Original article

Dimensional Differences of Evaluation Results of Instructors' Teaching Effectiveness: A Case of Faculty of Education, Jimma University.

Mekonnen Gorfu*

Abstract

The study is conducted in the faculty of Education of Jimma University in the first semester of 2004/2005 academic year. It investigated the variability of instructional effectiveness as measured by students' ratings or evaluation. All instructors/lecturers (N = 79) and 2, 370 students were involved in the study. The students were asked to rate their instructors at the end of the semester. Factor analysis, means and standard deviation were employed to analyze the data. The use of factor analysis has succeeded in identifying eight distinct dimensions or units of factor analysis has succeeded in identifying eight distinct dimensions or units of teaching effectiveness: Preparation & organization, Group interaction, Task Responsiveness & Enthusiasm, Professional Ethics, Rapport, Assessment skills, Punctuality, and objectives & Content Clarity. And, there is high level of relative agreement ($\propto$ = 0.63 – 0.93) among the different items found to be included under the same dimension or factor of effective teaching. With regard to variability of instructional effectiveness, item 12 (knowledge of the subject matter got the maximum rating ($\bar{\chi}$ 4.66) and item 8(accessible to students) received the minimum rating($\bar{\chi}$ = 3.80). The instructors of the faculty are also received the highest rating on the dimension of preparation and organization and lowest rating on their punctuality. At stream level, both language and social sciences are rated the highest at their preparation and organization and lowest in their punctuality; whereas, natural science instructors are evaluated highest in their rapport (dimension five) and receive lowest ratings in their assessment skills. Based on these findings, the following recommendations were forwarded: (1) The summary reports given to instructors should be based on the eight identified factor patterns of the evaluation of the questionnaire. Otherwise, broad global ratings averaged across a collection of heterogeneous items provide little diagnostic feedback and are difficult to interpret. (2) The instructors should improve their availability during consultation hours; they need to have the courage in preparing teaching materials (texts, manuals, etc). (3) Social science departments have to improve themselves with these items: item 11 (welcoming those students seeking help and advice),item 16 (assignments and feedbacks), item 17 (clarified methods of assessment) item 18(inviting idea sharing), and item 26 (appropriate use of instructional materials).

---

*Lecture, Psychology dept., Educ. Faculty, JU, Jimma, Ethiopia

Introduction

Students' rating has along history and a wide spread use in faculty evaluation. Administrative decisions regarding salary, promotion and other incentives have been including student evaluation results since earlier times. For example, as early as 1951 Miller (cited in Weigel, 1971) reported that a large number of institutions had been involved with the ratings of instructors by their students.

Different types of student evaluation forms have been used by college instructors for many years and interest in formal evaluation made by students seem to be increasing currently; it is a fairly standard means of evaluation on many college and university campuses.

Due to the increasing popularity of student rating as a measure of instructional effectiveness, it has attracted a great deal of attention of many researchers on its validity (Cohen, 1981). There are, of course, a number of excellent research summaries on the most controversial issues involving student ratings. Some of these include the following: Cohen, 1980; Marsh, 1984, and Stumpf and Freedman, 1979. Thus, this area has been the object of many studies; opinions about its reliability and usefulness vary drastically.

Students' rating one of the three general strategies employed to assess the effectiveness of an instructor in teaching, has been widely used. Despite the fact that the use of student rating for administrative purposes triggered considerable debates among academic staffs, it is now practicing by all of our universities and colleges. To this effect, it is always essential, as Assefa (1999) recommended, conducting continuous researches and making use of the findings for the effeteness of teaching-learning processes.

When we see the rationale and purposes of evaluation, the process of evaluation is justified for different purposes: improvement of instruction, directing and guiding faculty efforts, and to gather data for research on teaching and learning, to mention but a few. Thus, measuring teaching effectiveness can be

cited as one of the fundamental purposes of evaluation.

Different scholars crafted the definition of effective teaching in different ways. The different notions of effective teaching pave a good way in order to identify a number of dimensions of effective teaching. Based on the complexity of instructional processes, it seems reasonable to assume that the teaching process is multidimensional and that evaluation instruments should attempt to measure these dimensions, and feedback to the instructors has to be given using these dimensions. Like the teaching they are designed to assess, student ratings vary along different dimensions: organization/ planning, interaction, enthusiasm, coverage, examination/grading, instructor's knowledge, and others.

