The Differential Impact of Various Assessment Parameters on the Medical Students Performance in the Professional Anatomy Examination in a New Medical School

El Impacto Diferencial de Varios Parámetros de Evaluación del Desempeño de Estudiantes de Medicina en el Examen Profesional de Anatomía en una Nueva Escuela de Medicina

*Shittu, L. A. J.; **Zachariah, M. P.; ***Izegbu, M. C.; ****Adesanya, O. A. & *Ashiru, O. A.

SUMMARY: There is a dearth of knowledge on the level of agreement between all the various assessment tasks on the same content area, in order to test this hypothesis we adopted the concept of convergent validity and also to isolate area of academic weakness among the students and to readjust the curriculum content to balance the weakness. A blinded cohort retrospective study was carried out on a total of sixty-six third year medical students who had sat for their first professional examination in anatomy in the new medical college of Lagos State University. Using records of their grades in the various assessments parameters- the average end-in course assessment, short essays question (SEQ), multiple-choice questions (MCQ), and practical (Steeple-chase). The mean + S.D. Pearson's correlation with students t-test (p< 0.05 being significant) were analyzed using the computerized SPSS 11 (SPSS inc. Chicago, Illinois) software package.

The practical was significantly correlated with overall performance (r = 0.89, p< 0.01); with the student t-value of 6.15 (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.

The practical, MCQ, SEQ and end-incourse showed an overall rank order of relative performance in assessment tasks, therefore indicates that, in general, students performance in the Professional examination was better than in the end-incourse examination and hence, the need to reappraise this pattern and emphasize the role of incourse assessment in the curriculum.

KEY WORDS: Assessment parameters; Student performance; Anatomy; Nigeria.

INTRODUCTION

Medical education has been found to plays vital role in nation building. Hence, the review and formulation of curriculum component is an ongoing process in several medical schools all over the world (Ashiru *et al.,* 1984; Ashiru,1977).

However, the effectiveness of any educational innovation in itself remains undetermined unless evaluative research accompanies curriculum change (Walton, 1966).

Research into the various areas of assessment in higher education have also gained much attention since 1990,

(James & Fleming, 2004).This include concerned with the philosophy of assessment and of assessment practice (Miller *et al.,*1998; Swann & Eccelestone, 1999) focused on promoting good practice (Baume & Baume, 1992; Brown, 2001, among others).

Assessment is seen as one of the most powerful drivers of innovation and change in education, as it defines the goals for both learners and teachers.'(DFES, 2003). this is also linked directly to effective teaching and learning by rewarding understanding and achievement of learning outcomes (Jenkins, 2004).

* Department of Anatomy, Lagos State University, College of Medecine (Lasucom), PMB 21266 Ikeja, Lagos, Nigeria.
** Department of Anatomy, Department of Psychiatry, Lagos State University, College of Medecine (Lasucom), PMB 21266 Ikeja, Lagos, Nigeria.
*** Department of Morbid Anatomy, Lagos State University, College of Medecine (Lasucom), PMB 21266 Ikeja, Lagos, Nigeria.
**** Department of Anatomy, Igbinedion University, Okada, Edo-State, Nigeria.

More recently, qualitative studies have emphasized the importance of understanding the way students respond to innovations in assessment (Sambell & McDowell, 1998), as assessment is of singular importance to the student experience and learning (Rust, 2002; Habeshaw *et al.*, 1993; Brown & Glasner, 1999).

No doubt, the research for an ideal method of evaluating medical students and post graduates has undergone considerable change. We need to consider not just what we are assessing and how we are doing it (particularly which method and approaches), but also why? - Our rationale for assessing of any particular occasion and in what context.

However, 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 to perform manageable, relevant and developmental, where possible we must make best use of the available technologies to make assessment more efficient (Brown *et al.*, 1994).

James & Fleming have argued that 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. However, when there is more than one assessment task in a particular module, it may well be thought of as the combination of different assessment tasks (whatever the weighting 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

There is a dearth of knowledge on the level of agreement between all the various assessment tasks on the same content area.

The assessment component of the curriculum include Essays/short essay question (SEQ), multiple choice questions (MCQ), problem solving practical (steeple-chase) and viva voce/oral examination.

In order to test this hypothesis, the concept of convergent validity is used which implies that a construct should be correlated with theoretically similar construct (David & Richard, 2001), in this case 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.

To isolate area of academic weakness among the students and to readjust the curriculum content to balance this weakness.

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.

MATERIAL AND METHOD

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), and are then scheduled for 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 & 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, embryology, etc. 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.

Subjects. We carried out a blinded cohort retrospective study on the third year medical students, (2000 set). Thus, all 66 students in the class who took the first professional examination in anatomy had records of their grades in the various assessment tasks retrieved for this study. Students' grades/scores can range from 0 to 100 percent.

This set 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 nature 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. Computerized analysis of the data was done using SPSS11 (SPSS inc. Chicago, Illinois), software package.

The mean + S.D; Pearson's correlation and paired student t-test were used for student percentage grades. p< 0.01 was considered significant.

