

## ASSESSMENT PROCEDURES USED IN ART AND DESIGN MAIN STUDY AT TEACHER EDUCATION LEVEL IN ZIMBABWE: VALIDITY AND RELIABILITY.

**Dairai Darlington Dziwa**

Curriculum Studies Department, Faculty of Education  
Great Zimbabwe University, Masvingo,  
ZIMBABWE.

[dairaidziwa@yahoo.co.uk](mailto:dairaidziwa@yahoo.co.uk)

### ABSTRACT

*An assessment of an assessment' is done here to ascertain validity and reliability of measuring the capabilities of Art and Design learner teachers in Zimbabwe. The study sought to evaluate the worthiness, quality and effectiveness of the assessment procedures being used for Art and Design as a Main Study subject at teacher education level. Participating teachers' colleges for the survey were drawn from Masvingo province. The sample population was chosen using stratified random procedure. Thirty (30) Art and Design main study students and nine (9) Art and Design lecturers participated. Inspirations from both qualitative and quantitative designs were employed in the collection and analysis of data. Interviews, Document analysis and questionnaires were the chief data collecting methods used. The study revealed that the assessment procedures used at the time of research had a propensity of bias, inconsistency and subjectivity. The assessment also lacks content validity. The study has therefore recommended in-service training of lecturers and adoption of other assessment strategies for this practical subject.*

**Key words:** Assessment, Art and Design, Reliability and Validity

### INTRODUCTION

Assessment is an integral component of any learning programme without which no reflection is made on the mastery and competency of the learners. Assessment is inevitable in any learning programme for screening, diagnostic, counseling and evaluation purpose (Sattler, 1992). It is imperative therefore that assessment has been for a very long time and is still being carried in Art and Design at any level of education. This paper evaluates and unearths how assessment is done in Art and Design main study course at teacher education level in Zimbabwe.

### BACKGROUND

In 1977 the primary teacher education syllabus and regulations in the then Rhodesia proposed for the use of a continuous assessment in main study assessment at teacher education level. It further suggested that no formal final examination had to be set for assessment purposes. Therefore all colleges from that time had to adhere to these nationally set regulations of using a continuous assessment mode. On-going or continuous assessment is key to improving quality but however not very effective to ascertain level of cumulative proficiency at the end of a course.

In light of such shortfalls, the Teacher Education Review Committee (TERC) 1986 chaired by Dr Siyakwazi proposed new recommendations on the assessment of main study areas in teachers colleges. The committee clearly stated under Art and Craft syllabus guidelines, item 6.1, "assessment of students' work will be based on coursework and end-of course

examinations”. The committee further reiterates that ‘students were required to complete satisfactorily both the course work and the examinations in order to have passed the course’ (page 93, item 21.1). This then implied the use of continuous (formative) and summative modes of assessments. However the report was silent on the tools and methods that were to be used in assessing the various subjects. Therefore the issues of criterion and measurement tools were left to the colleges to decide. For instance Gwanda ZINTEC main study syllabus 1999 states that assessment is continuous based on assignments works. While Masvingo Teachers’ college 2001 states that assessment is both continuous and summative basing on a portfolio and a final examination. Therefore the assessment of Art and Design at teacher education was not uniform throughout Zimbabwe, though all colleges were and are associates of the University of Zimbabwe. Some colleges based assessment on one essay, one test and selected course work artifacts without a final exam, while other colleges depended on a folder or portfolio with artifacts produced during the three years together with written assignment and final examination.

All teachers colleges in Zimbabwe private or state owned are associates of the University of Zimbabwe through the scheme of association under the Department of Teacher Education. The Department of Teacher Education Handbook DTEH (2006), states that assessment should be a reflection of final achievement only. There are some assignments that are administered for developmental purposes only and should not be included for final assessment. Thus DTEH (2006) stipulates that only marks of assignments deliberately designed to measure the final achievement in certain intellectual skills or in certain areas of content should be used. The Handbook further suggests that assessment can be done using instruments such as written examination, selected course work, projects, practical works and tests. The college syllabus should be precise to state the structure and nature of the written examination on the tests used in measuring achievement of learning goals.