As long as the main purpose of evaluation is to give feedback for instructors so that they can see their weaknesses and strong sides, the need to summarize student ratings by dimensions is unquestionable. To this effect, when Jimma University revised the evaluation questionnaire of different faculties, the instructor performance evaluation format of education faculty is seems to comprise eight dimensions of teaching: organization, group, interactions, enthusiasm, task responsiveness & enthusiasm assessment skills, professional ethics, rapport, and punctuality. However, summary reports given to instructors are not still based on the identified dimension. Broad global ratings averaged across a collection of heterogeneous items provide little diagnostic feedback and difficult to interpret at an individual level. Above all, there is no evidence that shows to which dimensions of effective teaching instructors are rated low or high. Thus, taking into consideration this problem, the present study investigated the picture of ratings of instructors of the faculty for each items and dimensions.

More specifically, the research attempted to answer the following basic questions:

- How many factors or dimensions of effective teaching do JUIPEQ measure? How much is the agreement among the different items

which are found to measure the same factor or dimension?

- Among the identified dimensions, which dimension (s) is/are rated highest or lowest?
- Which (is) of the evaluation questionnaire is/are rated high or low?
- Are there rating differences among social sciences, natural sciences, and language departments? If so, which items or dimensions of effective teaching are rated high or low?

Depending on the effectiveness or approaches of teaching, all instructors couldn't have identical ratings across all dimensions of teaching effectiveness. The main objective of his study is to assess the pattern of student ratings of instructors of the education faculty across some dimensions of effective teaching: organization, interactions, enthusiasm, task responsiveness, assessment skills, ethics, rapport, and punctuality.

In the light of the questions stated previously this study tried to identify:

- Dimensions or factors of effective teaching measured by JUIPEQ,
- Items rating differences of instructors' effective teaching,
- Dimensional variability of instructors' teaching effectiveness, and
- Dimensions of effective teaching that are rated high or low.

The following hypotheses were also formulated to examine dimensional variability of instructor's teaching effectiveness:

- In the evaluation results of the instructors involved in this study, there would be differences of ratings across the twenty-eight evaluation items.
- In the evaluation results of the instructors involved in this study, there will be differences of ratings across the dimensions of effective teaching. And, the difference could be attributed to deficiency of skills in such areas.
- There will be variability of ratings as a matter of nature of courses or types of stream.

Students' evaluation of teaching is proved to be one important element of the teaching-learning processes (Centra, 1973). The formative function of the evaluation is important for the Improvement of instructors' performance. The evaluation results provide basis for self-improvement. The feedback attained by the evaluation results is not only important to faculty members, at individual level, but it is also very essential for the faculty or for a given institute in general. The evaluation results will guide the faculty or institute to take some corrective measures in those dimensions of teaching effectiveness rated low. And, in order to fill the gap by short or medium-term trainings or other measures, the faculty needs to see where the weakness lies.

Thus, it seems imperative that research on empirical bases should be conducted in order to see the pattern of evaluation on items and identified dimensions of teaching effectiveness of education faculty. It is hoped that the result of this study will help:

- To pinpoint the dimensions of teaching effectiveness to which instructions, on the average, show weakness,
- To pave a way to faculty evaluation that instructors should get feed back about their teaching not on the aggregated mean but on each dimension of teaching effectiveness, and
- As a source of general feedback to the faculty's teaching staffs, and other concerned bodies.

Moreover, the result of this study may be of use to other educational researchers who may need some picture of the state of the art in research on the students' evaluation of teaching.

METHODOLOGY AND DESIGN OF THE STUDY

Subjects

The data for this study were collected from faculty of education, Jimma University. All instructors/ lectures of the faculty (N =79) were taken as study subjects. Among these 26 were graduate assistants and 53 were lectures.

Data of the first semester of 2004/2005 academic year were processed and used in this study. When the data were tallied

or processed each subject remained anonymous.

