

# Student Evaluation of Teaching: A Case Study from School of Medicine (UNZA)

**Sekelani S. Banda**

Department of Medical Education Development  
School of Medicine, University of Zambia  
Lusaka, Zambia

## ABSTRACT

**Purpose:** This paper presents a case study of an academic department's experience with evaluation. The purpose is to review the impact of student evaluation of teaching. The paper also introduces a new evaluation scoring method: the University of Zambia Staff Appraisal System (UNZASAS) method.

**Method:** Anonymous 5-point Likert scale evaluation rating forms were administered to 134 third-year medical students in two consecutive years to measure students' perceptions of teaching quality of four faculty members in an academic department at School of Medicine. The rating forms were scored using the UNZASAS method.

**Results:** The response rate averaged 83%. The group average lecturer rating improved from 120 to 141.5. Individual performance of three lecturers improved while that of one declined. The UNZASAS method was effective in summarizing data to identify the areas of weakness and strength in the faculty members. The specificity and sensitivity of this method were fundamental to its success as a diagnostic tool for formative evaluation of teaching quality.

**Conclusions & Implications:** The quality of teaching in the academic department improved after evaluation of teaching was introduced. The

UNZASAS method proved to be an effective tool for scoring evaluation data and for providing faculty with useful and specific formative feedback. Health professionals training institutions can consider using the UNZASAS method for their evaluation of teaching contribution by educators.

## INTRODUCTION

In 2006 the University of Zambia Senate approved the school of medicine's roadmap, which articulated policies, governance and administration, and curriculum goals for the MB ChB programme. The handbook aims to instigate the School to meet the World Federation for Medical Education (WFME) basic standard<sup>1</sup>. Programme evaluation is one of the nine standards described by the WFME basic standard. Like each standard, it has sub-areas and a defined standard describing the level of attainment. In the quest to meet the basic standard, among other activities, the School must introduce programme evaluation. The components of programme evaluation include a) mechanisms for programme evaluation, b) teacher and student feedback, c) student performance, and d) involvement of stakeholders. This paper focuses on teacher and student feedback, in particular, student evaluation of teaching. It presents a case study of an academic department's experience with evaluation. The purpose is to review the impact of student evaluation of teaching. The information can be useful to health professionals training institutions preparing to implement large scale evaluation of the teaching contributions of lecturers and honorary lecturers.

**Key words:** Evaluation, Impact, Teaching, Accountability, Quality.

**Corresponding Author:** Dr. Sekelani S. Banda  
Department of Medical Education  
School of Medicine, University of Zambia  
PO Box 50110, Lusaka, Zambia  
E-mail: [ssbanda@zamnet.zm](mailto:ssbanda@zamnet.zm)

Evaluation of teaching has been around for a long time<sup>2</sup> and is usually concerned with accountability - to assure the public and relevant authorities of the standards of teaching at the institution. Regrettably, evaluation of teaching has had a lot of opposition and lots of controversy, even to date. For example, it has been blamed for grade inflation<sup>3,4</sup> to mention but one. However, unlike North America, Europe and Australia where evaluation has got accepted and is widespread<sup>5</sup>, in Africa, and in Zambia in particular, it is yet to be established. Although some opponents of student evaluation of teaching cite issues related to reliability and validity of the evaluation research indicates that students' ratings are adequate in terms of validity and reliability<sup>6,7,8,9</sup>. Ratings of a given instructor are reasonably consistent over courses, years, rating forms, and groups of raters. Other evidence has shown that students taught by highly rated teachers tend to learn the subject matter better than those taught by lower rated teachers<sup>8</sup>. However, it must be admitted upfront that student evaluation of teaching does not tell the whole story and can be lacking in scope to measure the worth of academic standards of health educators, and must always be supplemented by other sources of data on teaching and academic contribution.

## A CASE STUDY

One academic department was identified for this case study because it was one of the few, perhaps the only one, which conducted evaluation of teaching as a departmental routine. This case study explores whether student evaluation made a difference in improvement of quality of teaching in that department. This exploration is a most important because improvement of teaching is the ultimate purpose of teaching evaluation, and should be most justifiable reason for evaluating teaching. The case study also presents a new evaluation scoring system that was used.

