World Applied Sciences Journal 3 (1): 01-09, 2008
ISSN 1818-4952

The Hidden Value of Student Perspectives: Student Reasons Behind Their Ratings of Set Questionnaire Items

Nurdan Kalayci

Department of Educational Science, Faculty of Education, Gazi University, Besevler-Ankara- Turkey

Abstract: Student evaluations of teaching (SET) questionnaires are widely used in higher education to enhance the quality of teaching. More study in the area of qualitative analysis of student reasons for rating items may produce rich, multi-dimensional information, beyond quantitative ratings data, to give the institutions conducting SETs a better idea of student perception on quality of teaching. This paper reports the findings of a major Turkish state university's student evaluation of teaching questionnaire completed by 192 junior students from the Education Faculty.The qualitative data has been analyzed to determine the reasons given by the students for their ratings of closed-end items, the relationships between these reasons and the students' own perception of their initial interpretations of the questions put to them. Such rigorous analysis of student interpretations should be of value in developing better understanding of student perception about questionaire items and relationship between the reasons and the ratings given by students.

Key words: Student evaluations of teaching (SET) · student perception · open-ended questions · content analysis · ill-structured problem solving

INTRODUCTION

Student evaluations of teaching (SET) questionnaires are widely used throughout higher education establishments to enhance the quality of faculty teaching. Many studies were related to their importance in the academic performance environment [1-6]. Yet, despite the more than 2000 mainly quatitative studies [7] the credibility of the questionnaires remains a matter of concern.

Student evaluation feedback may be used to improve tecahing [8]. Taking evaluation as a part of a change process,Geis, [9] used an instructional feedback model to identify elements that had potential influence on feedback use by the recepient.

Geis Model: The questionnaire designer sends the message, via the questionnaire, to the student who then sends the message to the evaluator by means of the completed questionnaire. Geis [10] argued that the sender and the recipient of the message might alter the validity and interpretation of the message strongly. Therefore the stages, including the question formulation, the rating by the student and the analysis of the questionnaire items, have a structure which by its nature leaves considerable probability for different and subjective interpretations by participants (designer, student and evaluator).

The reliability and validity of the message exchange process depends on conveying the message in an intended manner. Thus in this study, the evaluation process of teaching was considered as the process for the exchange of messages. During a communication process, conveying and interpreting the message in an intended manner can increase the credibility of the communication process.

The key message process begins before the student [the information processing system) completes a SET questionnaires and incorporates two stages which are not contained in Geis' model – the stage during which questions are designed and the stage during which the SET are distributed (communicated) to the students [the 'task environment'). An Extended Instructional Feedback Model was therefore utilised for explaining of message exchanging process during questionnaire completion by students.

Student Evaluation of Teaching

Quantitative literature: The literature about SET questionnaires can be classified into four aspects: Students; Faculty; Courses; and Other Variables

Corresponding Author: Dr. Nurdan Kalayci, Department of Educational Science, Faculty of Education, Gazi University, Besevler-Ankara- Turkey

[administrative, environmental, etc.) Considerable literature exists in respect to quantitative research on student-related variables [11, 14-22].

However, student-related variables such as the impact of personal interest, motivation, personal characteristics, potential bias that might vary depending on the individual personal differences, have been subjected to fewer research studies [18, 23-29]. The principle method of these studies, however, has been quantitative analysis.

Qualitative literature: There are a limited number of other researchers who have investigated how students interpret the items of questionnaires qualitatively [30, 33-25]. Aiming to close this qualitative gap in the literature, Alderson [35] has been among the most pioneering researchers in understanding and explaining how students interpret the items in questionnaires.

The research of Low [30, 33] provides further explanations among the studies on expression, style, format and the scaling of items in questionnaires. Alderson [35] and Low [32] agreed that research on this subject remained insufficient.

Alderson, [35] provides further valuable insight to the problems that might be encountered during the process of exchanging messages between sender 2 and recipient 2, from step 3 to step 5. Block [36] conducted a similar study as a result of his interest in Alderson's findings, leading him to explain:

"While I cannot possibly know what the original designers of this particular questionnaire meant when they worded items as they did. I do know that their interpretation would surely be more narrow than the collection of multidimensional interpretations provided by the respondents involved in this study [424]".