Table 1: Number of instructors in the study

Department	Male	Female	Total
Psychology	12	-	12
Geography	4	-	4
History	7	-	7
English	15	-	15
Amharic	3	-	3
Afan Oromo	5	-	5
chemistry	4	-	4
Physics	8	-	8
Biology	3	-	3
Math	9	-	9
Pedagogy	7	-	7
Total	79	-	79

Using the available data from all study subjects the necessary descriptive analyses were made. The data were analyzed using factor analysis, means and standard deviations.

Operational Definition of Terms

Major concepts used in this study are defined as follows:

Evaluation: Evaluation is a broad term covering all forms of judgment. In this study, however, the term evaluation" is used specifically to imply rating of instructors by their students.

Rating: Rating refers to a subjective, qualitative judgment of an instructor by rater or by his or her students.

Reliability: Reliability, in this study, refers to the extent of the evaluation questionnaire being free from random error in evaluating instructors by their students.

Dimension of Teaching: Refers to taxonomies of teaching behaviors as measured by the 28 evaluation items the faculty.

Results

The data collected have been analyzed in view of the basic research questions, and the forth coming results have been found.

Dimension/Factors of Effective Teaching Measured by JUIPEQ

In this part of the study the research questions to be answered were:

- How many factors or dimensions of instructors' effectiveness do the JUIPEQ measure?
- How much is the agreement among the different items which are found to measure the same factor or dimension of effective teaching?

Factor analysis was used to determine the factor patters that JUIPEQ is designed to measure. Analyses were performed for the total rating forms (N =2,370). The analyses were conducted with the SPSS program with the following specific steps: First, the correlation matrix for all variables (28 items of JUIPEQ) was computed. At this step, variables (items) that do not appear to be related to other variables were identified. Second, the factor extraction-the number of factors necessary to represent the data and the method of calculating them-was determined. At this step, an initial inspection of the Eigen values indicated eight dimensions or factors greater than 1.0. Third, OBLIMIN rotation, i.e., the method of oblique rotation was performed. It focused on transforming the factors to make them more interpretable. Last, scores for each factor were computed. Table 2 shows the set of 28 evaluation items of the IPEQ and the factors of dimensions they were designed to define.

Table 2: Item loading on Eight Factors of JUIPEQ

Evaluation items (paraphrased [a]	Factor pattern Loadings I	II	III	IV	V	Vi	VII	VIII
I. preparation and Organization								
6. Clear presentation of the subject.	.827	.120	.065	.400	.480	.058	.435	.565
12. Knowledge of the subject	.791	.156	.118	.074	.160	.112	.540	.440
25. Preparation for class	.535	.435	.069	.343	.102	.130	.080	.154
II. Group Interaction								
10. Encourage Qs & answer	.069	.802	0.98	.035	.062	.050	.170	.191
18. Invites sharing of ideas	.139	.800	.256	.340	.263	.391	.156	.256
28. Encourage class discussion	.074	.382	.325	.074	.124	.191	.120	.110
III. Task Responsiveness & Enthusiasm								
13. Covers Content properly	.001	.028	.391	.025	.154	.531	.120	.070
19. Use the class period properly	.039	.300.	.269	.019	.057	.416	.044	.110
24. Give list of reference materials	.101	.263	.531	.024	.046	.397	.312	.044
11. Welcomed seeking help/advice	.033	.041	.529	.025	.026	.072	.783	.029
20. Text/manual/Module paper.	.101	.263	.263	.024	.046	.397	.312	.121
23. Interested in teaching	.238	.331	.254	.145	.321	.423	.433	.343
26. Appropriate use of inst. Materials	.059	0,62	.416	.083	.161	.529	.070	.197
IV. Professional Ethics								
4. Respect for law and order	.078	.046	.029	.541	.082	.098	.669	.156
7. Loyalty	.193	.132	.083	.780	.083	.254	.091	.263
21. Respect for students	.230	.112	.121	.665	.121	.110	.263	.102
15. Trusted by students	.437	.395	.276	.502	.276	.375	.456	.395
V. Rapport								
2. Reaction to students' Qs	.069	.052	.585	.041	.746	.060	.141	.065
27. Welcomed seeking help/advice	.003	.041	.025	.026	.699	.072	.378	.241
VI. Assessment Skills								
9. Tests emphasize course content	0.30	.563	.017	.032	.213	.620	.026	.107
14. Fair time allocation for tests	0.77	.470	.010	.061	.009	.608	.072	.047
16. Homework/assignments/feedbacks	.052	.041	.051	.081	.120	.607	.032	.023
17. clarified method of assessment	.039	.026	.056	.111	.122	.563	.032	.017
22. Eva. Methods fair/appropt.	.060	.111	.024	.015	.094	.529	.080	.061
VII. Punctuality								
3. Non-absenteeism	.096	.056	.007	.236	.029	.240	.254	.025
8. Accessible to students	.230	.112	.121	.121	.110	.091	.263	.091
VIII. Objectives and content clarity								
1. Clarification of sp. Objectives	.051	.013	.134	.714	.009	.002	.120	.858
5. Clarification of course plan & general objective	.074	.325	.061	.032	.077	.065	.221	.733

