

Analysis of Evaluation Data Collected on Likert Type Items: Humanities-Courses

Raghavendra Dwivedi^{1*}, Nagendra Nath Pandey²

¹Shri Jagdish Prasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan

²P.D.Lions College of Commerce & Economics [MS41], Malad (West), Mumbai

Email: *raghavendra.jjtu@gmail.com

How to cite this paper: Dwivedi, R. and Pandey, N.N. (2021) Analysis of Evaluation Data Collected on Likert Type Items: Humanities-Courses. *Journal of Data Analysis and Information Processing*, 9, 102-121. <https://doi.org/10.4236/jdaip.2021.92007>

Received: September 19, 2020

Accepted: May 4, 2021

Published: May 7, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

To improve high quality and/or retain achieved high quality of an academic program, time to time evaluation for quality of each covered course is often an integrated aspect considered in reputed institutions, however, there has been little effort regarding humanities courses. This research article deals with analysis of evaluation data collected regarding humanities course from a College of Commerce & Economics, Mumbai, Maharashtra, India, on Likert type items. Appropriateness of one parametric measure and three non-parametric measures are discussed and used in this regard which could provide useful clues for educational policy planners. Keeping in view of the analytical results using these four measures, regardless of the threshold regarding satisfaction among students, overall performance of almost every subject has been un-satisfactory. There is a need to make a focused approach to take every course at the level of high performance. The inconsistency noticed under every threshold further revealed that under such poorly performing subjects globally, one needs to analyze merely at the global level item. Once the global level analysis reveals high performance of a course, then only item specific analysis may need to be focused to find out the items requiring further improvements.

Keywords

Evaluation Data, Humanities Courses, Likert Type Items, Parametric Measure, Non-Parametric Measures

1. Background

The monitoring of a developed as well as a developing educational program requires reliable evidence drawn through inferences from analytical results of a number of evaluations of activities under that educational program. For a sustain-

able quality of an educational program, every academic institution including those with accredited educational program needs to derive policy oriented clues through such evaluations from time to time. The institutional evaluation studies might use various evaluation programs including course evaluation survey (CES). Such institutional evaluation studies are often adapted in most of the high quality institutions including those from developed countries as well as developing countries like India. As such, to ensure the quality in higher education, students' course evaluation surveys may not be avoided [1] [2].

Under CES, often semi-structured questionnaires involving a mix of mostly Likert type items and few open-ended questions are used. The appraisal of the analytical results involves three major issues namely accuracy, easiness of understanding of the analytical results and regular assessments of the prevailing evaluation system [3] [4] [5]. The present paper mainly deals with analytical as well inferential issues related to items. Course evaluation surveys mostly involve data collection on Likert type items, for example, from 1 (Strongly Disagree) to 5 (Strongly Agree). Before addressing the analytical approaches often used in item by item analysis under course evaluation surveys, one needs to differentiate between "Likert Type Item" and "Likert Scale". The data on a Likert type item will remain to be on an ordinal scale. On the other hand, in case of a number of Likert type items on the same number of agreement grades, summation of grading on every individual Likert type items in an evaluation questionnaire provides data on a likert scale [6] [7]. Hence, to ensure appropriateness in analysis and accuracy in related inferences, one needs to differentiate between "Likert Type Item" and "Likert Scale" [3].

To analyze Likert type items individually to derive policy driven inferences, an inappropriate use of arithmetic mean even by reputed global universities may often be seen in the literature. Also, the usefulness of mean of scores on a Likert type item becomes questionable. As a mixed practice, some of the researchers report mean with standard deviation whereas others report mean even without standard deviation. To analyze data, the used analytical method needs not to be guided by the fact that everyone is using it [7]. Rather, to be theoretically appropriate, analytical methods need to be driven by scales of measurements of that data [8]. Further, there are researchers who have been emphasizing the need of statistical measures which not only follow statistical principles but also might be reliable, easier to interpret and stable regarding policy implications.

As obvious, being on ordinal scale due to considered Likert type items [9] [10], item by item analysis may be more appropriate if non-parametric approach (e.g. Median, Quartiles, Percentiles, Deciles) is used. Otherwise, the analytical results might be deceptive. For example, just to emphasize inappropriateness of the parametric approach (Mean, Standard Deviation), the mean of ordinal scores (strong disagreements to strong agreements) may provide a misleading impression of average agreement [3]. In contrary to this, consideration of measures of location like Median, Decile, Percentile, Quartile or Cumulative percentage with

specific grade points will be able to provide more clarity regarding agreement level. Obviously, this approach will provide clues with clarity and high policy implications. Hence, from the academic planning point of view, the non-parametric measures are expected to provide better understanding of results. As a matter of fact, the above-described appropriate methods have already been adopted by some of the universities. Some of them analyze the data on Likert type item appropriately using non-parametric approach along with inappropriate parametric measures. Likewise, the analysis of data on Likert scale might be appropriately analyzed by using parametric approach (e.g. Mean, Standard Deviation). Recently, there have been suggestions to involve correct measures in this regard [3].