Problem Solving: Kitchener, [37] states that when contradictory assumptions, evidences and ideas, conclude with different solutions, this amounts to ill-structured problem solving. In the questionnaires, whether they may contain closed or open ended items or both, the fact that the students suggested alternative solutions with alternative reasons for items in the questionnaire and that they stated divergent types of thinking, bears the hallmarks of ill-structured problem solving during rating the items.

As Jonassen, [38] emphasised, in the process of solving ill-structured problems, the individual differences in students, such as familiarity, blurriness, knowledge about the process they are participating in, personal cognitive styles, self awareness and self-confidence, motivation and determination, result in significant variations in their solutions. When Jonassen's[38] definition of a complex problem is taken into consideration, the task of rating items in the questionnaire is also complex. The complexity becomes apparent when all things are considered around on the subject matter of the items and all the particular features relating to the field of the items which are to be rated. The students are required to put forward a reason or reasons to support the solution they have chosen during this complex decision-making process. These reasons might be of cognitive and or non-cognitive origin.

Jonassen and Tessmer [39] suggest the cognitive and metacognitive processes are necessary in solving ill-structured problems. However, they also emphasize that as these are insufficient on their own, so emotional and connotative processes are also necessary.

By using this point of wiev, the studies of Low [30, 33]; Solas [34], Alderson [35] and Block [36] can be considered as steps towards using ill-structured problem solving explanation for the results based on qualitative data obtained from reasons of questionnaire ratings. Moreover, Robertson [25] used the problem solving to analyze the reasons behind the ratings given by students to questionnaire items adding further understanding of the research problem at issue.

Low [33] explained the phenomena:

"While it is certainly important to know if particular questions do or do not cause statistically significant response biases, it is rather more important to find out what leads respondent to answer as they do [2]."

Based on the approach that Robertson [25] used, the qualitative data obtained from this study is also interpeted by using ill-structured problem solving approach.

Purpose: The purpose of this study is to make a qualitative assessment of the reasons given by educational faculty junior class students for their ratings to the closed-ended items in the questionnaire to determine perceptions of the students for items.

To achieve this purpose, the study focuses on two sub-problems:

- To what extend a specific item wording effected the students perceptions ?
- What is the relationship between reasons and ratings given by students for a specific item?

Such rigorous analysis of reasons is important as it leads to multi-dimensional and rich information, beyond quantitative ratings data, to provide the researchers studying SETs a better understanding of student perceptions.

METHODS

A qualitative case study design was used to answer the research question presented above. The topic to be studied was complex human and institutional relationship, so qualitative method was more preferable to qualitative methods and also such phenomena was not easily translated to numbers [40].

Sample: The SET questionnaire of the Gazi University Education Faculty was conducted among 192 junior students from the three faculty departments of Elementary Education Class Teaching, Pre-School Education and Social Sciences Teacher Education., differing from the study sample of Robertson [25]. The students were selected and sampled using a purposive sampling approach which is suitable for qualitative method, during the Fall Semester of 2004-2005. The junior students' more extensive experience of the faculty teaching was preferred to the freshmen and sophomore students in evaluating effective teaching. Senior students were not used in the study as the low number of volunteers from the group could not produce a credible analysis of their perspectives.

Instrument The study used an evaluation questionnaire for obtaining the qualitative data. Twenty items were used in the questionnaire which is exactly the same as the university used as an instrument for evaluation of teaching (Table 1). Likert's 5-point scale was used in the study for the students to rate items. The questionnaire used in this study, different from the original instrument used by the university, contained a space next to each item in which students were able to explain or rationalise the reasons for their ratings, thus producing the qualitative data and the items contained in both instruments are the same. At the end of the semester, in the course time, the students were given the questionnaire to complete.

The purpose of the study were explained to the students before they were asked to complete the questionnaires. They were asked not to rate the items that they did not understand, had no knowledge of, or which were irrelevant to the content of their courses. No personal identifiers of the students were requested as the study was based on anonymity. The students were requested to provide the reasons, by writing, for the choices they made, in the free spaces allocated on the questionnaire.