Note; Factor loadings in boxes are loadings to items designed to measure each factor. [a] All items descriptions are paraphrased.

Coefficient Alphas of JUIPEQ Factors

After the facto patterns were identified, the next important thing was the computation of coefficient alphas for each evaluation factor. Table 3 shows the value of coefficient alphas for each of the identified dimensions.

Table 3: Coefficient Alphas for the Six Dimensions of JUIPEQ

Evaluation Factor	Coefficient Alphas
I. preparation & interaction	0.90
II. Group interactions	0.93
III. Task Responsiveness & enthusiasm	0.80
IV. Professional Ethics	0.91
V. Rapport	0.69
VI. Assessment Skills	0.87
VII. Punctuality	0.63
VIII. Objective & content clarification	0.78

Item Rating Differences of instructors' teaching

In responding to the question, "which item(s) of effective teaching was/were rated high or low? Results were seen from global ratings for each item and evaluation results based social science, natural science, and language departments. Accordingly, the item to which instructors were rated highest was item number 12 which was about instructors' knowledge of the subject matter. The lowest rated item 8, was concerning accessibility to students.

In this study, variability of instructors' teaching effectiveness was also seen based on streams. The results were summarized below.

Table 4: Summary Ratings of instructors in different departments

Streams	Lowest rating	highest rating	Grand mean (for the 28 items)
Natural sciences	4.10	4.73	4.49
Social Sciences	3.46	4.48	4.15
Language	3.77	4.62	4.38

Dimensional Differences of Instructors' Evaluation Results

This research investigated the dimensional differences of instructors' performance evaluation of the faculty of education (JU) as the main objective of its study. The results are given below.

Table 5: Means and Standard Deviations of instructors' Ratings under the Eight Dimension of JUIPEQ

Streams		Dimensions							
		D1	D2	D3	D4	D5	D6	D7	D8
Natural Sciences	$\bar{\chi}$	4.56	4.54	4.48	4.54	4.60	4.36	4.47	4.49
Social sciences	$\bar{\chi}$	4.36	4.19	4.10	4.31	4.26	4.02	3.98	4.16
Language	$\bar{\chi}$	4.56	4.46	4.34	4.45	4.51	4.23	4.11	4.47
Totals	X	4.53	4.43	4.33	4.45	4.47	4.25	4.15	4.14
	δ	.444	.447	.582	.432	.492	.559	.667	.480

DISCUSSION

In studying the dimensionality of student ratings Marsh (1984) noted that student ratings are multidimensional. He argued that both common sense and a considerable body of empirical research indicate the multidimensionality of students' evaluations. Like the teaching they are designed to assess, student ratings vary along such dimensions as enthusiasm/skill, learning /accomplishments, structure, clarity, group interaction, individual rapport, grading/ examinations, breadth of coverage, workload/difficulty, instructor's knowledge, and others (Cohen, 1981). Dimensionality of students' evaluations is an important point to be considered in relation to student rating forms. Different surveys reviewed provide clear support for the multidimensionality of students' evaluations. Thus, instruments used to collect students' evaluations of teaching effectiveness should be designed to measure separate components of teaching effectiveness. This is in line with Jimma University Instructors Performance Evaluation Questionnaire (JUIPEQ).JUIPEQ (for education faculty) appears to measure some evaluation factors or dimensions of effective teaching. As it can be seen from Table 2, the questionnaire appeared to measure eight evaluation factors or dimensions. They could be labeled as: Preparation and Organization, Group interactions, Task Responsiveness and Enthusiasm, Professional Ethics, Rapport, Assessment Skills, Punctuality, and Objectives and Content Clarification. From the table it can be seen that items are loaded higher on the

factors that were designed to measure than other factors.