## METHODS

Anonymous 5-point Likert scale rating forms with 12 items (in year 1) and 13 items (in year 2) were used in two consecutive years to measure students perceptions of teaching quality of four faculty members in the selected academic department. Students were requested to indicate the extent (strongly agree, agree, undecided, disagree, strongly disagree) to which they agreed or disagreed with

statements on the evaluation rating form concerning individual lecturers. The rating forms were administered to 134 (75 in year 1 and 59 in year 2) third-year medical students who had earlier completed two years of university pre-med courses. In both years of the study the rating forms were administered at the end of the second semester when all teaching activities were completed. The following were the attribute statements evaluated: A1 - Communicates effectively; A2 - Teaching style makes note taking difficult; A3 - Enthusiastic about teaching his/her courses; A4 - Stimulates my interest in his/her courses; A5 - Interested in students; A6 - Accessible to students outside classes; A7 - Encourages students to participate; A8 - Well organised; A9 - Gives clear explanations; A10 - Confident; A11 - Helpful in laboratory; A12 - Attends all his/her scheduled lectures; A13 - Uses audio-visual aids effectively. The rating forms were scored using a purpose developed scoring method, that is, the University of Zambia Staff Appraisal System (UNZASAS) method shown in figure 1.

### UNZASAS Computation Steps:

1. Teaching attributes to be evaluated were selected (object attribute) and an evaluation statement was developed for each attribute.
2. A decision was made on whether the agreement statement on each attribute was positive or negative.
3. A 5-point agreement scale was constructed for each attribute.
4. The option values (weighting) were assigned as follows: for a positive statement = strongly agree = +2; agree = +1; undecided = 0; disagree = -1; strongly disagree = -2. For negative statements the weightings were: strongly agree = -2; agree = -1; undecided = 0; disagree = +1; strongly agree = +2.
5. The evaluation rating forms were administered to students.
6. The frequency on each of the options (i.e., SA, A, U, D, SD) for each attribute was computed. The frequency tally for each option was converted into a percentage.
7. The following **frequency values** were assigned to the following percentage ranges:  
0-9% = 1, 10-19% = 2, 20-29% = 3, 30-39% = 4, 40-49% = 5, 50-59% = 6, 60-69% = 7, 70-79% = 8, 80-89% = 9, 90-100% = 10.
8. Multiply the frequency value by the option value to obtain an **option score** at each option.
9. Compute the sum of the option scores at each attribute to obtain an **attribute score**. The sum of all the attribute scores yields the **Staff Appraisal Score**.

Example

<ol style="list-style-type: none"> <li>1. Evaluation statement of attribute: <i>Lecturer encourages students to participate in class</i></li> <li>2. Agreement Statement positive or negative: <i>Positive</i></li> <li>3. <i>5-point agreement scale:</i></li> </ol>						
	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree	
Option values	+2	+1	0	-1	-2	
Frequency values	(n=50) = 50% = 6	(n= 20) = 20% = 3	(n=1) = 1% = 1	(n=19) = 19% = 2	(n=10) = 10% = 2	
Option Score	+2 x 6 = 12	+1 x 3 = 3	0 x 1 = 0	-1 x 2 = -2	-2 x 2 = -4	
Attribute Score	+12	+3	0	-2	-4	9
In the example the attribute score is 12+3+0+[-2]+[-4] = 9.						

Figure 1, Steps for UNZASAS Computation and example of scoring using the UNSAS method.

Interpretation of scores

For a ten-attribute rating form the maximum (best) score possible using the UNZASAS method would be 200 and the minimum (worst) score possible would be -200. Some of the ways a score of zero could be obtained are as follows:

1. All (100%) the responses are the option ‘undecided’.
2. Fifty percent (50%) of the responses are at the option ‘Strongly Agree’ and the other 50% are ‘Strongly Disagree’.
3. Fifty percent (50%) of the responses are option ‘Agree’ and the other 50% are ‘Disagree’.

For that reason, the UNZASAS scoring method has the scale midpoint at zero. The UNZASAS is a compensatory scoring procedure; higher scores from some raters will ‘compensate’ for lower scores from other raters.

Feedback to lecturers consisted of a UNZASAS record card indicating the score achieved at each attribute and the overall SAS (together with maximum possible scores respectively). With regard standard setting, the case study department agreed that it was acceptable for 20% of the students to be dissatisfied with a lecturer. Given this scenario:

1. If 80% of the students are very satisfied (responses rated ‘strongly agree’ on positive statements) and 20% of them very dissatisfied (responses rated ‘strongly disagree’ on positive statements), the Staff Appraisal Score on a ten item evaluation questionnaire will be 120 (180-60). If 80% of the students are very satisfied (responses rated ‘strongly agree’ on positive statements) and 20% of them dissatisfied (responses rated ‘disagree’ on positive statements), the Staff Appraisal Score on a ten item evaluation questionnaire will be 150 (180-30).
2. If all the 80% of the students that are satisfied give ratings of ‘agree’ to positive statements and the dissatisfied maintain the strongly disagree, the SAS on a ten item evaluation questionnaire is 60 (90-30). However, if the dissatisfied students were very dissatisfied (responses rated ‘strongly disagree’ on positive statements) the SAS is 30 (90-60).