Qualitative Data Analysis Method: Conventional content analysis was used for the analysis of the qualitative data. Van Manen [41] denoted that the researcher in that kind of analysis tries to discover the 'experiences' of the students from their comments. The students provided a

Table 1: The 20 items used in the Gazi SET questionnaire

1.	He/she clearly explains the content and the objectives of the course and subjects at beginning of the term.
2.	He/she makes preparation for the content of the course before she/he comes to the class.
3.	He/she is meticulous about beginning and ending time of classes.
4.	He/he is master of the content of the course.
5.	He/ she use proper voice tone, language and expressions during the course.
6.	He/she makes the class interesting and enjoyable.
7.	He/she uses teaching objectives properly and effectively.
8.	He/she never mentions too much about the subject which are out of scope of the content.
9.	He/she manages the class without using grades as threat.
10.	He/she creates a democratic environment allowing students to ask questions and mention their opinion during the course.
11.	He/ she has successful communication skills.
12.	He/ she treated the students equally.
13.	He/she never mentions and is never under the influence of his/her personnel problems.
14.	He/she is always a role model with his/her physical appearance and his/her behaviours.
15.	He/she leads the students to research activities.
16.	At the end of each class time, he/she always summarize the subjects and daily content and mention about important points.
17.	If it is necessary, he/she devotes time to the students out of the class.
18.	Examination questions reflect the course content and objectives.
19.	He/ she evaluated the exams objectively.
20.	He/she grades examinations and assignments on time and warns students about their mistakes and deficiencies.

Table 2: The matrix of category-ratings for Item 6

	Class A						Class B						Class C						General Total					
Reasons	1	2	3	4	5	T	1	2	3	4	5	T	1	2	3	4	5	T	1	2	3	4	5	T
s/he sometimes tells about an idea or one of her/his nice memories related with the subject	0	0	0	3	21	24	0	0	0	8	8	0	0	0	0	0	6	6	0	0	0	3	35	38
s/he tells about practical jokes during class that do not disturb objectives of the course	0	0	0	0	5	5	0	0	0	0	3	3	0	0	0	0	2	2	0	0	0	0	10	10
it will be better if s/he instructs faster	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2
s/he uses the materials, hands out Xeroxes	0	0	0	1	1	2	0	0	0	0	9	9	0	0	0	1	0	1	0	0	0	2	10	12
s/he gives educational examples	0	0	2	1	4	7	0	0	0	0	5	5	0	0	0	0	0	0	0	0	2	1	9	12
her/his methods are suitable with the course	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
s/he does not use the materials; there is no need for this indeed	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	2
s/he does not use the materials, but s/he is genial and charming	0	0	1	3	2	6	0	1	1	0	1	3	0	0	1	0	0	1	0	1	3	3	3	10
s/he gives examples (e.g. related with Fenerbahche football team) and relates course subject with daily life	0	0	2	0	2	4	0	0	0	0	23	23	0	0	0	3	6	9	0	0	2	3	31	36
s/he is genial and sympathetic	0	0	0	0	1	1	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	4	4
s/he has very nice mimics and jests	0	0	0	0	1	1	0	0	0	0	2	2	0	0	0	0	1	1	0	0	0	0	4	4
s/he has no temper and s/he does not act formally	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1
her/his jokes are not funny	0	0	0	0	1	0	0	0	0	2	1	3	0	0	0	0	0	0	0	0	0	2	2	3
s/he does not offend us	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	3	3
her/his voice is not monotonous, but attractive	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1
her/his motivation is high	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1
s/he is so good at lecturing, but s/he does not get involved during our presentations	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	0	0	0	0	4	4
s/he gives interesting examples related with the course, but we prepare the learning materials	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1
our instructor acts like a player and play different roles during class hours	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	2	2
s/he acts naturally and makes jokes	0	0	0	1	2	3	0	0	0	1	8	9	0	0	1	1	0	2	0	0	1	3	10	14
s/he provides information as if s/he is in a conversation with us, but s/he sometimes depart from the subject	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
s/he is open to different ideas and suggestions	0	0	0	0	2	2	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	2	3
s/he signs songs or read poems and addresses us by our names	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
course is too monotonous	0	0	0	0	0	0	0	0	1	0	0	1	0	4	4	3	0	11	0	4	5	3	0	12

total of 3517 comments against the 20 items in the questionnaire. The term "unit" is used in this study to mean each original comment provided for each item from each student. The term "category" refers to each obtained criterio (reason) for the ratings deducted from the comments. Emergent Haney *et al*., [42], Latent, Frankel & Walen, [43] coding approach was used to find categories for ratings arising for each item. Ideas or 'meanings' in the units were naturally and inductively obtained from comments for each item.

