of facts even after the test | | | | | | | | | | |
| 3 | Tests encourage students to do collective work to solve real life problems | | | | | | | | | | |
| 4 | Tests promote sustainable learning as they give room for verifiable/researchable responses | | | | | | | | | | |
| 5 | Tests serve a means of communication between the student and teachers on what they have learnt and what they can do naturally | | | | | | | | | | |

# Appendix 2

## QUESTIONNAIRE TO BIOLOGY TEACHERS

**Dear Respondents,**

I am a doctorate student of the University of Yaounde 1, in the Faculty of Education, department of Curriculum and evaluation. I am a holder of a Bachelor Degree in education (B.Ed.) in Biology, Masters in Educational Measurement and Evaluation (M.Ed.), now researching for a Doctorate/PhD degree in Psychometrics. Here to gather data for the research.

This questionnaire serves to gather information on "***Test Structure Analysis for Creative Student Development in Biology in Secondary schools in Cameroon.***". This exercise is purely for academic purpose and a requirement in partial fulfilment of the requirements for the award of a Doctor of Philosophy (PhD) Degree in Curriculum and evaluation with a concentration on Psychometrics. Your objective responses will provide useful information on the best test Structure Analysis that will help us improve in our classrooms test practices. Try to be as frank as possible as the information you give will be treated with utmost confidentiality. I will use the data collected for research purpose only.

### Section A: Demographic Information

*Kindly fill in the blank or mark an "X" in the appropriate box:*

1.Subject taught/Department……
2. Class taught…….
3. School: □ Government □Mission □Lay private
4. Sex: □Male □Female
5. Age range: □26-30 □31-35 □36-40 □41-45 □ 46-50 □51-55 □ 56-60
6. Highest academic qualification: □A/L □First degree □Others, Specify: ……

7. Years of teaching: □1-5 □6-10 □ 11-15 □16-20 □ 21-25 □ 26-30 □30 +
8. Number of years of marking the GCE Exam: □Zero □1-5 □6-10 □11+

**Instructions for sections B to F**

Please rate how strongly you agree or disagree with each of the following statements by placing a tick that corresponds to one of the numbers selected from the table below.

| ***Strongly Disagree*** ***SD*** | | ***Somewhat disagree*** ***SWD*** | | | ***Somewhat agree*** ***SWA*** | | | ***Strongly agree*** ***SA*** | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

**Section B*: Test Item Analysis (TIA)***

| **SN** | **STATEMENT** | **Strongly Disagree SD** | | **Somewhat Disagree SWD** | | | **Somewhat Agree SWA** | | | **Strongly Agree SA** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Item difficulty index is calculated before the item is validated | | | | | | | | | | |
| 2 | Item discrimination is ensured in the selection of test items | | | | | | | | | | |
| 3 | The mean score after item trial survey is considered in validating an item for a test | | | | | | | | | | |
| 4 | The response time for each item is not calculated before it is considered as part of a test | | | | | | | | | | |
| 5 | The comprehensibility of an item is ensured before using it in a test | | | | | | | | | | |

## Section C: *Item Distractor Analysis (IDA)*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | The cognitive level of the class is considered in choosing the distractors of responses | | | | | | | | | | |
| 2 | The distractor efficiency is considered in each item during its selection | | | | | | | | | | |
| 3 | A fixed number of distractors is considered per item | | | | | | | | | | |
| 4 | Item writing flaws are ignored when choosing an item for a test | | | | | | | | | | |
| 5 | Functional and nonfunctional distractors are part of the responses | | | | | | | | | | |

## Section D: *Test validity analysis (TVA)*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | The presentation of the test items is appealing to the students' zeal to take the test. | | | | | | | | | | |
| 2 | The items address mainly what the students have been taught or are expected to be taught | | | | | | | | | | |
| 3 | The test follows laid down rules and criteria provided by the examination board or laid down format of a previously validated test | | | | | | | | | | |
| 4 | The test items provoke logical reasoning and coherent outline of facts to be given | | | | | | | | | | |
| 5 | A pre-test of the instrument is carried out from which the validity coefficient is calculated and tagged to the test pack. | | | | | | | | | | |

## Section E*: Test reliability analysis (TRA)*

| SN | STATEMENT | Strongly Disagree SD | | Somewhat Disagree SWD | | | Somewhat Agree SWA | | | Strongly Agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Test items are relooked upon before using them in the test | | | | | | | | | | |
| 2 | Rejected items are thrown away or not used in the test | | | | | | | | | | |
| 3 | Items are retained from a variety of options available or from an item bank | | | | | | | | | | |
| 4 | The selected test items are arranged in an logical order as to ease comprehension and coherence in the flow of ideas | | | | | | | | | | |
| 5 | A calculated value of the reliability coefficient is tagged against each test pack | | | | | | | | | | |