Taking into consideration the above assumptions the academic department studied set 100 as the minimum satisfactory score for an evaluation instrument measuring ten attributes, even if statistically the midpoint is zero. Scores under 100 were considered as requiring remedial action and were discussed in the feedback meeting. The results of the survey of all faculty members, including those of the head of department, were distributed to each faculty member and later discussed openly at a departmental meeting in a non-threatening but with fact-of-the-matter approach. The feedback itemized the scores at each attribute measured, so each faculty member could discuss possible explanations for their ratings for specific attributes. Other faculty members also offered suggestions on how problems could be addressed.

**RESULTS**

The response rate was 84% (63/75) in year 1 and 83% (49/59) in year 2. The student ratings in the academic department in year 1 were compared to those of year 2 for the average lecturer ratings and for each lecturer. The group average lecturer rating improved from 120 to 141.5. In both years the average teaching quality was above the set 100, however, clearly the quality improved markedly in year 2 (figure 1).

Regarding individual lecturer performance, lecturer 2 scored 14/240 which was well below the set standard of 100. The department expressed great dissatisfaction with this performance and the lecturer was made aware of this impression in the feedback meeting. In year 2, however, lecturer 2 scored 108/260 above the set 100 mark. In general the standards for lecturer 4 marginally declined over the two years. The performance of lecturer 1 was comparatively consistent across all the attributes but generally recorded an improvement. The most improvement for lecturer 1 was recorded on attendance of scheduled lectures. Lecturer 2 improved his/her performance on all the attributes and recorded a very significant seven-fold improvement in the evaluation rating score. The students were no longer disagreeing with the positively structured evaluation statements about lecturer 2. Lecturer 3 the best performing lecturer in both years maintained high performance scores on all attributes. Lecturer 4, on the other hand, was rated lower on many attributes and most significantly concerning enthusiasm, stimulating student interest, interest in students, and being well organized. The longitudinal performances of the lecturers are illustrated in figure 2, 3 and 4.