Firstly, two researchers agreed on qualitative content analysis and then decided to determine the categories to serve for the classification of ideas in the units, together with counting their frequency, thereby agreeing on a common view before reading the data.

The researchers then, independently from each other, read all 3517 student comments to identify the general categories based on each item, facilitating coder agreement. Five of the student participants agreed to assist the two researchers interchangeably in order to aid understanding and analysis of the student comments in building the categorizations (collaboration). The researchers compared the categories and frequencies they had determined for each item during the above process and then jointly agreed the main categories for each item. The agreed categories form the 'criteria' which the students used as the basis for their ratings for each item.

A sample of twenty randomly selected questionnaires and their list of main categories were submitted for peer debriefing to two professors from the Faculty to check whether the categories of every item properly represented the main ideas of units from which these categories were obtained.

In line with this peer debriefing, the researchers read the qualitative data and related categories independently from each other for the second time within a period of three months. They added new categories if needed.

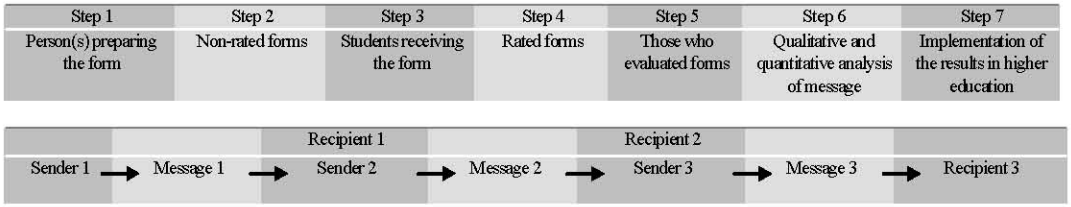

Fig. 1: The Extended Instructional Feedback Model

For the purpose of analyzing qualitative data, matrices of category-rating analysis, which resemble the matrix suggested by Lewis [44] for open-ended items, were used for open-ended items in the survey. The researchers entered the frequencies of rating in relation to related categories in the matrices.

The researchers compared their matrices of category-rating analysis based on every item, thereby using the different categories for each item and different frequency figures. Consequently, they established a common matrix of category-rating analysis for every questionnaire item. At the end of this process, the differences which appeared in the matrices were removed with the agreement of the researchers and the five assisting students.The matrices were designed to explicitly present the data of the three separate classes. The items to be analyzed were written down at the top of every matrix. Each line written for every category in the matrix displayed the frequencies equal to the scores in the Likert scale and how many times the related category appeared. The last column group showed summation of frequency data from all three classes.

For each item in the questionnaire, 15-25 reasons (categories) were identified. Despite the fact that all the reasons within the framework of behaviour that the items intended to measure were analyzed, the researchers focused mainly on data producing difference on interesting variables or on data that could not be explained with Likert's scale.

RESULTS

Results obtained from this study can be classified into two sets namely, "wording of the items related" results and "students related" results. To give a clear and coinsize point of view about the results and also to stress the importance of the item wording in questionnaries, the results given here are constrained with "wording of the items" related results. Since, Coffey and Gibbs, [45] sated that 'virtually all student feedback questionnaires used in UK are "home-grown" '. The pharase 'home-grown' is also used by Kolitch and Dean, [46] when they refer to the range of SET instruments used in much of the USA.