## Section F*: Test grading analysis (TGA)*

| SN | STATEMENT | Strongly disagree SD | | Somewhat disagree SWD | | | Somewhat agree SWA | | | Strongly agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Tests are graded according to the type of test involved | | | | | | | | | | |
| 2 | Each item is assigned related score | | | | | | | | | | |
| 3 | Each item in a test is accompanied by a marking guide | | | | | | | | | | |
| 4 | All test takers of the same level but different classes or teachers take the same test | | | | | | | | | | |
| 5 | The results off a test are discussed in a board or committee and success/failure level fixed in common | | | | | | | | | | |

## Section G: *Creative student development (CSD)*

| SN | STATEMENT | Strongly disagree SD | | Somewhat disagree SWD | | | Somewhat agree SWA | | | Strongly agree SA | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Test items arouse curiosity in the students who feel they are responding to practical issues | | | | | | | | | | |
| 2 | Tests permit students to do experimentation of facts even after the test | | | | | | | | | | |
| 3 | Tests encourage students to do collective work to solve real life problems | | | | | | | | | | |
| 4 | Tests promote sustainable learning as they give room for verifiable/researchable responses | | | | | | | | | | |
| 5 | Tests serve a means of communication between the student and teachers on what they have learnt and what they can do naturally | | | | | | | | | | |

# Appendix 3

## Semi-Structured Interview Guide

1- What theories do you or your colleagues consider as basis for test or item development?

2- Where do you or your colleagues get items for a test?

3- When developing a multiple-choice test item, what considerations do you make in selecting the distractors?

4- What procedures do you follow in validating a particular item or a test in general?

5- How do you consider the reliability of a test as a whole?

6- Could you please enlighten me about the processes involved in test grading in your department?

7- Do you think item statistics, distractor analysis, test validity and reliability together with test grading analysis could have any correlation with creative student development in Biology?

# Appendix 4

## Observational Checklist

This instrument served the purpose of indicating the availability of various resources or infrastructure necessary to enhance the direct application of the acquired knowledge by student and teachers on or off campus, such that the development of creativity in learners could be sustained as required by the research.

| Researched elements | | Available | Not available |
|---|---|---|---|
| Item bank | -Personal<br>-Departmental<br>-All | | |
| Test bank | -Personal<br>-Departmental<br>- All | | |
| Test type | -Multiple choice<br>-Essay<br>- Structural<br>- All | | |
| Test validity coefficient | -Face<br>-Content<br>-Criterion-related<br>-Construct<br>-Translation<br>-Predictive<br>-Concurrent<br>-Convergent<br>-Discriminant | | |
| Test reliability coefficient | -Test-retest<br>-Parallel (equivalent form)<br>-Interrater<br>-Internal consistency | | |
| Test grading | -Marking guide<br>-Score sheet<br>-Marks per item | | |
| Practice site | -Laboratory<br>-Farmland<br>-Accessible land | | |
| School location | -Urban centre<br>-Rural area<br>-Fenced<br>-Unfenced | | |

# Appendix 5

## Sample test structure analysis in Ordinary level Biology

### Introduction

Test Structure Analysis provides the teachers and learners with valid and objective information on which each one should base his/her actions. It gives a vivid guide to the examiner as to the domains and competencies involved while at the same time it provides a clue to the responses of the learners; whether the responses fall in the right domain or in other different domains of the same study or out of the study.

Ngoh and Mbangwana (2001, p 52) provide as function of Test Specification that it provides content validity, shows the number of test items from each syllabus section and per cognitive objective among others.

Kempa (1986) cited in Ngoh and Mbangwana (2001: p 53) explained that "a complete specification of an examination or assessment schedule should consider a range of abilities/skills to be tested, the activities with respect to which such abilities/skills are to be demonstrated, and the subject matter areas to which the various activities relate". Thus test structure analysis has as objective; to analyse the various instruments of a test instrument,

To determine the proportion of each content domain covered in the test,

To determine the number of test competencies in each content domain,

To determine the proportion of each content domain covered in test takers' responses.

This sample has been presented under the sub headings; overview, content specification, test structure table and test specification table (or Test Blueprint).

## Biology Test (Cameroon General Certificate of Education- Ordinary Level)

### 1- Overview

The test structure analysis presented in this paper is on the ordinary level (O/L) Biology of the Cameroon General Certificate of Education (CGCE). The syllabus used is that developed in 2009 by the Cameroon General Certificate of Education Board (CGCEB).

The syllabus presents three (3) main domains as follows;

1- Nature and Variety of Living Organisms

2- Functioning of Living Organisms

3- Environmental Biology

- The syllabus ranges from the classification of organisms through their functioning to their way of life, habitats and effects to the environment.
- The testing is mainly based on the cognitive domain of learning with the Bloom's Taxonomy as the main point of reference in the setting of the questions and therefore the responses. The test at this level of education aims at testing the lower level of the taxonomy; Knowledge, Comprehension and application. The implication of the level of Analysis, synthesis and evaluation of the taxonomy may be tested but to a very low degree or percentage.
- The examination generally has two main parts or papers; the paper one comprising of objective or multiple choice questions (MCQs), and the paper two comprised of essay (constructed response) type questions.