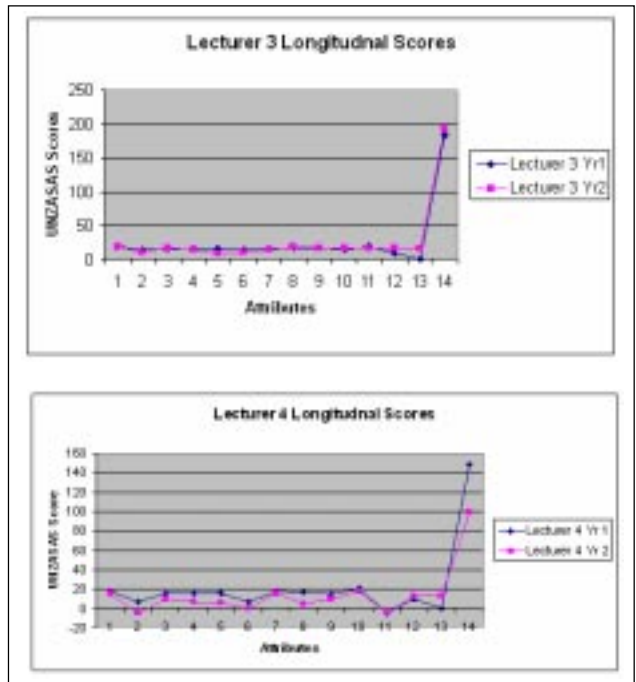


Figure 3 Longitudinal Rating Score Performance for Lecturer 1 and 2

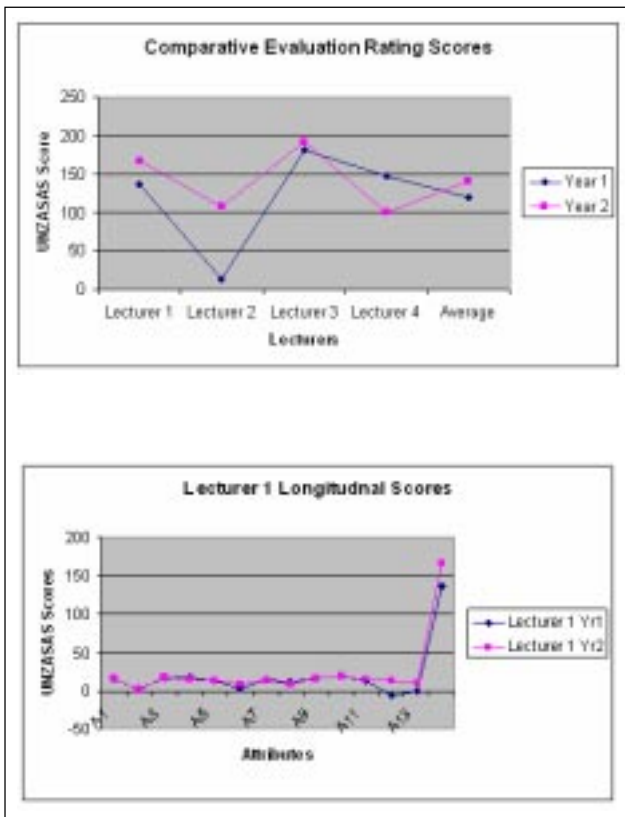


Figure 2 – Comparison of Evaluation Rating Performance in Year 1 and Year 2

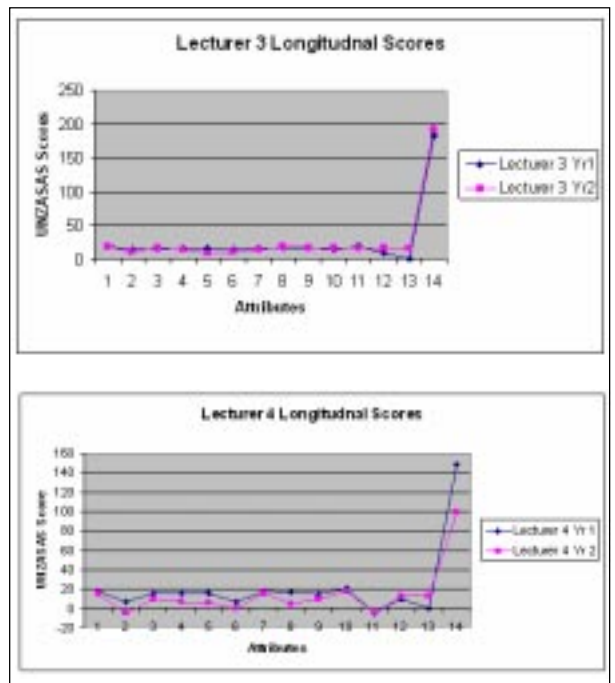


Figure 4 Longitudinal Rating Score Performance for Lecturer 3 and 4

## DISCUSSION

Generally, evaluation contributed to the improvement of teaching in the academic department as evidenced from the improved average lecturer ratings from 120 to 141.5. The feedback was also specific enough to allow the lecturers to work on specific attributes and improve on them. This is clearly evident from the improvement of lecturer 2's performance across all attributes. Lecturer 1 and 3 maintained what they were doing well and could track their performance on specific attributes requiring correction. Regrettably, the performance of lecturer 4 declined and it is not clear why. A speculation is that lecturer 4's was generally not actively involved in departmental activities and this hurt his/her ratings.

The UNZASAS method was effective in summarizing data to identify the areas of weakness and strength in the faculty members. The feedback was meaningful and effective, as demonstrated by the success of remedial action taken to improve the performance of Lecturer 2, whose SAS score rose from 14 in year 1 to 108 in year 2. The UNZASAS is a sensitive method because it uses negative and positive scores on the Likert scale to indicate positive and negative tendencies of the perceptions of the learners. The method also takes into account the percentage of the respondents who agree or disagree on ratings of each attribute (frequency value). The specificity and sensitivity of this method was fundamental to its success as a diagnostic tool for formative evaluation of teaching quality. Another advantage of the method is that it does not violate statistical principles for manipulating non-parametric data. Data obtained from instruments that measure attitudes with 5-point or 7-point Likert scales are non-parametric; however, the literature is replete with reports that have nonetheless ignored this rule<sup>10,11,12,13</sup>. Other workers<sup>14,15,16</sup> have avoided this mistake by reporting their findings as frequencies for each measured item. The UNZASAS method avoids the flawed approach of using parametric statistical manipulations by not using means, variance, standard deviations and analysis of variance; instead, it arbitrarily assigns integer values to the scale item options. Furthermore, the method harnesses the power of frequency distribution at each scale option to indicate the magnitude of positive or negative tendency for each measured attribute. This method

does not address quantitative contributions of the faculty, which could be measured by other methods. The method by Bardes<sup>17</sup>, which assigns a relative weight to each teaching activity and calculates scores according to the formula:  $RVU = \text{hours} \times \text{weight}$  is one such a method for quantitative measurement. Student ratings are valid<sup>6,7,8,9</sup>, however, and should be used for evaluation of teaching<sup>18</sup>. Although the study was done prior to the adoption of the School Roadmap which aims to instigate the School to achieve WFME standards the study remains pertinent because the School has not yet implemented any school-wide evaluation procedures.

