

Full Length Research Paper

Evaluating the Influence of Assessment Criteria on Medical Students' Performance in Anatomy Exams

L. A. J SHITTU¹, M. P. ZACHARIAH², O. A. ADESANYA³, M. C. IZEGBU⁴, and O. A. ASHIRU¹

¹Department Of Anatomy, ²Department Of Psychiatry, and ⁴Department Of Morbid Anatomy, Lagos State University, College Of Medicine (LASUCOM), PMB 21266 Ikeja, Lagos, Nigeria. ³Department Of Anatomy, Igbinedion University, Okada, Edo State, Nigeria.

Accepted 29 September, 2024

This study was designed to assess the convergent validity of the professional anatomy (a multifaceted examination) with other markers of success (the various interactive assessment components of the curriculum) in determining the overall performance of third-year medical students. The aim was to isolate area of academic weakness among the students and to readjust the curriculum content to balance the weakness. A total of 66 third year medical students with records of their grades in the various assessments criteria were analyzed. Parameters on the average End-in course assessment, Short Essays Question (SEQ), Multiple-Choice Questions (MCQ), and Practical (Steeple-chase) were considered. The Practical significantly correlated with overall performance ($r = 0.89$, $P < 0.01$). The Practical, MCQ, SEQ and End incourse showed an overall rank order of relative performance in assessment tasks and therefore indicate that, in general, students performance in the Professional examination was better than in the End-incourse examination.

Key words: Assessment, curriculum, Professional examination.

INTRODUCTION

The review of the curriculum and the formulation of curriculum is an ongoing process in several medical schools all over the world (Ashiru et al., 1984). In Nigeria of today, medical education plays a vital role in nation building (Ashiru, 1977). Since 1990, there has been increasing attention to assessment in higher education, with various attempts to inform the professional discourse of assessment in the sector (James and Fleming, 2004). Some have been concerned with the philosophy of assessment and of assessment practice (Miller et al., 1998; Swann and Ecclestone, 1999), while others have focused on promoting good practice (Baume and Baume, 1992; Brown, 2001) and the application of specific examples (Habeshaw et al., 1993; Race, 1995, 1996). Assessment is one of the most powerful drivers of innovation and change in education, as it defines the goals for both learners

and teachers (DfES, 2003). It has also been claimed that student learning is assessment-driven (Habeshaw et al., 1993), and even that assessment is of singular importance to the student experience (Rust, 2002).

The curricular model of LASUCOM has different structural phases (basic and clinical sciences) similar to what is obtained in other medical schools. In LASUCOM, the duration of the basic medical science courses span for an 18-month period with Anatomy having 35% of the total time allotted (486 hours out of 1406 hours with 200 hours in first semester and 234 hours in 2nd and 3rd semester respectively). According to Kolesnikov (2002), it is quite justified that so much time is spent on anatomy. He believed that anatomy of man is one of the basic educational medico-biological disciplines in higher medical formation that medical students first meet in their career. This is contrary to the view of Moxham (2002). One of the innovations in this new curriculum being implemented in the LASUCOM basic sciences program is the introduction of incourse assessment which forms part of the continuous examination exercises implicit in the curriculum.

*Corresponding authors E-mail: drlukemanjoseph@yahoo.com

riculum. Moreover, it ensures continuous study on the part of the students. However, the effectiveness of any educational innovation in itself remains undetermined unless evaluative research accompanies curriculum change (Walton, 1967). More recently, qualitative studies have emphasized the importance of understanding the way students respond to innovations in assessment (Sambell and McDowell, 1998).

Internationally, assessment is changing as the nature of teaching and learning in post-compulsory education changes (Brown, 2004). The research for an ideal method of evaluating medical students and post graduates has undergone considerable change. Any assessment strategy needs to be efficient in terms of staff time, cost-effective for the organizations concerned and should ensure that learners find the tasks they are set manageable, relevant and developmental. Kent (1987) noted that the presence of an assessment was closely related to the extent of teaching in the various areas (subjects) and the number of hours provided.