Under the present study, the developed course evaluation survey (CES) questionnaire includes 15 items covering different domains such as items related to start of the course (items 1 - 3), those related to its teachers (items 4 - 10), those related to the institutional infrastructure (items 11 - 13), that related to acquired knowledge by the student (item 14) and overall satisfaction with the course quality (Global item, item 15). The institutional policy planners may often obtain valuable evidence through item by item analysis of the Course Evaluation Survey (CES) for related educational developments [11]. Therefore, the present paper describes appropriate approaches to item by item analysis of likert type items through applications on actual data sets. A continuing Academic development program may be strengthened further through support from quality evidence through evaluation results. For translational/transformational evidence, the analytical methods to deal with evaluation data need to be theoretically robust and also should have problem-solving potentials. In other words, the used analytical methods need to have appropriateness in application, accuracy in results and related inferences, ease of understanding, policy-driven inferences for educational improvements and helpful to optimize the use of allocated resources for quality in higher education [3] [12].

To begin with, to develop and sustain high quality academic programs within an institution, competitiveness among various courses within each of the academic programs needs to be encouraged. Obviously, for this, inferences drawn from evaluation results of individual courses for various academic programs may be genuine evidence. An academic institution may adopt a universally used evaluation results without fail. Such option may not be useful only for institutions that develop sustainable system to achieve accreditation but also for the accredited institutions to retain it as well as further develop for higher accreditation.

2. Methodology

2.1. Data

The used data were collected on three academic programs Bachelor of Commerce (BCOM), Bachelor of Mass Media (BMM) and Master of Commerce (MCOM). The first two academic programs were at undergraduate level and the

third were at postgraduate level. The evaluation data could be collected for three courses under BCOM, two courses under BMM and two courses under MCOM. As obvious, the number of students under various courses in an academic program may not be the same. Also, it gets further changed due to attendance on the time of data collection. Accordingly number of students completing evaluations under BCOM ranged from 21 to 42; those under each course of BMM remained 18 whereas those under MCOM remained 28.

While planning for Course Evaluation Survey (CES), the focus was to cover at least two courses from each academic program. Accordingly number of students (n) and their relative % out of total covered in three courses under COM were (n = 92, 50%); two under BMM (n = 36, 20%) and two under MCOM (n = 57, 30%). Likewise, under the BCOM, three courses were covered: Maths/Statistics (n = 21, 23%); BC (n = 42, 46%) and MHRM (n = 29, 31%). Under the BMM, two courses were covered: History (n = 18, 50%); and Sociology (n = 18, 50%). Under the MCOM, two courses were covered: Economics (n = 28, 49%); Financial Accounts (n = 29, 51%). The number of students covered under each Course Evaluation Survey obviously relied on attendance of the students on the day of survey. The data collection for this study was carried out during February 2015.

2.2. Analytical Methods

As described earlier, one parametric and three non-parametric evaluation measures were used in the present analysis [3]:

2.2.1. Parametric Methods

Average evaluation score, that is, average agreement by the students for each item was calculated using arithmetic mean. Further, the perceived performance of each of the 15 items by the students was expressed in three ranges:

- 3.6 & above - High quality
- 2.6 - 3.6 - Acceptable
- Less than 2.6 - Improvement required

As mentioned earlier, use of arithmetic mean to analyze individual Likert type items is known to be inappropriate. It may be highly influenced by extreme values of evaluation scores. Further, it is likely to provide underestimation, or overestimation of the mean due to known observed skewness in such data. As a result, performance grading of individual items based on average score alone might be either underestimated or overestimated. In addition, average of strong disagreements reported by few students to strong agreements by others for an item may reveal a misleading impression of average agreement. In other words, average grading based on average score may neither have clarity nor ease of understanding by the educational policy planners.