Ethics statement. This research was approved by the University's research and ethic committee. In particular, data on student performance were presented such that identification of individual student performance were made possible, bearing in mind the Data Protection Act requirement with restricted access to the data permissible 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 etc).

RESULTS

The male to female ratio for the set was 1:2. 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 as shown in Table I below.

Table I. The scored grades in Anatomy at the Final Professional Exam for the sets of 66 students.

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

For the overall grade performance at the professional exam, the mean + S. D. value was found to be 54.5 +9.7 as compared to other assessment parameters. 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, as shown in Table II.

661

The Pearson's correlations values (r) for all the assessment tasks were significant (p<0.01) with the overall performance in the professional examination were found to be significant (p<0.001), as shown in Table III.

Correlation was consistently highest with the practical examination. The practical has the highest significant predictive value with the overall performance t-value = 6.15

(p< 0.01) as shown in Table III.

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, as reflected in Table IV.

Table II. Crude Statistics of the data generated for the performance of 66 medical students (3rd Year) in Anatomy.

Variables	Mean	Standard deviation	S. E. of mean
End-incourse	16.6	4.6	0.6
Practical (professional) exam	59.8	14.0	1.7
MCQ (professional) exam	51.9	11.8	1.5
Essay (professional) exam	54.9	13.9	1.7
Overall performance	54.5	9.7	1.2

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

Assessment parameters	Incourse	Practical	MCQ	Essay	Overall performance
End-incourse	1.00	0.365**	0.498**	0.268**	0.450**
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<0.01 (2-tailed).

Table IV. Comparison of the various examination/assessment parameters with the overall professional examination performance of 66 medical students (3rd Year) in Anatomy, using Pearson correlation matrices and paired sample t-test.

Paired assessment parameters	Correlation(r)	t-value
End-incourse- overall	0.450**	-35.6**
Practical – overall	0.891**	6.2**
MCQ – overall	0.833**	-3.2**
Essay – overall	0.722**	0.4

Correlation coefficient was considered significant at p<0.01 (2-tailed).

DISCUSSION

Contrary to finding in other study (Ashiru *et al*.), a high pass rate of 91% was recorded; although, a lot still need to be done to reduce the failure rate further. It was found that computer-based instruction (which utilizes tomographic images) also did enhance the performance of the students on the test when such images are combined with dissection; this finding suggests that innovative methods should be adopted (Stanford, 1994).

Furthermore, we found out that the true-false type of items used here in our study has been shown to be as reliable as the multiple-choice format (Green, 1979).

Modified essay questions (MEQ) is significantly and positively correlated with MCQ performance in similar content area, with a high magnitude of their correlation

enough to predicting student performance in the professional Anatomy examination. MCQ was found to be a better prediction (with a higher correlation value of 0.83) of overall professional examination performance than the MEQ. This is similar to findings in other studies (Davis, 1972; Prout & Hoy, 1976; Rabinowitz, 1987).

Also, MCQ carried more weight than other parameters except the practical examination in predicting students' outcome in the overall examination.

We 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 incourse examination was an innovation that form part of the continuous examination exercise and found to be highly significant with the overall professional examination (-35.6, p<0.01).

From the study, we observed that the objective assessments (practical and MCQ) were more sensitive than the subjective assessment task (SEQ) in determining the overall performance of students in the final professional examination. The Practical examination was two times more sensitive than the MCQ (0.67: 0.34) in the same content as shown in Table III. This is in consonant with the general believed and it further drive home the fact that the level of the examination conducted by the system is of high quality and sensitivity.

Practical examination was significant and consistently has the highest correlation value with all the assessment parameters. This was similar to the findings of Watt (1987).

One possible reason for this in our study may be attributed to the fact that despite the unstable nature of school calendar as a result of incessant strikes that have characterized our educational sector lately, the structured programme and routine supervisors we have on ground in the department compensated for the academic lag phase between teaching and practical.

Hence 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 student's performance.

The practical, MCQ, SEQ, and end incourse showed an overall rank order of relative performance in assessment tasks on the same content area, therefore indicates 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 further demonstrated that the practical, MCQ, essay, and end-incourse assessment are predictors of students' performance.

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 over all performance. Additional studies are being planned to test this hypothesis further.

CONCLUSIONS

It is envisage that the level of agreement/interaction between performance outcomes on the different tasks may further elucidate the nature of overall student performance.

ACKNOWLEDGEMENTS

SHITTU, L. A. J.; ZACHARIAH, M. P.; IZEGBU, M. C.; ADESANYA, O. A. & ASHIRU, O. A. El impacto diferencial de varios parámetros de evaluación del desempeño de estudiantes de Medicina en el examen profesional de Anatomía en una nueva Escuela de Medicina. *Int. J. Morphol., 24(4)*:659-664, 2006.