Miller (1969) emphasized that evaluation of the results of teaching is considered to be fundamental to any educational system and Mosel (1964) had earlier stated that the judicious use of different assessment methods (including MCQ) resulted in greater acquisition of knowledge, hence advocated the administration of these during training as a means of adding to the efficiency of teachings. These changes in students following teachings were similarly noted by other authors (Burra et al., 1982; Wilkinson et al., 1983a). However, Dobson (1987) emphasized that the study must be well designed and sufficiently detailed to possibly assess these changes in the students following teaching. Hence, the assessment tasks need to be integrated into the learning process. Assessment is recognized as probably the most influential factor related to how students learn (Brown and Glasner, 1999), directly linked to effective teaching and learning by rewarding understanding and achievement (Jenkins, 2004).

Assessment of a student's performance on a particular module may often be thought of as a single evaluation of the extent to which the student has met some or the entire module's learning outcomes. More helpfully, however, when there is more than one assessment task in a particular module, it may be thought of as the combination of different assessment tasks (whatever the weighing attached to each of them). In this sense, the level of agreement between performances on the different tasks may elucidate the nature of overall student performance further still. Typically, assessment tasks within a module tend not to be of the same kind; often they are complementary, sometimes through the use of different media (James and Fleming, 2004).

Stefani (2004) have long argued that assessment needs to be 'fit-for-purpose'; that is, it should enable evaluation of the extent to which learners have learned and the extent to which they can demonstrate that learn-

ing (Brown and Smith, 1997). One of the components of the curriculum being studied is the variety of assessment methods used which include essays/short essay question (SEQ), multiple choice questions (MCQ), problem solving practical (steeple-chase) and *viva voce*/oral examination. Other forms may include self assessment materials, as well as evaluation of workbooks being used by the medical students.

In a similar study done by Ashiru et al. (1987), a 12 months old programme in Anatomy for a particular set at the College of Medicine, University of Lagos (CMUL) was considered. However, LASUCOM operates an 18 months curriculum programme. This present study was designed to give a detailed appraisal of the differential interaction of all the various assessment parameters as stated above on the overall students' performance in the professional examination at the end of the third semester. The major objective of this study was to assess the convergent validity of the professional anatomy (a multifaceted examination) with other markers of success (the various interactive assessment components of the curriculum) in determining the overall performance of third-year medical students as well as to isolate area of academic weakness among the students and to readjust the curriculum content to balance the weakness.

The concept of convergent validity implies that a construct should be correlated with theoretically similar constructs (Moffatt et al., 1971; David et al., 2001). The convergent validity of the professional examination was assessed by correlating the obtained results on the professional examination with other measures thought to be predictors of student performance, such as SEQ, MCQ, Steeple-chase and *viva voce*/oral examination. It is hoped that the current investigation would contribute towards appraisal of the efficiency of the curriculum with the view to introducing possible adaptation as appropriate taking into cognizance the dynamism of the evolving educational technology. It is also expected to shed greater light on the relevance and value of different assessment procedures in the areas of basic medicine in developing country such as Nigeria.

MATERIAL AND METHODS

Course logistics

The Gross Anatomy course at the LASUCOM is taught for three semesters, during the second and third year of the students' curriculum. The class usually contains about 65-68 second year Medical Students (MS) depending on the set admitted and their year of admission. The students normally receive lectures as a group (average of five hours/week), with additional five hours of laboratory time. In a traditional curriculum like ours, all 65-68 students are assigned to dissect and required to be in the laboratory for all six hours, every week, for the 17 weeks per semester for the whole three semesters.

Ancillary teaching materials

In the course, additional tools available to the students for the study

of anatomy include: Textbooks and atlases, course guides on the various units of anatomy (gross-anatomy, neuro-anatomy, embryology, and microscopic anatomy), prosections (prepared by the faculty members and postgraduates students in the department); computer assisted learning programs (e.g., Human Anatomy - Anatomy trainers 3B Scientifics and histology CD-Rom from University of South Carolina, USA. used to compliment the students' histological study), radiology and osteology seminars and access to additional learning tools (bones, models, museum pots/cross-sections and video dissections).