Overlapping in the item caused uncertainty: Item 20 *He/she grades examinations and assignments on time and warns students about their mistakes and deficiencies*, it is apparent that students were being asked to measure with only one rating up to six different behaviours. The principle behaviour intended to be measured is an unknown to the students. This created uncertainty in the student responses: 'He/she reads the exams on time. He/she does not give the feedback individually. He/she gives the feed back when He/she remembers'; 'He/she reads the exams early and gives feedback', 'He/she reads the exams late'; 'He/she gives/does not give feedback', 'He/she gives homework, He/she does not give feedback', 'He/she does not read the homework on time, but gives feedback'; 'There is no homework for this course', 'He/she reads the exams on time, but gives feedback when requested'. The important point here is that the students rated a single behaviour or some behaviour contained in the overlapping items, which was important for them individually so they did not all rate the same behaviour. The same problem was seen for items 1, 7, 14 and 16. The questionnaire designers and evaluators might have thought that the students rated based on the entire item. The results have contradicted this assumption. The overlapping items might have caused the emergence of individual differences during the ill-structured problem solving process. Hong, [47] explained that problem solvers had to organize and adjust their cognitive efforts with different degrees towards different directions when the objectives and process alternatives are not clear. Cashin,[21] defined "validity" in such questionnaires as the extent to which the questionnaire items identify some aspects of effectiveness of teaching. Based on this definition, the validity of items in 1, 7, 14, 16 and 20 can be questioned. When the reasons for these items are analyzed, it can be surmised that each student rated one of the parts of the

items, or that they measured subject matter which was beyond the intended dimension to be measured in the item. The same problematic wording can be encountered in the questionnaires and their items in the American and British universities. For example, in the question bank of the American University, Washington D.C., item 201 and 212 under section "instructor ratings" are worded as "The instructor speaks clearly and can be heard without difficulty" and "Instructor is skilful in observing and responding to students' reactions", respectively.

Some parts of the conditionally worded items were focused and rated: Students concentrated on some parts of the conditionally worded questionnaire items and rated accordingly. For example the students concentrated on the conditional part of item 9 and relative part of item 10. In item 9, students were asked whether the faculty manages the class without the anxiety of grade. The reason of seven students is, 'Grade is not more important than learning'. The reason of seventeen students is, 'I do not have an anxiety of grade as I learn the lesson very well'. The comment of one student is, 'The Faculty says that solving this question shall bring 5 points to the midterm grade. This gives me a negative feeling'. The reason of another student is, 'Anxiety of grade appears during the midterms'. And finally another student reasoned, 'Since there is no oral exam, there is not anxiety of grade'. The students rated in accordance with these reasons and also were observed to rate based on the first section of the item, that is, the practices concerning grades and on their feeling about this issue. Item 10 *He/she creates a democratic environment allowing students to ask questions and mention their opinion during the course.* Out of 100 reasons given, the reasons of twenty seven students included, 'He/she enables the students to reveal their ideas and answers the questions enthusiastically'. 32 students reasoned, 'We can ask him/her what we don't understand'. These students therefore rated these items based on the relative part of the relevant item. For both items, the students rated by focusing on the different parts of items instead of whole items. While rating, the students concentrated on the parts concerning "anxiety of grade" and "opportunity to ask questions", which are concrete and relevant experiences on which they have knowledge or strong feelings.

Ambiguous wording in some items increased the number of reasons: Students gave the items 2, 4, 6 and 11 an excessive number of reasons compared to other items. For example, the student reasons for item 4 *He/she is master of the content of the course*, included: 'He/she answers our questions clearly'; 'He/she relates the daily life affairs with the course subject, this makes me understand the course well'; 'He/she is fluent in lecturing'; 'He/she begins to explain the easiest problems to the most complicated one'; 'He/she updates himself/herself about the new developments'; 'He/she has many sources', 'I do not believe that explanations are thorough'; 'He/she lacks course material during the class'; 'Student lectures, peer readings and the use of un-qualified or assistant teachers gives us a superficial understanding of the course', 'His/her teaching method suits me, He/she is knowledgeable in 'His/her field'.

A great majority of the reasons explained are the behaviours that might take place in the literature within the scope of a teacher being knowledgeable. The important result here is that there were 29 different reasons given for the item The excessive number of reasons for rating might be explained by the item being worded ambiguously, which can trigger the use of unknown problem solving elements and, as Wood, [48] and Kitchener,[49] argue, this might have created the probability of multiple solutions or non-solutions in ill-structured problem solving.