These variations and influences in the assessment of Art and Design in Zimbabwean teacher education therefore posed a lot of challenges on identifying which precise knowledge domains were important for a learner teacher in Art and Design. Which range of skills and content knowledge is satisfactorily comprehensive to ascertain competency in an Art and Design learner teacher? How accurate is the marking, scoring or grading done to practical work? Is the assessment at teacher education effective in helping students identify weaknesses and strengths and ways in which they can improve?

## CONCEPTUAL FRAMEWORK

Arends (1988) define assessment as the process of gathering information about students’ performance through observation, interviews, homework, tests and written reports for use in making judgments. Assessment in Art and Design has taken many forms including projects, production of artifacts, pencil and paper examinations and practical examinations.

The assessment of college learning to date can be traced to have evolved through four eras, firstly the origin of standardized tests of learning, secondly, the assessment of learning for general and graduate education, thirdly, the rise of test providers and lastly the era of external accountability (Shavelson, 2007). While these eras are distinct and traceable in the developments in the western world, it is not a replicable situation the world over. However assessment at whatever level or place has taken similar strides in focus and intent to what it is today.

The first third of the twentieth century (1900-1933) marked the beginning of the use of standardized, objective testing to measure learning in higher education in America (Shavelson, 2007). The Carnegie Foundation in 1916, tested students in arithmetic, spelling,

penmanship, reading, and English composition using recognized tests, procedures, and scales, and a statistical treatment that though comparatively crude was indicative” (Savage 1953:284). E. L. Thorndike’s study of university students tested students on all or parts of several objective tests (multiple tests) in mathematics, English, physics and the arts (Savage 1953:285). These tests focused on content knowledge, largely tapping facts and concepts (declarative knowledge) and mathematical routines (procedural knowledge). The early tests were “objective”; students responded by selecting an answer where one answer was correct. Compared to the widely used essay examination, these tests gained reliability in scoring and content coverage per unit of time.

This study is noteworthy because it laid out a conception of what was meant by undergraduate achievement and learning, assuming that achievement was the result of college learning defined as the accumulation of breadth and depth of content knowledge. It also focused heavily and comprehensively at the knowledge level, especially on declarative and procedural knowledge. Mpofo (1991), observed that most of the instruments used for assessment lack content and construct validity. The pencil- paper test tends to reveal only whether the students can recognise, recall or plug in what has been learnt out of context. Validity has been defined by Kubiszyn & Borich (2003) as the accuracy in which a test measures what it is suppose to measure. There are three forms of validity evidenced namely content, construct and concurrent validity.

The Assessment of Learning in General and Graduate Education (1933–1947) emerged with its focus on the whole student. The attempt was to measure not only cognitive outcomes across the spectrum but also the personal, social, and moral outcomes of general education. Assessment at this stage diverged from rather than adopted the Carnegie Foundation’s view of education and learning assessment. University of Chicago College program reflected thinking in the progressive era. In the Chicago program, a central university examiner’s office, rather than individual faculty in their courses, was responsible for developing, administering, and scoring tests of student achievement in the university’s general education program (Frodin 1950). Whereas the Pennsylvania Study assessed declarative and procedural knowledge, the Chicago examinations tested a much broader range of knowledge and abilities: the use of knowledge in a variety of unfamiliar situations; the ability to apply principles to explain phenomenon; and the ability to predict outcomes, determine courses of action, and interpret works of art (Shavelson; 2007). The Chicago comprehensive exams were characterized by open-ended essays and multiple-choice questions demanding interpretation, synthesis, and application of new texts. The year 1949 saw a landmark in student learning assessment, in a shift from testing content to testing general reasoning in effectiveness of expression