Dissection protocol

Basically, the traditional all students' dissection protocol is used where all the students are assigned alphabetically (by last names, A-Z) to 6 dissection tables in the laboratory. Thus, there are ten (or eleven) students at each table. The students' dissection manual is the Cunningham manuals which are reflected in the dissection programmed time-table seen in the departmental course guide in which the students have been enjoined to strictly obey. On the average, six hours per week of assigned dissection time was scheduled, with both the faculty (usually two) and postgraduate students (usually four) available in the laboratory (one person for one or two tables)

Teaching/seminar session protocol

Topics treated include the related osteology, radiology, clinical correlations and embryology. The students were given a comprehensive list of all structures to be dissected (in line with their dissection manuals and course guide) and all topics to be covered in the teaching/seminar session generated by the faculty. If for any reason the group did not recognize the problem, the grader usually the Head of Department or a faculty member assigned to that group will intervene, clarify the material, and instruct the group to perform the entire session again at a later time; this has occurred only once since the entire program begun.

Professional examinations protocols

The Gross Anatomy and Embryology subject examinations was administered at the end of the third semester as the students' only final examination in Gross Anatomy, this accounted for 70% of the final grade. The exam is made of two parts namely, part I- paper 1 which is MCQ of 200 questions and the part II- paper 2 which is a short essay questions (SEQ) of 8 questions covering the various units and aspects of the course. In addition the practical examination was timed (40 seconds per item), 50-question identification examinations with about 80% of the questions being two-part questions ('a' and 'b'). Examined material included tagged cadaver items, in addition to radiology, pots/cross-section, neuro-anatomy, osteology, embryology and histology questions. Students had prior knowledge of the regions upon which they would be assessed during the practical examination, but were randomly assigned to one of the stations on arrival for the practical.

The end incourse assessment carries 30% while the other professional examination assessment parameters (MCQ, SEQ and practical) carried 70% of the overall score. Each of the assessment parameters (such as essay, MCQ and practical) carries 100% of the 70% student's total grade in the final professional examination. Faculty members are encouraged to contribute questions in a scenario format questions, based on their areas of expertise. These questions are sent to an external examiner for moderation.

Subject

This investigation utilized a convenience sample and was a blinded cohort retrospective study carried out on the third year medical students, (2000 set). Thus, all 66 students in the class took the first professional examination in anatomy and records of their grades in the various assessments criteria were used for the study. Students' grades/scores can range from 0 to 100 percent. This set (2000) got admitted through Joint Admission Matriculation Board Examination (JAMB) entrance into the institution, hence was considered suitable for the study in view of their homogenous background. The professional examination is a multifaceted, dynamic examination designed to assess proficiency and performance related to each required course taken at LASUCOM during the prior 18 months.

Data analysis

Student performance data (i.e. percentage marks/grade) for each assessment point were acquired from central student records of electronic module results. The data were initially cleaned by removing student marks when no attempt was made at an assessment point, and were then sorted by student identification number in order to match students across the module. Computerized analysis of the data was done using SPSS11 (SPSS inc. Chicago, Illinois), software package.

Based on the convergent validity principle, analysis of data was correlated with the markers of success (predictor variables-assessment parameters) using Pearson's product moment of correlation and modified in this study using the rank spearman's correlation. Regression analysis of predictor variables and Mean \pm S.D for student percentage grades was equally done. A paired sample t-statistic was used to test for statistically significant differences in performance on all the assessment parameters. 0.001 was considered significant.

Ethics statement

The University's principles and procedures on research ethics were adhered to throughout the study. In particular, data on student performance were presented such that identification of individual student performance was impossible, thereby complying with the requirements of the Data Protection Act. Restricted access to the data is permitted only to those who have administrative (e.g. data collation and processing) and academic functions (e.g. management roles with teaching, learning and assessment responsibilities, roles overseeing pastoral responsibility and course leaders).