Another possible explanation for the result is the lack of a common list of definitions and peculiarities for the behaviours within the scope of the items in question [21]. In other words, each student may give comments more in line with 'His/her background rather than common attributes.

Some items were misunderstood because of unclear wording: Item 7 *He/she uses teaching objectives properly and effectively.* The item's construction seems to create missed meaning or misunderstanding. For example, student reasons included, 'He/she does not threaten us with grades'; 'He/she is an effective teacher with both discipline and sincerity'; and, 'He/she gave advice on what to be careful about when we became teachers'.

We can see that these reasons do not appear to be related to the item. The first reason is more about the characteristics of a successful teacher, the second about successful class management and the third about the course content. 16 students giving these reasons rated 5. The behaviours the students chose to rate and gave high ratings were not the criteria on the effective and timely use of course objectives. A clear result for this item is the fact that although twenty nine students said they did not understand this item, these students all rated the

item - four giving a rating of 3, fifteen giving a rating of 4 and ten giving the top rating of 5. The comments of three of these students are significant in showing how the students decided on their scores: 'The teacher is successful in general. I did not understand this item, but I rated because of 'His/her successful way of lecturing'; 'He/she is good in communication, but the question is not understandable. However, I rated because He/she likes us'; and, 'I wanted to rate because He/she is successful in general'. Other student responses to this item included twelve who did not rate and did not comment and twenty students who said they had not rated because they didn't understand the item. The number of students who rated without understanding are higher than those who did not understand and did not rate. The same result was observed in Robertson's study, [25] but in that study the students rated the item low because of misunderstanding, contrary to the higher scores in this study.

The items in the Gazi questionnaire can be likened to those evaluated by Tagomori and Bishop p.75, [49] who identified that that three quarters of the items in questionnaires were complicated, hardly certain, or subjective. The finding that some of the Gazi students who misinterpreted the item rated nonetheless based on characteristics of the teacher that impressed them, tends to support the view of Gick, [50] and Jonessen, [38] that when students faced a problem they were not familiar with, the students interpreted the problem area based on their experiences and then tried to solve it.

CONCLUSIONS

As Robertson, [25] stated, understanding and rating a questionnaire item can be considered as an ill-structured problem solving process. Bearing in mind this point of wiev, definiton or description of the problem, which can be considered as sending a message to the receiver, is an important step for understanding of the problem. In other words, according to the Geis, [9]model, sending a clear message to receiver is an important factor for the message to be understood in an intended manner by the receiver.

The use of overlapping, conditional and broad meaning wording for items, or the use of wording which was difficult to understand, is revealed to have had a strong effect on the student perception process during the rating of questionnaire items. This process can be explained as ill-structured problem solving and is determined to be the cause of the variation of the rating scale, the emergence of features about which items did not intend to measure and the halo effect. This assessment reaffirms the findings of Marsh and Roche, [51] that items with inappropriate or bad wording may not provide valuable information and what is to be measured will be unclear with the average ratings obtained from such items. According to these results, accurate wordings of items are important for the credibility of questionnaires in order to accumulate the student rating and rating criteria on one common point. If the results of the questionnaire were examined using the "Extended Instructional Feedback Model" (EIFM) significant differences could be found among students with their message perceiving (problem understanding) manner For example, the issue or issues intended to be measured in the items by the questionnaire designers are constrained and limited, compared to what the students sought to measure in rating and reasoning the same items. It is unlikely that the designers and senders of the original messages planned to measure the extent of variable and extreme behaviours produced by the items in the questionnaire. However, this un-intended outcome has the advantage of bringing out data about the features of the teaching which the students paid most attention to or valued. The opinions of students, who are the main source of the data derived from the evaluation questionnaires, could be valued and directly included into an evaluation process. In light of the findings of this study, it is recognised that requesting students to state their reasons when rating the items enables the formative evaluations to be more functional and helpful to the overall objective of such surveys which was one of the main goals of Geis, [9]Model i.e. to improve the quality of faculty and teaching and by implication, the quality of the learning experiences and academic outcomes for the students.