RESUMEN: Existe escasez de conocimiento sobre el nivel de concordancia entre todas las varias tareas de evalución sobre los contenidos de una misma área. En orden a evaluar esta hipótesis, adoptamos el concepto de validez convergente y también de área aislada, de falta de solidez académica entre los estudiantes y readecuar el contenido curricular para balancear esa falta de solidez. Se llevó a cabo un estudio retrospectivo al azar sobre un total de 66 alumnos de Medicina de Tercer Año, quienes tuvieron su primer examen profesional de Anatomía en la nueva Escuela de Medicina de la Universidad del Estado de Lagos, Nigeria. Usando registros de sus grados en varios parámetros de evaluación: promedio final de la evaluación del Curso, preguntas cortas de redacción, ítemes de selección múltiple y prácticas. La media + desviación estándar, correlación de Person con el t-test de Sudents (p< 0.05) fueron analizados con el programa SPSS 1l (SPSS inc. Chicago, Illinois).

La parte práctica fue significativamente correlacionada con todo el rendimiento (r = 0.89, p< 0.01); con un valor del t-Students de 6.15 (p< 0.01). Aunque las preguntas cortas mostraron correlación significativa para la totalidad de lo hecho (r =0.72; p<0.01), el valor de t de 0,4 no fue significativo, pero estuvo dentro de un rango aceptable.

La parte práctica, selección múltiple y preguntas cortas de redacción y final del curso, mostraron un orden de clasificación de relativo desempeño en las tareas de evaluación, lo que indica que, en general, la actividad de los estudiantes en el examen profesional fue mejor que en el examen del curso y de ahí, la necesidad de readecuar este patrón y enfatizar el rol de evaluación dentro del curso en el curriculum.

PALABRAS CLAVE: Parámetros de evaluación; Desempeño de estudiantes; Anatomía; Nigeria.

REFERENCES

Ashiru, O. A.; Cressie, C. Noronha & Adegoke O. The value of incourse- assessment as predictors of medical students' performance in the Professional examination. August, pp. 7-11, 1984.

Ashiru O. A. The role of Medical education in the development of Nigeria. *Anat. Record., 18*:535, 1977.

Baume, D. & Baume, C. *Assessing students' work, Oxford.* Oxford Brooks University, Oxford Centre for Staff Development, 1992.

Brown, S.; Rust, C. & Gibbs, G. Strategies for Diversifying Assessment, 1994.

Brown, G. LTSN *Generic Centre Assessment Series Nº.3: Assessment: A Guide for Lecturers.* New York. Learning & Teaching Support Network, 2001.

Brown, S. & Glasner, A. (eds) *Assessment Matters in Higher Education: choosing and using divers approaches.* Buckingham, Open University Press, 1999.

DFES (Department for education and skills) Towards a Unified e-Learning Strategy: consultation document, Nottingham: DES Publications, 2003.

David, A. Latif & Richard Stull. Relationship between an Annual Examination to Assess Student Knowledge and Traditional Measures of Academic Performance. *Am. J. Pharm. Educ. 65*:346-9, 2001.

Davis C. O. Validity: quantitative, qualitative and quasi. *J. of Dental Education, 36*:42-4, 1972.

Green, K. Multiple choice and true-false: reliability and validity compared. *J. of Experimental Education, 48*:42-4, 1979.

Habeshaw, S.; Gibbs, G. & Habeshaw, T. *Interesting Ways to Assess your Students*. 3rd. ed. Bristol, Technical and Educational Services, 1993.

James D. & S. Fleming. Agreement in Student Performance in Assessment Learning in Higher Education, 1:31-50, 2004.

Jenkins, M. Unfulfilled Promise: formative: assessment using computer- Aided Assessment. Learning and Teaching in Higher Education, Issue1, 2004.

Miller, A. H.; Imrie, B.W. & Cox, K. *Student Assessment in Higher Education: a handbook for assessing performance*. London, Kogan Page, 1998.

Prout, R. E. S. & Hoy. T.G. University examination performance of Dental students' correlation between entry qualifications and non-clinical and clinical examinations. *British Dental Journal, 141*:141-5, 1976.

Rabinowitz, H. K. The modified essay question: an evaluation of its use in a family Medicine Clerkship. *Medical Education 21*:114-8, 1987.

Rust, C. 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 Education, 3(2)*:145-58, 2002.

Sambell, K. & Mcdowell, L. The construction of the hiddencurriculum: messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education, 23(4):*391-402, 1998.

Swann, J. & Eccelstone, K. Improving lecturers' assessment practice in higher education: a problem-based approach. *Educational Action Research, 7(1):*63-87, 1999.

Stanford, W.; Erkonen, W. E.; Cassell, M. D.; Moran, B. D. Easley, G.; Carris, R. L. & Albanese, M. A. Evaluation of a computer-based program for teaching cardiac anatomy. *Invest. Radiol., 29*:248-52, 1994.

Watt, M. E. Retention of Pre-clinical knowledge by medical students. *Medical education; 2(1)*:119-24, 1987.

Walton, H. J. The measurement of medical students` attitudes. *British Journal of Medical Education, 1*:330-40, 1966.

Correspondence to:
Dr. Shittu L. A. J.
Department of Anatomy
Lagos State University
College of Medecine
(Lasucom), PMB 21266
Ikeja, Lagos
NIGERIA

Email: drlukemanjoseph@yahoo.com