In 1959 tests were provided by Testing Organizations to assess student learning. These 1960s and 1970s testing programs included ETS’s Undergraduate Assessment Program, which incorporated the Graduate Record of Examination GRE, and American College Testing ACT’s College Outcomes Measures Project (COMP). The GRE evolved into today’s Measure of Academic Proficiency and Progress (MAPP), and the latter, (COMP) evolved into today’s College Assessment of Academic Proficiency (CAAP). However, several developments in the late 1970s, reminiscent of the progressive era, augured for a change. They wanted to get at broader abilities such as the ability to communicate, think analytically, and solve problems in a holistic manner. Tests tapped communication skills, analytic thinking, synthesizing ability, and social/cultural awareness (Warren 1978).

American College Testing (ACT) experimented with open-ended performance-based assessments in 1959 that sought to measure skills for effective functioning in adult life in

social institutions, in using science and technology, and in using the arts. For a short period, these learning assessments set the mold. But due to time and cost limitations, as well as difficulties in securing and training people to score responses and in achieving adequate reliability, it was discontinued. Omrod (2000) in Shavelson (2007) suggests that the reliability of an assessment tool such as a pencil-paper test or essay test or practical examination is the extent to which it yields consistent information about knowledge, skills or other abilities being measured. This can be tested through test-retest, alternative testing and checking internal consistency

The performance based tests were costly in terms of time and scoring. The quantification of the levels of skill and content mastery measured from performance based tests in assessment can be problematic (Kubiszyn and Borich; 2003). A study carried out by Mamvuto in 2002 at Hillside Teachers college in Bulawayo, Zimbabwe showed that even when marking was based on objective qualities inherent in any artwork such as elements and principles of design, the end result tilted towards subjectivity. Mamvuto (2002) noted that two subsequent marking of same piece of work done at reasonable interval yield surprisingly different results or grades.

There are basically two assessment approaches used, these are formative and summative. Castiglione (1996) identifies four assessment techniques that can be used in any one of the approaches stated above. These are Criterion referenced, norm- referenced, portfolio and performance based assessment. In Criterion referenced assessment, the student's level of assessment is judged in relation to specific body of knowledge or list of agreed objectives. Norm referenced assessment the student's performance is judged in relation to the performance of peers. Performance-based assessment requires a student to demonstrate in a practical situation to show mastery of a skill (Tacit knowledge). Portfolio assessment is based on a collection of students work, drafts, sketches, notes and self-reflective notes and the final product over a period of time. The process and product are central parts in arts assessment.

By the end of the 70's in America objective testing was incompatible with the way practical subjects lecturers either assessed student learning or wanted student learning to be assessed. For them, life is not a multiple-choice test. Rather, lecturers like the open-ended, holistic, problem-based assessments exemplified by, for example, Tasks in Critical Thinking. Intuitively, they suspected that the kind of thinking stimulated and performance assessed by multiple-choice and other highly structured tests is different from that stimulated and assessed by more open-ended tasks. While a multiple-choice test and a "constructed-response" test may produce scores that are correlated with each other, this correlation does not mean that the same kind of thinking and reasoning is involved (Martinez, 1999; National Research Council; 2001). Student performance varies considerably depending upon whether a task is presented as a multiple-choice question, an open-ended question, or a concrete performance task (Baxter and Shavelson 1994). For example, Lythcott (1990: 248) in Shavelson (2007) found that "it is possible, though not our intention, for [high school and college] students to produce right answers to art problems without really understanding much of the creativity involved." Performance-based assessment requires a student to demonstrate in a practical situation to show mastery of a skill (Tacit knowledge). Portfolio assessment is based on a collection of students work, drafts, sketches, notes and self-reflective notes and the final product over a period of time. The process and product are central parts in arts assessment.