RESULTS

The male to female ratio for the two sets combined was found to be 1 to 2. 9% of the students failed while 91% passed, out of which 2% had distinction, 23% with credits grades and 66% with passes in the professional examination at first sitting (Table 1). For the overall grade performance at the professional exam, the mean \pm S.D value was found to be 54.5 ± 9.7 as compared to other assessment parameters. Although the overall performance has the same value with the Essay, Practical has the highest value of 59.8 ± 14.0 with End -incourse assessment examination having the lowest value of 16.6 ± 4.6 (Table 2).

Table 1. The Scored grades in Anatomy at the Final Professional Exam for the sets of 66 students as shown in the table below.

Code	Grade level	Absolute Frequency	Relative Frequency
1	Fail	6	9
2	Pass	44	65
3	Credit	14	23
4	Distinction	2	3
Total		66	100

Table 2. Crude Statistics of the data generated for the performance of 113 medical students (2nd Year) in Anatomy are shown below.

Predictor variable (marker of success)	Mean	Standard Deviation	S.E of Mean	No. of Candidate
End-incourse Exam	16.6	4.6	0.60	66
Practical (Professional) Exam	59.8	14.0	1.7	66
MCQ (Professional) Exam	51.9	11.8	1.50	66
Essay (Professional) Exam	54.9	13.9	1.7	66
Overall Performance	54.5	9.7	1.20	66

All the assessment parameters studied, their pearson's correlations values (r) with the overall performance in the professional examination were found to be significant ($P < 0.001$). The Essay ($r = 0.72$); End-Incourse ($r = 0.45$). MCQ ($r = 0.83$) and Practical ($r = 0.89$) are correlated with the overall performance in this order (Table 3). Cor-relation was consistently highest with the practical examination. The practical was positively correlated with overall performance (student t -value = 6.15 and $P < 0.001$ is significant at two-tailed point) (Table 4).

The End- incourse has the lowest t -value of -35.64 while practical has the highest value of 6.15 as reflected in Table 4b. Both the Practical and MCQ were closely and significantly correlated with the overall performance ($r = 0.89$ and 0.83 , respectively; $P < 0.01$). Although, the SEQ showed significant correlation to overall performance ($r = 0.72$; $P < 0.01$), the t -value of 0.4 was non-significant, but, within an acceptable range.

DISCUSSION

It was obvious that the pass rate was high (91%) and a lot need to be done to reduce the failure rate further more. This finding was contrary to that of Ashiru et al. (1987) where a high failure rate was recorded despite the fact that their students spent more hours in the course especially at the practical/dissection session. Stanford (1994) found that computer-based instruction (which utilizes tomographic images) did enhance performance of their students on the test when such images are combined with dissection; this finding suggests that innovative methods should be adopted.

In this context, it is of note that Davis (1972) found out that objective tests could indicate probable performance

in the oral biology professional examination. Further-more, Green (1979) emphasized that the True-False type of items used here in our study has been shown to be as reliable as the multiple-choice format. Support for this perspective emerged in the present study, where the difference between overall professional examination and multiple choice questions was significant ($t = -3.20$, $P < 0.001$) and highly correlated (0.83 , $P < 0.001$). Hence, MCQ carried more weight than other parameters except the practical examination in predicting students' outcome in the overall examination.

Modified essay questions (MEQ) is positively correlated with MCQ performance in similar content area but the magnitude of their correlation is not high enough in predicting student performance in National board Examination for Medicine, MCQ was found to be a better prediction (with a higher correlation value) of NBME part I and II performance than the MEQ (Rabinowitz, 1987). We observed this same trend in our study where the correlation of theory versus MCQ was found to be 0.38 ($P < 0.001$). Prout et al. (1976) also predicted similar correlation in student performance during biochemistry examination.

Rabinowitz (1987) noted that the essay questions though time consuming provides a unique evaluation tool particularly suited for the undergraduate settings. The essay utilizes the effects of biological, psychological and sociological factors; it is able to predict some of the attitudinal aspects of performance. The difference in t -values for incourse and overall professional examination was highly significant (-35.6 , $P < 0.001$). This finding is generally in line with findings reported by some authors like Pettingale et al. (1976) and Ashiru et al. (1984) who noted that it was of a great value at predicting student

Table 3. Correlation matrices for the various assessment parameters (examination models).