- The paper one has sixty (60) multiple choice questions with a weight of 40 percent while the paper two has four questions with a weighting of 60 percent.
- Candidates are expected to answer all the sixty MCQs in paper one and two out of four questions in the paper two. Each question in paper one is scored one mark giving the total on sixty. On the other hand, each paper two essay question is on twenty marks, two giving forty marks. At the end, the score on sixty is brought to forty percent
- That is score/60 multiplied by 40. Then the paper two scored on forty is brought on sixty percent. That is score/40 multiplied by 60.
- The final score of a candidate is got by adding the first score on forty to the second score on sixty, having a score on 100. The candidates' scores are then graded as A, B, C, D and U. A, B, C being pass grades while D and U are failed marks.
- Though the competencies of each domain are not well outlined, they can be sought out of the domains and the recurrence of each competency helps in the weighting of the competencies.
- The number of question for each competency depends on the level at which it is found in the cognitive levels of the Bloom's Taxonomy. The higher the level of cognitive domain of a competency, the lesser the number of questions in the selected response section on that competency, and vice versa. The competencies at higher level of cognition are tested mostly in the constructed response section. Nevertheless, the test should cover all the domains of the syllabus.
- The papers one takes duration of one and half hour to write, while the paper two takes two and half hours for the candidates to complete.

## 2- Content Specification

The following content domain and their weighting are presented below:

| Domain | weighting |
|---|---|
| 1- Nature and Variety of Living Organisms | 20% |
| 2- Functioning of Living Organisms | 60% |
| 3- Environmental Biology | 20% |

The major competencies per domain are presented as follows:

### 1- Nature and Variety of Living Organisms

| Competency | weighting |
|---|---|
| 1.1 Ability to differentiate organisms by classification | 8% |
| 1.2 Ability to explain cellular functions and role | 2% |
| 1.3 Ability to appreciate some organs/organisms | 4% |
| 1.4 Ability to List out or name particular taxon/ their members | 2% |
| 1.5 Ability to illustrate links between species | 4% |

### 2- Functioning of Living Organisms

| | |
|---|---|
| 2.1 Abilities to demonstration certain experiments | 8% |
| 2.2 Abilities to appreciate certain roles of some organs | 6% |
| 2.3 Matching/associating organs and their functions | 4% |
| 2.4 Abilities to compare similar organisms and functions | 10% |
| 2.5 Ability to Relate/Relationship structures and functions | 4% |

| | |
|---|---|
| 2.6 Illustration (Draw and Label) with the aid of diagrams | 12% |
| 2.7 Stating/listing/naming of concepts and organisms | 6% |
| 2.8 Situating/Identifying specific structures of organisms | 3% |
| 2.9 Abilities to differentiate organs and their functions | 4% |
| 2.10 Explanation of certain functions | 3% |

**3- Environmental Biology**

| | |
|---|---|
| 3.1 Defining of terms and concepts | 5% |
| 3.2 Illustrating of relationships and impact | 12% |
| 3.3 Abilities of explaining certain notions | 3% |

# 3-Test Structure Table (O/L Biology Test)

Table 1

*Test Structure Table on Biology Ordinary Level*

| Content Specifications Domain | Number of Competencies | Approximate Weighting | Approximate number of Multiple Choice Items | Number and type of constructed-response items | |
|---|---|---|---|---|---|
| | | | | Focused educational problem and instructional tasks | Case study |
| **Domain 1** Nature and variety of living organisms | 5 | 20% | 15 | 1(illustrative and diagrammatic presentation of responses) | 0 |
| **Domain 2** Functioning of living organisms | 10 | 60% | 35 | 2(experimental presentations, balanced equations and arrow diagrams) | |
| **Domain 3** Environmental Biology and human activities | 3 | 20% | 10 | 1(Problem solving approach to natural and man-made environment) | |
| **Total** | 18 | 100% | 60 | 4 | 0 |