Wiggins & McTighe (2001), define a portfolio as a "representative collection of one's work." A portfolio communicates what a student has learnt. Portfolio assessment is in two forms,

summative and formative. When a portfolio is summative, it is an overview of what a student accomplished. A summative portfolio will typically feature the student artist's finest work. Formative portfolios recognize that the artistic process is as valuable as the artistic product. When a Formative portfolio assessment allows a student to present the development of their work such as the process folio, developed by Arts PROPEL at Project Zero (Gardner; 1987).

## RESEARCH QUESTIONS

The study was therefore guided by the following research questions;

1. Are the precise Art and Design learning outcomes measured and observed in the assessment approaches used in Zimbabwe teacher education?
2. Which range of skills and content knowledge is satisfactorily comprehensive to ascertain competency in a Art and Design learner teacher?
3. How accurate and objective is the marking, scoring or grading done to practical work?
4. How reliable and valid are the lecturer-made tests used in the assessment of Art and Design at teacher education level?

## METHODOLOGY

Qualitative methodology was used in his study especially the descriptive case study. Borg and Gall, (1996) suggests that case studies are done to produce detailed descriptions of phenomena, to develop possible explanations of it or to evaluate a phenomena. The study made judgments, merit, value or worthiness of the assessment procedures being used in teachers colleges. The study targeted Art and Design Main study students at teachers colleges in Zimbabwe. Teachers colleges in Masvingo province were sampled purposively as they offer Art and Design as a main study subject and their proximity to the researcher. Stratified random sampling was used to select participants who were mostly likely to be knowledgeable about the phenomena under study. The final year students and their lecturers were the most appropriate population to provide the assessment information. Thirty (30) Art and Design main study students and nine (9) Art and Design lecturers participated in the research study. Inspirations from both qualitative and quantitative designs were employed in the collection and analysis of data. Interviews, Document analysis and questionnaires were the chief data collecting methods used. Observations were also done to validate the participants' verbal claims. However participants often pretend behavior whenever they are aware that they are under observation (Borg and Gall, 1996) the researcher collected data as self no research assistants were used.

## RESULTS

Document analysis of syllabus and examination papers was conducted to investigate content validity. Two past examination papers administered consecutively were sampled from each college. The frequency of the examination questions in satisfying the demands of the learning outcomes and objectives of the Art and Design main study syllabus was analysed. These learning outcomes according to Bloom's Taxonomy in Anderson & Krathwohl, (2001) are Knowledge (explicit/codified), application, analysis, Synthesis and evaluation. Learning outcomes help focus the assessment on the right criteria. The question indicators stated in Table 1 below were used for the document analysis purpose. The indicators were used to categorise the learning outcomes tested by each examination question. Tallies were made for each question in the examination paper against the learning outcomes tested and the frequencies in Table 2 were obtained. Three primary teacher training colleges participated, Masvingo (X), Bondolfi (Y) and Morgenster (Z) teachers colleges.

**Table 1: Learning outcomes and Indicators**

<i>Learning outcomes</i>	<i>Indicators in a question</i>
Knowledge	Name, Identify, Describe, Label, List, Mention, State
Application	Determine, Relate, Solve, Find
Analysis	Classify, Sequence, Categorise, and Demonstrate, Contrast.
Synthesis	Illustrate, Design, Formulate, Modify, Reconstruct, Simplify.
evaluation	Justify, Compare, Evaluate, Defend, upraise

**Table 2: The Frequency of Learning Outcomes Tested in the theory Examination**

<i>Learning outcomes</i>	<i>Frequency in examination papers per college</i>					
	X		Y		Z	
	2009	2010	2009	2010	2009	2010
<i>Knowledge</i>	26(72%)	25(74%)	27(79%)	21(88%)	42(84%)	45(76%)
<i>Application</i>	2(6%)	3(8.5%)	1(3%)	0	0	5(8%)
<i>Analysis</i>	4(11%)	0	2(6%)	1(4%)	8(16%)	9(16%)
<i>Synthesis</i>	0	3(8.5%)	2(6%)	1(4%)	0	0
<i>Evaluation</i>	4(11%)	3(8.5%)	2(6%)	1(4%)	0	0
<i>Total</i>	36	34	34	24	50	59

An average of 79% of the questions included in all the examination papers presented in Table 2 above requires learners to reproduce content taught. This promotes rote learning. Very few questions demand learners to show application, analysis, synthesis and evaluation and yet all these learning outcomes form the core of the syllabus expectations.