REFECENCES

1. Seldin, P., 1993. The use and abuse of student ratings of professors. The Chronicle of Higher Education, 39(3): 40.
2. Abrami, P.C., 1989. How should we use student ratings to evaluate teaching? Research in Higher Education. 30: 21-227.
3. Wagenaar, T.A., 1995. Student evaluation of teaching: some cautions and suggestions. Teaching Sociology, 23 (1): 64-8.
4. Abrami, P.C., S. d'Apollonia and P.A. Cohen, 1990. Validity of student ratings of instruction: what we know and what we do not know. Journal of Educational Psychology, 82(2): 219-231.

5. Hobson, S.M. and D.M. Talbot, 2001. Understanding student evaluations, College Teaching, 49(1): 26-31.
6. Grant, H. 1998., Academic contests: merit pay in Canadian Universities, Relations Industrielles / Industrial Relations, 53(4): 647-664.
7. Centra, J.A., 2003 Will Teachers receive higher student evaluations by giving higher grades and less course work? Research in Higher Education, 44(5): 495-518.
8. Murray, H.G., R.B. Jelley and R.D. Renaud, 1996. longitudianal Trends in Student Insructional Ratings. Paper Presented at the Anual Meeting of the American Educational Reserh Association. New York.
9. Geis, G.L., 1991. The moment of truth: feeding back information about teaching. New Directions for Teaching and Learning, 48: 7-19.
10. Geis, G.L., 1986. Formative feedback: the receiving side. Performance and Instruction, 25(5): 3-6.
11. Feldman, K.A., 1976. The Superior college teacher from the student's view, Research in Higher Education, 5: 243-288.
12. Feldman, K.A., 1977. Consistency and variability among college students in rating their teachers and courses: a review and analysis. Research in Higher Education, 6(3): 223-274.
13. Feldman, K.A., 1978, "Course characteristics and college students ratings of their teachers: what we know and what we don't". Research in Higher Education, 9: 199-242.
14. Feldman, K.A., 1993. College students' views of male and female college teachers: part ii- evidence from students' evaluations of their classroom teachers. Research in Higher Education, 34(2): 151-211.
15. Howard, G.S., and S.E. Maxwell, 1980. Correlation between student satisfaction and grades: A case of mistaken causation? Journal of Educational Psychology, 72: 810-820.
16. Aleamoni, L.M., 1981. Student Ratings of Instruction. In J. Millman (Ed.), Handbook of teacher evaluation. 110-145. Beverly Hills, California: Sage.
17. Centra, J.A., 1993. Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. San Francisco: Jossey-Bass.
18. Marsh, H.W. and M. Dunkin, 1992. Students' evaluations of university teaching: A multi-dimensional perspective. In J. C. Smart (Ed.), *Higher Education*: *Handbook on Theory and Research,* New York: Agathon Press. 143-234.
19. Braskamp, L.A. and J.C. Ory, 1994. Assessing Faculty Work: Enhancing Individual and Institutional Performance. San Francisco: Jossey-Bass.
20. Cashin, W.E., 1990. Student ratings of teaching: recommendations for use kansas state university. IDEA Paper no:22. Kansas State University, Center for Faculty Evaluation & Development.
21. Cashin, W.E., 1995, Student ratings of teaching: the research revisited. IDEA Paper no: 32. Kansas State University, Center for Faculty Evaluation and Development.
22. Sixbury, G.R. and W.E. Cashin, 1995b. IDEA Technical Report no. 10: Comparative Data by Academic Field. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
23. Kamber, D. and A. Wong, 2000. Implication for evaluation from a study of students' perceptions of good and poor teaching, Higher Education, 40(1): 69-97.
37. Robertson, I., 2004. Student perceptions of student perception of module questionnaires: questionnaire completion as problem solving, Assessment and Evaluation in Higher Education, 29(6): 662-669.
24. Chen, Y. and L.B. Hoshower, 2003. Student evaluation of teaching effectiveness: An assessment of student perception and motivation. Assessment & Evaluation in Higher Education, 28(1): 71-88.
26. LaForge, M.C., 2003. Student mood and teaching evaluation ratings. Journal of the Academy of Business Education, 4.
27. Abrami, P.C., R.P. Perry and L. Leventhal, 1982. The relationship between student personality characteristics, Teacher ratings and student achievement. Journal of Educational Psychology, 74(1): 111-125.
28. Hoffman, F.E. and L. Kremer, 1980. Attitudes toward higher education and course evaluation, Journal of Educational Psychology, 72(5): 610–617.
29. Abrami, P.C. and D. Mizener., 1983. Does the attitude similarity of college professors and their students produce 'bias' in course evaluations? American Educational Research Journal, 20: 123-136.
30. Low, G., 1988. The semantics of questionnaires rating scales' Evaluation and Research in Education, 2: 69-79.
31. Low, G., 1991. Talking to questionnaires pragmatic models in questionnaire design. In B Heaton, P Adams and P. Howarth (eds.) 1991 Sociocultural Aspect of English for Academic Purposes. London: Macmillan.