Although it is possible to categorize items into appropriate dimensions they are measuring, factor analysis is important in the development of student evaluation instruments and to test whether students are able to differentiate among different components of effective teaching (Marsh, 1984). And, the presentations of separate components of student ratings or evaluations enhance the diagnostic feedback to tractors.

Coefficient alphas consider the relative agreement among different items designed to measure the same factor (Marsh, 1982). Thus, it can be used as a sort of assessing whether the identified units of instruction are ideal enough as a source of feedback for the instructors and for other purposes.

As it is seen from Table 3, the coefficient alphas for the different evaluation factors of JUIPEQ vary between 0.63 and 0.93.

Teaching effectiveness may not be equally rated for all items of instructors' performance evaluation questionnaire (IPEQ). Instructors, in general, could be good at some items but not at other. According to this line of reasoning, means and standard deviations of instructors' evaluation vis-à-vis the 28 items of JUIPEQ (for education faculty) were computed. To this end, as it was seen from table 4, the item to which instructors are raged highest was item number 12 which was about instructors' knowledge of the subject matter. The lowest rated item, item 8, was concerning accessibility to students.

In this study, variability of instructors' teaching effectiveness was also seen based on streams. In all departments, item 8 (availability during consultation hours) was rated lowest whereas, natural science instructors are evaluated highest in their interests in teaching (item 23), social sciences in knowledge of the subject (item12) and language instructors in classifications of specific objectives of the subject in concern. In terms of broad global ratings averaged across the collection of the items, natural science instructors were rated better ($\bar{\chi}$ = 4.49) than the other departments; and language instructors stood second while social science instructors were rated least. Of course, in all items, natural science instructors were evaluated above 4.09, whereas there were seven items rated below 4.00 in social science departments.

As dimensions of teaching measures taxonomy of teaching behaviors they are the best indicators of effective teaching to which instructors should be given feedback they are also good indicators for the faculty or any other concerned body to see the essence of teaching-learning. So, understanding the dimensional differences, i.e., the dimensions of effective teaching to which instructors are evaluated good or bad, is very essential.

Teaching effectiveness couldn't be evaluated equally in all instructional units. Instructors could be rated good at one or more dimensions and poor at another. As results in Table 5 revealed it, instructors of the faculty were generally rated highest with respect to their preparation and organization (D_1) and evaluated lowest with their punctuality (D_7). This rating was also holds true for

social sciences and language departments (at stream level) while natural science instructors were evaluated lowest at their assessment skills (D_6).

CONCLUSION AND RECOMMEDATION

Jimma University has been using students' evaluation of instructors for feedback, academic and administrative decisions. However, there is no reported evidence for the dimensionality of the evaluation questionnaire, and at which dimensions and items of effective teaching instructors are good and weak. Thus, the major purpose of this study was to examine the variability of instructors, teaching effectiveness (in faculty of education) across the dimensions and items of effective teaching. Accordingly, the following research questions were posed for investigation.

- What dimensions or factors of effective teaching does the JUIPEQ measure?
- Is there a high level agreement among the different items which are found to be included a given dimension or factor?
- Which dimension(s) of effective teaching is/are rated high or low both at the faculty and departments' level?

Seventy nine instructors were involved in the study. Semester evaluation results of instructors in the 2004/05 academic year were taken. The collected data were analyzed using factor analysis, means and standard deviations. Results of the analyses revealed the following findings:

- The instructors' performance evaluation questionnaire appears to measure eight dimensions of instructor effectiveness: Preparation &

organization, Group Interactions, Task Responsiveness & Enthusiasm, Professional Ethics, Rapport, Assessment skills, Punctuality, and Objectives & content clarification.