### Frequency of Syllabus Content Examined

There are twelve (12) topics in the Art and Design main study syllabus from all the colleges studied. The sampled examination papers were analysed to investigate content by tallying the frequency of each syllabus topic in the examinations.

Data collected and presented above in Table 3 shows the frequency by which Art and Design main study syllabus content is tested in the examination per topic. Painting is the most frequently tested topic (41%) in both practical and theory examinations in all the colleges studied. Some topics are occasionally left out for examination purposes such as photography, printmaking, sculpture and the so called “Art theory”.

Results also show that at college Z in 2009 have 4.7% of the questions on Child art and another 1.5 % on curriculum issues of planning and scheming which are not in the main study syllabus. College Y also in 2009 set 6% of the questions on a topic “Child Art” which was not in the syllabus.

**Table 3: Frequency of syllabus content examined**

Syllabus content	Frequency per college											
	X		Y				Z					
	Practical (%)		Theory (%)		Practical (%)		Theory (%)		Practical (%)		Theory (%)	
	2009	2010	2009	2010	2009	2010	2009	2010	2009	2010	2009	2010
Drawing	2(40)	2(66)	4(10)	5(14)	3(33)	1(11)	3(9)	2(8)	2(25)	2(25)	10(20)	9(15)
Painting	2(40)		7(19)	7(21)	3(33)	6(66)	9(27)	2(8)	3(36)	3(36)	11(22)	12(20)
Printmaking								2(8)	1(13)	1(13)	3(6)	6(10)
Textiles			5(13)	3(6)	1(11)		1(3)	3(12)	1(13)	1(13)	4(8)	5(8)
Graphic Design	1(20)	1(33)	3(8)	3(9)	2(22)	2(22)	3(9)	5(21)	1(13)	1(13)	4(8)	9(15)
3 D Studies			6(18)	4(12)			2(6)	2(8)			6(12)	4(7)
Sculpture			3(8)				2(6)				3(6)	2(3)
Photography									1(4)			
Zimbabwean Art			5(13)	2(6)			3(9)	1(4)				
Art Movements			2(6)	8(24)			3(9)	3(12)			8(16)	7(12)
African art			1(3)	1(3)			3(9)	1(4)				1(2)
Art Theory							4(12)	2(8)				4(7)
Total	5	3	36	34	9		34	24	8	8	50	59

### Content validity

The ability of the examinations to measure what they are suppose to measure was analysed quantitatively basing on the frequency of syllabus content presented above. The Standard Deviation (Sd) as a measure of variability was calculated to determine distribution of syllabus content in the examination papers and hence infer content validity. The formulae:  $Sd = \frac{\sqrt{(x-x)^2}}{N}$  was used to calculate the standard deviation. The s.d of all the examination papers is very high. A high Sd indicates a greater score variability. This implies that the syllabus content is not given equal or uniform propotion in the testing of Art and Design main study syllabus content mastery. Therefore it implies that content validity of the examinations is very low.

### Reliability

Alternate reliability has been used to estimate the ability of the examinations administered to measure the expected syllabus content outcomes consistently by the two forms of examination papers sampled from each college. The Spearman–Brown prediction formula, also known as the Spearman–Brown prophecy formula was used to predict psychometric reliability of the final examinations of the three colleges. The reliability coefficient of 0.00 expressed in 2 decimal places was obtained from the three colleges. Thus the exams are not reliable instruments to accurately and consistently measure students’ mastery of content

outcomes expected from the colleges' syllabuses, they vary in length and content year after year.