Assessment parameters	Incourse	Practical	MCQ	Essay	Overall Performance
Incourse	1.00	0.365**	0.498**	0.268**	0.45**
Practical		1.00	0.681**	0.667**	0.891**
MCQ			1.00	0.376**	0.833**
Essay				1.00	0.722**
Overall Professional Exam					1.00

N = 66; **. Correlation coefficient was considered significant at $P < 0.001$ (2-tailed).

Table 4a. Comparison of the various examination/assessment parameters with the overall professional examination performance using paired sample t-test correlation matrices.

Paired assessment parameters	Correlation (r)	Spearman's	No. of Students
Incourse – Overall	0.450 ($p < 0.001$)	0.755 ($p < 0.001$)	66
Practical – Overall	0.891 ($p < 0.001$)	0.831 ($p < 0.001$)	66
MCQ – Overall	0.833 ($p < 0.001$)	0.672 ($p < 0.001$)	66
Essay – Overall	0.722 ($p < 0.001$)	0.636 ($p < 0.001$)	66

$P < 0.001$ is significant

Table 4b. Students t-test analysis of the various assessment parameters. The overall performance parameter (professional examination) is the dependent assessment parameter.

Markers of success	t- value
Incourse – Overall	-35.6 ($p < 0.001$)
Practical – Overall	6.2 ($p < 0.001$)
MCQ – Overall	-3.2 ($p < 0.001$)
Essay – Overall	0.4 (NS)

(t-value is significant at $P < 0.001$).

performance in professional anatomy examination.

One consistent finding in this present study was that practical examinations consistently has the highest correlation value and significant t-value of 6.15, ($P < 0.001$) of all the assessment parameters. This was similar to the findings of Watt (1987) and Ashiru et al. (1984).

Conclusion

The Practical, MCQ, SEQ, and End incourse showed an overall rank order of relative performance in assessment tasks, indicating that, in general, students' performance in the Professional examination was better than in the End-incourse examination. Hence the need to reappraise this pattern and emphasize the role of incourse assessment in the curriculum. This study has also demonstrated that the Practical, MCQ, Essay, and End-incourse assessment are predictors of students' performance.

Therefore, a review of the curriculum with emphasis on the practical examination being an essential tool for proficiency of the future medical doctors is advocated.

There is also need for emphasis on the practical course guide already being introduced to complement the available dissecting manual. Furthermore, incorporation of innovation instructional methods such as computer based instructions is also advocated as a means to improving students' performance.

Limitations

However, the number of students in the analysis was small. The scope of the study did not cover the relative agreement/contributions of specific content area to performance. Additional studies are being planned to test this hypothesis further.

ACKNOWLEDGEMENTS

The secretariat assistance of Remilekun Shittu and Olofinlade Francis were appreciated.

REFERENCES

- Ashiru OA, Cressie C, Noronha, Adegoke O (1984). The value of incourse-assessment as predictors of medical students' performance in the Professional examination. August, pp. 7–11.
- Ashiru OA (1977). The role of Medical education in the development of Nigeria. *Anatomy Record*; 18: 535.
- Baume D, Baume C (1992). Assessing students' work, Oxford: Oxford Brooks University, Oxford Centre for Staff Development.
- Brown G (2001). LTSN Generic Centre Assessment Series No.3: Assessment: A Guide for Lecturers, York: Learning and Teaching Support Network.
- Brown S, Glasner A (1999). (eds) Assessment Matters in Higher Education: choosing and using diverse approaches, Buckingham: Open University Press
- Brown S, Smith B (1997). Getting to Grips with Assessment, Birmingham:

- Burra P, Kaliu R, leichner P, Waldron JJ, Handforth JR, Jerrette FJ, Amara B (1982). The ATP 30: a scale for measuring medical students' attitude to psychiatry. *Medical Education* 13: 185-192.
- DFES (2003). department for education and skills) Towards a Unified e-Learning Strategy: consultation document, Nottingham: DfES Publications.
- David A, Latif, Richard Stull (2001). Relationship between an Annual Examination to Assess Student Knowledge and Traditional Measures of Academic Performance Am(r). *J. Pharm. Edu.* 65: 346-349.
- Davis CO, (1972). Validity: quantitative, qualitative and quasi. *J. Dental Edu.* 36: 42-4.
- Dobson B, (1987). Medical student opinion of methods used to teach Child and adolescent psychiatry. *Med. Edu.* 21: 143-150
- Green k (1979). Multiple choice and true-false: reliability and validity compared. *J. Exp. Edu.* 48: 42- 4.
- Habeshaw S, Gibbs G, Habeshaw T (1993). Interesting Ways to Assess your Students, 3rd ed, Bristol: Technical and Educational Services.
- James D, Fleming S (2004). Agreement in Student Performance in Assessment Learning in Higher Education, issue I: 31-50
- Jenkins M (2004-05) (Unfulfilled Promise: formative: assessment using computer-Aided Assessment. *Learning and Teaching in Higher Education*, Issue 1. Kent G. (1987) Behavioural sciences teaching in the U K dental school *Medical Education*; 21: 105-10 8.
- Kolesnikov L, Sapin MR, Nikityak DB, Kulikor VV (2002). Features on teaching human anatomy in Russia; Proceeding of the International conference on Morphological Sciences (ICMS) Romania
- Miller AH, Imrie BW, Cox K (1998). Student Assessment in Higher Education: a handbook for assessing performance, London: Kogan Page
- Miller GE (1969). The study of medical education. *Br. J. Med. Edu.* 3: 5-10.
- Moffatt AJ, Jacobs AW, Metcalf WK (1971). Predictors of academic performance in gross anatomy. *J. Med. Edu.* 46: 545- 548
- Mosel JN (1964). The learning process. *Journal of Medical Education* 39: 485-96.5.
- Moxham BJ (2002). What is the purpose of Anatomy in the medical curriculum? Anatomists' Perceptions of the relationship between course aims and teaching methodology. Proceeding of the International conference on Morphological Sciences (ICMS), Romania, pp. 214.
- Pettinghale KW, Tomliss RW, Ryan KC, Anderson J, Gillian B, Clark (1976). The prediction of medical students performance using in course assessment. *Medical Edu.* 163 – 166.
- Prout RES, Hoy TG (1976). University examination performance of Dental students' correlation between entry qualifications and non-clinical and clinical examinations. *Br. Dent. J.* 141: 141 – 5.
- Rabinowitz HK (1987). The modified essay question: an evaluation of its use in a family Medicine Clerkship, *Medical Education* 21: 114 – 118.
- Race P (1995). The art of assessing 1, *New Academic*, vol. 4, no. 2, pp. 3-6.
- Race P (1996). The art of assessing 2, *New Academic*, vol. 5(1) 3-6.
- Rust C (2002). The impact of assessment on student learning: how can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices ?, *Active Learning in Higher Edu.* 3(2): 145-158.
- Report of the Advisory Board for the Establishment of Lagos State University, college of Medicine (LASUCOM), (September, 1987): 34 – 39.
- Sambell K, Mcdowell L (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning, *Assessment and Evaluation in Higher Education*, vol. 23(4): 391-402.
- Swann J, Eccelstone K (1999). Improving lecturers' assessment practice in higher education: a problem-based approach. *Edu. Action Res.* 7(1): 63-87.
- Stanford W, Erkonen WE, Cassell MD, Moran BD, Easley G, Carris RL, Albanese MA. (1994). Evaluation of a computer-based program for teaching cardiac anatomy. *Invest Radiol.* 29: 248-52.
- Stefani LAJ (2004-05). Assessment of Student Learning: Promoting a scholarly approach *Learning and Teaching in Higher Education*. Issue 1.
- Watt ME (1987). Retention of Pre-clinical knowledge by medical students. *Medical Edu.* 2(1): 119-124.
- Walton HJ (1967). The measurement of medical students' Attitudes. *Br. J. Med. Edu.* 1: 330-40.
- Wilkinson DG, Greer S, Toone BK (1983a). Medical students' Attitudes to psychiatry. *Sychological Med.* 13: 185-92.