- There is high level of relative agreement ($\alpha = 0.63 - 0.93$) among the different items found to be included under the same dimension or factor of effective teaching.
- Instructors of the faculty are evaluated across the 28 items in the following descending order: 12,2,23,13,28,4,6,3,25,21,24,10,5,19,15,1,27,7,14,22,18,17,9,26,16,11,20, & 8 – with maximum rating ($\bar{\chi}$ = 4.66) for item 12 and minimum rating ($\bar{\chi}$ = 3.80) for item 8
- In terms of broad global ratings across all items, natural science instructors are evaluated higher ($\bar{\chi}$ = 4.49) than language ($\bar{\chi}$ = 4.38) and social science ($\bar{\chi}$ = 4.15) instructors.
- Although the lowest rating (item 8) is the same in all departments, the highest evaluated items are some how different in the three streams. In natural sciences, item 23 is rated the highest ($\bar{\chi}$ = 4.73) whereas item 1 ($\bar{\chi}$ = 4.62) and item 12 ($\bar{\chi}$ = 4.48) have received highest ratings in language and social science departments, respectively.
- With regard to variability in dimensions of effective teaching, the instructors, generally, are evaluated highest with respect to their preparation and organization (dimension one). And, regarding lowest rating, natural science and language departments are evaluated lowest with their punctuality (dimension seven) while natural sciences are evaluated lowest at their assessment skills (dimension six). See the items included under these dimensions on Table 2.

Finally, though this research could not be considered perfect and final, the following recommendations could be made on the basis of the finding.

1. One of the reasons for evaluation of teaching effectiveness is that evaluation results provide bases for self-improvement by the faculty. Just as feedback is important for students in order to correct their errors, so also is feedback essential to faculty members. Therefore, the summary reports given to instructors should be based on the eight identified factor patterns of the evaluation questionnaire. Otherwise, broad global ratings averaged across a collection of heterogeneous items provide little diagnostic feedback and are difficult to interpret.
2. Although instructors of the faculty received good ratings in almost all items, there are some items and dimensions of effective teaching for which they need to have self improvement:

- The instructors should improve their availability during consultation hours; they need to have the courage in preparing teaching materials (texts, manuals, etc).
- Compared to other departments, social science departments received lowest overall ratings. So, they need to improve their teaching in some instructional units. Specially, they have to improve their teaching in some instructional units. Specifically, they have to improve themselves with these items: item 11 (welcoming those students seeking help and advice), item 16 (assignments and feedbacks), item 17 (clarified methods of assessment) item 18 (inviting idea sharing), and item 26 (appropriate use of instructional materials).
- At dimensions of effective teaching level, both social science and language

instructors should improve their punctuality; whereas natural science instructors are expected to improve their assessment skills.

ACKNOWLEDGEMENTS

Different people, in one way or another, have assisted me while doing this research. Specially, I would like to express my deepest gratitude to the staffs of faculty of education for their cooperation in the data collection processes. I equally wish to extend my indebtedness to the research and publication office (RPO) of education faculty for the financial support it rendered to carryout the study. I am still indebted to W/rt Hana Fekede for he devoted collaboration in typing and printing the manuscript.

REFERENCES

Assefa Berhane (1999). College students' evaluation of University teaching effectiveness. **IER, FLAMBEAU, 6, 32-47**.

Centra, J.A., (1973). Effectiveness of student feedback in modifying college education. **Journal of Education Psychology. 65, 345-401.**

Cohen, P.A. (1980). Effectiveness of student – rating feedback for improving college instruction: A meta- Analysis of findings. **Review of Educational Research,** 72,468-475.

_____ (1981). Student ratings of instruction and student achievement: a meta- analysis of multisection validity studies. Review of **Educational Research,** 51 (3), 281-309.

Marsh, H.W., (1982). SEEQ: A Reliable, Valid, and useful instrument for collecting students' evaluations of university teaching. **Journal of Educational Psychology**, 52, 77-95.

_____. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student course/inst characteristics. **Journal of Educational Psychology,** 75, 150-156.

_____ (1984), Students evaluation of university teaching: dimensionality, reliability, validity, potential bias and utility. **Journal of Educational Psychology,** 76, 707- 754.

Stumf, L. & Freedman, K. (1979). The Influence of student Evaluative Feedback on the Improvement of Teaching. **Review of Educational Research,** 50 (3), 123-200.

Weigel, J. (1971). Validity Concerns and Usefulness of Student Ratings of Instruction**. American Pschologist**. 52 (11), 1182-1186.