### Standardisation

Data collected through questionnaires indicated that generally students agree that there is standardization in the assessment process. Students are all notified of tests on time, given adequate and equal material resources, as well as equal working conditions in examinations. The majority 86% however expressed dissatisfaction in using coursework artworks for final assessment purposes arguing that they are created under different working conditions. Coursework practical assignments are assigned as homework assignments done without lecturer supervision therefore other students are aided by experts at home while others are not. Thus different performance levels and competencies exhibited reflect a direct result of the different working conditions created and not the learner teachers' performance.

### Moderation

The moderation of examinations is done at three levels. Firstly, when the examination and scoring rubrics are set. The question paper and the marking guides are moderated by experts appointed by the Department of Teacher Education (DTE) University of Zimbabwe through the scheme of association. This should ideally enhance the validity and reliability of the examination at the preparation stage. Secondly there is moderation of marked scripts or assignments by fellow lecturers in the same departments. All distinctive and failing scripts are moderated.

At the third level, external examiners appointed by the DTE moderate all the course and final examination work presented for final examination.

### Assessment methods

The study revealed that all the lecturers are familiar with four assessment methods which are, Criterion-Referenced, Norm Referenced, Performance- Based and Portfolio assessment methods.

**Table 4: Assessment weighting**

<i>college</i>	<i>Course work (on-going)</i>	<i>Examination (Terminal)</i>	<i>Total</i>
X	60%	40%	100%
Y	50%	50%	100%
Z	60%	40%	100%

The majority 66.6% is very familiar to portfolio and performance-based assessment methods. Portfolio assessment is the most commonly used but misused in all the three colleges studied. Students make a portfolio of all the artworks done throughout the three years and this forms part of the final assessment mark. Results have shown that though the lecturers claim to use performance based assessment during the terminal practical examinations it's not effectively done. There is no scoring done to the process but rather only the end product. Both the course work mark and the final examination marks are used for final assessment with varying weighting as shown in the Table 4 below.

## SUMMARY OF FINDINGS AND CONCLUSIONS

The tests used do not examine all and only those abilities that they are designed to access. Some questions included in the exams are not derived from the syllabus content while some



content from the syllabus are excluded. Eisner (2010) reiterates that *'not everything that matters can be measured, and not everything that is measured matters.'* The lecture – made tests used in the terminal assessment of Art and Design main study lack content validity.

Findings from the study also revealed that the examinations used in the art and Design in all the colleges studied have little probability of providing same results when measuring Art and Design competencies or mastery of Art and Design content. The test papers sampled showed that the question items set year after year are not consistent with the syllabus content and objectives to measure the same skills each year. The large differences in the frequency of question items on the two forms of examination papers sampled, which should measure the same competencies on the Art and Design teacher, indicates an unreliable test. Thus alternate reliability is very low.

The precise expected Art and Design learning outcomes are hardly measured in the assessment approaches used in teacher education. There is over emphasis on the domain of explicit / codified knowledge retention that promotes rote learning at the expense of other domains pertinent to Art and Design as a practical subject. Domains like evaluation, synthesis, analysis and application are hardly tested.

Findings also show that there is standardization in the administration of final examinations but however it is not reflected during coursework assignments. There is inconsistency in the working conditions created.

The lecturers as the assessors are not very versatile and knowledgeable about the precise psychometric procedures suitable for Art and design. Some of these lecturers have indicated lack of knowledge on the performance based assessment methods which are central to art and Design.

## RECOMMENDATIONS

In light of the findings above, recommendations are made to improve the assessment practices of Art and Design at teacher education level in Zimbabwe.