## CONCLUSION

The teaching in the academic department was of better quality the following year and the evaluation of teaching evaluation contributed to that improvement because students provided useful feedback for the lecturers. The UNZASAS method because of identifying tendencies about specific attributes, and by being sensitive proved to be an effective tool for interpreting students' perceptions on the quality of teaching given by academic faculty, and providing faculty with formative feedback. The UNZASAS method is unique to this study and it may be first time it is being reported. More detailed validation studies are required before generalizations can be made about its usefulness. However, it performed exceptionally well for this study and health professions training institutions can consider using the UNZASAS method for widespread implementation of their evaluation of teaching programmes.

## REFERENCES

1. World Federation for Medical Education (2003). Basic Medical Education Global Standards for Quality Improvement. World Federation for Medical Education.
2. Tomlinson, L. (1955). Pioneer Studies in the Evaluation of Teaching. *Educational Research Bulletin*, Vol. 34 (3): 63-71.
3. Johnson, V.E. (2003). *Grade Inflation: A Crisis in College Education*. New York: Springer-Verlag.

4. Worthington, A.G. and Wong, P.T. (1979). Effects of Earned and Assigned Grades on Student Evaluations of an Instructor. *Journal of Education Psychology*, 71: 764-775.
5. Murray, H.G. (1997). Does Evaluation of Teaching Lead to Improvement of Teaching? *International Journal of Academic Development*, 2: 8-23.
6. Greenwald, A. (1997). Validity Concerns and Usefulness of Student Ratings of Instruction. *American Psychologist*, 52(11): 1182-1186.
7. Murray, H.G. (1985). Classroom Teaching Behaviours Related to College Teaching Effectiveness. In J.G. Donald and A.M. Sullivan (Eds), *Using Research to Improve University Teaching*. San Francisco: Jossey-Bass.
8. Cohen, P.A. (1981). Student Ratings of Instruction and Student Achievement: A Meta-analysis of multisection validity studies. *Review of Educational Research*, 51:281-309.
9. Cohen, A.P. (1980). Effectiveness of Student-rating for Improving College Instruction: A Meta-analysis of findings. *Research in Higher Education*, 13: 321-341.
10. Cook, D., Brown, L., Skajun, E. (1997). Factors which Influence the Outcome of Student Evaluation of Teaching. In Scherpbier, A., van der Vleuten, C., Rethans, J., van der Steeg, A. (Eds). *Advances in Medical Education*, pp 545-547.
11. Prideaux, D., Henry-Edwards, S., Marshall, V. (1997). Evaluating the Outcomes of Changes in Undergraduate Surgery Education: Interns' Views on Their Preparedness for Surgical Rotations. In Scherpbier, A., van der Vleuten, C., Rethans, J., van der Steeg, A. (Eds). *Advances in Medical Education*, pp 581-584.
12. Sukchareon, I. (1997). Evaluation of Clinical Teaching Processes: The Development of a Questionnaire. In Scherpbier, A., van der Vleuten, C., Rethans, J., van der Steeg, A. (Eds). *Advances in Medical Education*, pp 590-593.
13. Buschbacher, R., Braddom, R. (1995). Resident versus program director perceptions about PM & R research training. *American Journal of Physical Medicine and Rehabilitation*, 74(2): 90-100.
14. Graber, D., Bellack, J., Musham, C., O'neil, E.(1997). Academic deans' views on curriculum content in medical schools. *Academic Medicine*, 72(10): 901-7.
15. Covinsky, K., Bates, C., Davis, R., Delbanco, T. (1996). Physicians' attitudes toward using patient reports to assess quality of care. *Academic Medicine*, 71(12): 1353-6.
16. McCashland, T., Zetterman, R., Ruby, E., McCashland, C., Wigton, R. (1996). Attitudes and expectations of 1995 gastroenterology graduates about gastroenterology. *American Journal of Gastroenterology*, 91(10): 2091-5.
17. Bardes, C.L., Hayes, J.G., Falcone, D. J., Hajjar, D. P., Alonso, D. R. (1998). Measuring Teaching: A Relative Value Scale in Teaching. *Teaching and Learning in Medicine*, 10 (1): 40-43
18. McKeachie, W. (1997). Student Ratings: The validity of Use. *American Psychologist*, 52(11): 1218-1225.