32. Low, G., 1996..Intensifiers and hedges questionnaire items and the lexical invisibility hypothesis. Applied Linguistics, 17(1): 1-36.
33. Low, G., 1999. What respondents do with questionnaires: accounting for incongruity and fluidity. Applied Linguistics, 20(4): 503-533.
34. Solas, J., 1990. Effective teaching as construed by social work students. Journal of Social Work Education, 26: 45-154.
35. Alderson, J.C., 1992. Validating questionnaires. Centre for research in language education. Working Papers. No. 15. University of Lancaster.
36. Block, D., 1998 Exploring interpretations of questionnaire items. System, 26(3): 403-425.
37. Kitchener, K.S., 1983. Cognition, metacognition and epistemic cognition: a tree-level model of cognitive processing, Human Development, 4: 222-232.
38. Jonassen, D.H., 2000 Integrating problem solving into instructional design. In R.A. Reiser & J. Dempsey (Eds.), Trends and Issues on Instructional Design and Technology. Upper Saddle River, NJ: Prentice-Hall.
39. Jonassen, D.H. and M. Tessmer, 1996/1997. An outcomes-based taxonomy for instructional systems design, evaluation and research. Training Research Journal, 2: 11-46.
40. Peterson, M.W. and M.G. Spencer., 1993. Qualitative and Quantitative Approaches to Academic Culture: Do They Tell Us the Same Thing. In Smart, J.C. (ed) Higher education: Handbook of theory and research. IX. New York.Agathon Press.
41. Van Manen, M., 1990. Researching Lived Experience: Humanscience for an Action Based Pedagogy. London: The Althouse Press.
42. Haney, W., M. Russell, C. Gulek and E. Fierros, 1998. Using student drawings to promote middle school improvement, Drawing on Education. School in the Middle, 7(3): 38-43.
43. Fraenkel, J.R. and N.E. Wallen, 2000. How to Design and Evaluate Research in Education. McGraw-Hill Companies.
44. Lewis, G.K., 1991. Making sense of student written comments, New Direction for Teaching and Learning, 87: 25-32.
45. Coffey, M. and G. Gibbs, 2001. The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education. Assessment & Evaluation in Higher Education, 26(1): 89-93.
46. Kolitch, E. and A.V. Dean, 1999. Students' Ratings of Instruction in The USA: Hidden Assumptions and Missing Conceptions About 'Good' Teaching, Studies in Higher Education, 24(1): 27-42.
47. Hong, S.N., 1998. The Relationship Between Well-Structured and Ill-Structured Problem Solving in Multimedia Simulation. Unpublished Doctoral Thesis. The Pennsylvania State University, The Graduate School College of Education.
48. Wood, P.K., 1983. An inquiry systems problem structures: implication for cognitive developments. *Human D Development,* (26), 249-265.
49. Tagomori, H. and L. Bishop, 1995. Sudent evaluation of teaching: flaws in the instruments, Thought and Action, 11: 63-78.
50. Gick, M.L., 1986. Problem solving strategies. Educational Psychologist, 21 (1&2): 99-120. UnLoi em noi cho tinh chung ta, nhu doan cuoi trong cuon phim buon. Nguoi da den nhu la giac mo roi ra di cho anh bat ngo... http://nhatquanglan.xlphp.net/
51. Marsh, H.W. and L.A. Roche, 1997. Making students' evaluations of teaching effectiveness effective. the critical issues of validity, bias and utility. American Psychologist, 52(11): 1187-1197.