# 4- Table of specification (Test Blueprint)

Table 2

*Test Specification Table (or Test Blueprint)*

| **Content** | | | **Cognitive Levels** | | | |
|---|---|---|---|---|---|---|
| **Domain** | **Competencies** | **Weight** | **Remember 45%** | **Understand 30%** | **Apply/Analyse/Evaluate/create 25%** | **Total Items** |
| 1 | 1.1 | 8% | 1 | 0 | 2 | 3 |
| | 1.2 | 2% | 1 | 0 | 1 | 2 |
| | 1.3 | 4% | 1 | 1 | 0 | 2 |
| | 1.4 | 2% | 1 | 2 | 3 | 6 |
| | 1.5 | 4% | 1 | 1 | 0 | 2 |
| 2 | 2.1 | 8% | 1 | 1 | 0 | 2 |
| | 2.2 | 6% | 1 | 1 | 0 | 2 |
| | 2.3 | 4% | 3 | 1 | 1 | 5 |
| | 2.4 | 10% | 1 | 1 | 0 | 2 |
| | 2.5 | 4% | 1 | 1 | 0 | 2 |
| | 2.6 | 12% | 0 | 1 | 1 | 2 |
| | 2.7 | 6% | 5 | 2 | 1 | 8 |
| | 2.8 | 3% | 4 | 1 | 1 | 6 |
| | 2.9 | 4% | 2 | 1 | 0 | 3 |
| | 2.10 | 3% | 2 | 1 | 0 | 3 |
| 3 | 3.1 | 5% | 1 | 1 | 1 | 3 |
| | 3.2 | 12% | 0 | 1 | 4 | 5 |
| | 3.3 | 3% | 1 | 1 | 0 | 2 |
| **Total** | 18 | 100% | 27 | 18 | 15 | 60 |

## Sample Biology Test from the test structure analysis in this study

*Place a tick (√) against the letter (A or B or C) corresponding to the right answer to the given description or Question from 1 to 20. (1x20Marks)*

*C1.1 (3 items)*

1. An example of an amphibian
   - A. gecko
   - B. fish
   - C. frog

2. Non pigmented plants-like growing organism
   - D. fern
   - E. fungus
   - F. spirogyra

3. A class of organism related to human by the manner of birth
   - D. mammalia
   - E. hummalia
   - F. humminidae

*C1.2 (2 items)*

4. Small pores in the epidermis of the leaf used for gaseous exchange
   - D. flaccids
   - E. porous
   - F. stomata

5. Components of air send out by plants during the day
   - D. $CO_2$
   - E. $O_2$
   - F. CHO

*C1.3*

6. A vital organ for gaseous exchange in man
   - D. lung
   - E. bronchus
   - F. nostrils

*C1.4*

7. Organism that feed on dead organic matter
   - D. parasites
   - E. holozoits
   - F. saprophytes

*C1.5*

8. One reason for classifying rat and human in the same taxon
   - D. they give birth to live young ones
   - E. they use their teeth to eat
   - F. their female care for the young

*C2.1 (3 items); Questions 9, 10 and 11are based on experiments on photosynthesis*

9. To prove the presence of starch in a leaf the colour changes to
   - D. brown
   - E. pure white
   - F. dark blue

10. A requirement to test if sunlight is necessary for photosynthesis
    D. aluminum foil
    E. cardboard paper
    F. polythene bag
11. A substance not use in the experiment to show that carbohydrate is necessary for photosynthesis
    D. iodine
    E. caustic soda
    F. ethanol

*C2.2 (2 items)*

12. An organ specialized in the removal of waste products of metabolism from the human body
    D. Skin
    E. Kidney
    F. Anus
13. The movement of plant food made In the leaves to other parts of the plats is known as
    D. Translocation
    E. Transportation
    F. Transpiration

*C2.4*

14. This organism carries out respiration through the gills.
    A. Tadpole
    B. Frog
    C. Toad

*C2.5*

15. This organism has an elaborate dentition for shearing of its prey for food.
    A. snake
    B. crocodile
    C. monkey

*C2.7*

16. A reptile adapted to camouflage in its environment.
    A. lizard
    B. earthworm
    C. chameleon

*C2.8*

17. Movement by this organism is done on eight legs.
    A. cockroach
    B. spider
    C. bee

*C2.10*

18. A pair of tactile organs for sensory perception in insects.
    A. nerves
    B. compound eyes
    C. antennae

*C3.3*

19. To naturally enrich the soil to boost plants growth, farmers are advised to apply
    A. fertilizer

B. wood ash
C. manure

*C3.2*

20. In a food chain relationship, the spider feeds on the housefly. In this case the spider is called the

A. predator
B. victor
C. prey

## Conclusion

As a continuation of the test structure analysis, the test items were developed for a classroom Biology test. The test items were supposed to be pilot tested and administered, then carry out analysis to determine which of the domains/ competencies were covered by the candidates' responses, whether they were representative or not. But for the purpose of this research, and following the limitations of the study, the researcher has ended at the level of the classroom test.

# Appendix 6

# Research Authorisation

# Appendix 7

## Endorsed Applications to access Classrooms in the various Schools of the selected regions

# Appendix 8

## The Krejcie and Morgan Table of populations and samples

# Appendix 9

## Cameroon map showing the four cultural zones


**Source:** *Google maps; labeled by the author*

# Appendix 10

# Conference attestations and certificates