1. First and Foremost the competencies of the assessors (the lecturers) to effectively prepare syllabus, set examinations, mark and moderate should be improved by means of workshops or in- service or seminars. Competencies of the assessors are central in the whole assessment process.
2. Standardization can be improved during coursework assessment by using same artwork for assessment.
3. Panel marking of art works can reduce bias and subjectivity.
4. The construction of examination paper should go under serious scrutiny before being administered. That is the paper construction can go through several stages which can aid modification for a valid paper. These are item writing, editing, trying out (piloting) and re writing. Item difficulty and discrimination should also be analysed at the construction stage hence content validity is improved. This process if done every year for every examination paper then reliability is improved.
5. The study also recommends the adoption of portfolio assessments. This includes regular formative assessments done during the development of the portfolio by the lecturers. Self-assessment is a necessary and important component of quality classroom portfolio arts assessment. Self-assessment is authentic assessment; it mirrors the “real-world” work done by professional artists. Self-assessment

contributes to student artistic autonomy (Hale & Green, 2009). Students' self-assessments can take many forms and are typically formative assessment. Journals, recordings, sketches, diaries and discussions are a few of the available tools. These document the procedures taken, challenges and strength achieved.

6. The study recommends that further study be done to establish the extent to which portfolio assessment can be used at teacher education level. Does portfolio assessment measure effectively competences of learner teachers in Art and Design given the few residential periods which now characterize teacher education in the 2-5-2 model?

## CONCLUSION

The assessment of art and Teacher education level is done has gone through several changes and innovations throughout the ages as highlighted in this paper. There are infinite opportunities to precisely assess art and design learner teacher's knowledge interests if there are constant evaluations and introspections to the contemporary practices.

## REFERENCES

- Anderson, L. & Krathwohl, D. A. (2001). *Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman
- Arends, R. (1998). *Learning to Teach*: 4th Ed, Singapore: McGraw- Hill.
- Baxter, G. P. & Shavelson, R. J. (1994). Science performance assessments: Benchmarks and surrogates. *International Journal of Educational Research* 21 (3): 279–98.
- Borg, W. R. & Gall, M. P. (1986). 4<sup>th</sup> ed. *Educational Research: An Introduction*; Auckland: Longman.
- Boyle, J. D. & Radocy, R. E. (1987). *Measurement and Evaluation in Musical Experiences*, London: Collier Macmillan Publishers.
- Castiglione, L.V. (1996). Portfolio Assessment in Art and Education; *Arts Education Policy Review*; 97(4); 2-9.
- Department of Teacher Education Handbook. (2006) *A Handbook on Teacher Education*. U.Z.
- Frith, E. W. & Macintosh, H.G. (1991). *A Teachers Guide to Assessment*. Leckhampton: Stanley Thrones Publishers limited.
- Frodin, R. (1950). Very simple but thoroughgoing. In *The idea and practice of general education: An account of the College of the University of Chicago by present and former members of the faculty*, ed. F. Ward, 25–99. Chicago: The University of Chicago Press.
- Gardner, H. (1989). Zero-Based Art Education: An Introduction to ARTS PROPEL. *Journal of Issues and Research*, 30(2): 71-83
- Gronlund, N. E (1981). *Measurement and Evaluation in Teaching*, New York: Macmillan,
- Hale & Green, 2009
- Kubisyn, T. & Borich, G. (1993). *Educational Test and Measurement: Classroom Application and Practice*, New York: Harper Collins College,

- Mamvuto, A. (2000). Involving Students in the Evaluation of Their Own Creative Art Studio Productions Through Mutual Evaluation, *The Zimbabwe Bulletin Of Teacher Education*, 111(2), 1-13.
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist* 34: 207–18.
- Mpofu, E. (1991). *Testing for Teaching*. Harare: Longman.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Sattler, J. M. (1992) . *Assessment of Children*. San Diego: Jerom Sattler Publishers Inc.
- Shavelson, R.J. (2007). *A Brief History of Student Learning Assessment*. Association of American Colleges and Universities. Washington: Teacher Education Review Committee (TERC) Report (2006).
- Warren, J. R. (1978). *The Measurement of Academic Competence*. Princeton, NJ: Educational Testing Service.
- Wiggins, G. & McTighe , J. (2006). 2<sup>nd</sup> ed. *Understanding by Design by Association for Supervision and Curriculum Development Alexandria, VA*.