2.2.2. Non-Parametric Methods

To overcome the problem of inappropriateness of using mean average grading, non-parametric three evaluation measures that have their added merits were

used:

1) Evaluation Measure Using Median

Median evaluation score for a Likert type item insinuates that at least 50% of the students marked that score or higher scores for that item. In case of the involved fraction in the calculated median score, although negligible, it was rounded off for better understanding to policy planners. For an educational institution which is newly started, to begin with an aim to attain satisfaction among at least 50% students, a threshold that at least 50 % students report agreement score 4 or 5 for each item needs to be targeted. Accordingly, keeping in view of the targeted threshold of the agreement, the grading of each item using the students' median agreement score was considered as follows:

- 4 & 5 - High quality
- 3 - Acceptable
- 1 & 2 - Improvement required

Intuitively, the grading of individual Likert type items using median score facilitates easy understanding by the educational policy planners.

2) Evaluation Measure Using First Quartile

On the line of median evaluation score, first quartile evaluation score insinuates that at least 75% of the students marked that score or higher scores for that item. In case of the involved fraction in the calculated first quartile score, it was also rounded off. To move towards higher quality, once a minimum target of satisfaction among at least 50% students is achieved, an educational institution may target an increased level of satisfaction among at least 75% students. Accordingly, with an increased target that at least 75% students report agreement score 4 or 5 for each item, grading of each item using first quartile score was considered as:

- 4 & 5 - High quality
- 3 - Acceptable
- 1 & 2 - Improvement required

As obvious, the educational policy planners understand it easily and use it for future improvements in their educational institutions. Intuitively, in comparison to median evaluation score, it can help in identifying additional areas covered in the CES questionnaire for further improvements.

3) Evaluation Measure using Cumulative Percentage

As already adopted by some globally reputed educational institutions, a cumulative percentage of students with satisfaction score 4 or 5 was worked out for each item. To improve further, an increased level of students' satisfaction may be considered. Moving above first quartile evaluation score as 4 or 5, it is considered as satisfaction among at least 80% students. With higher target that at least 80% of students' report agreement score 4 or 5 for each item, using the listed cumulative percentage against 4 or 5, grading of each item was considered as:

- 80 & above - High quality
- 60 - 80 - Acceptable
- Less than 60 - Improvement required

The consideration of cumulative percentage as evaluation measure may have additional advantages. In comparison with both evaluation measures, median and first quartile, it is a straight forward measure, and facilitates clear understanding to educational policy planners. Also, compared to first quartile, it can further yield additional areas for the attention of institutional administrators. Further, in case of repeated course evaluation surveys, it may help in quantifying increment/decrement in improvements of each item over a period of time.

2.2.3. Pooled Analysis

For poorly performing courses on various items, each of 15 items may be considered to be equally important. Accordingly pooled analysis was carried out at course level.

2.2.4. Comparative Results

Considering each of the four measures, pooled analysis was used to discuss comparative results for each course. The usefulness of this approach is emphasized through respective depiction in figures.

2.3. Results

The analytical results for the course BC (**Table 1** and **Figure 1**) reveal that based on average score students showed high agreement with 60% of the items, whereas remaining 40% showed acceptable performance. Based on median score, 93% items showed high performance. On the other hand, based on first quartile,

Table 1. Evaluation results of course BC (n = 42).

Items_Domains	Mean	Median	First Quartile	CF (≥ 4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.9	4.0	3.0	64.3	0.0	4.8	31.0	35.7	28.6
item2_D1	3.5	4.0	3.0	59.5	4.8	4.8	31.0	52.4	7.1
item3_D1	3.8	4.0	3.0	66.7	0.0	16.7	16.7	40.5	26.2
item4_D2	4.0	4.0	3.0	71.4	0.0	9.5	19.0	35.7	35.7
item5_D2	4.2	5.0	4.0	78.7	0.0	7.1	14.3	26.3	52.4
item6_D2	3.6	4.0	3.0	61.9	4.8	14.3	19.0	40.5	21.4
item7_D2	3.9	4.0	3.0	69.1	4.8	7.1	19.0	28.6	40.5
item8_D2	3.7	4.0	3.0	52.4	2.4	16.7	28.6	16.7	35.7
item9_D2	3.6	4.0	3.0	54.8	4.8	11.9	28.6	26.2	28.6
item10_D2	3.5	4.0	3.0	54.7	7.1	11.9	26.2	33.3	21.4
item11_D3	3.4	4.0	3.0	54.8	11.9	4.8	28.6	42.9	11.9
item12_D3	3.4	4.0	3.0	54.8	4.8	16.7	23.8	42.9	11.9
item13_D3	3.1	3.0	2.0	45.2	11.9	19.0	23.8	35.7	9.5
item14_D4	3.8	4.0	3.0	62.0	7.1	2.4	28.6	31.0	31.0
item15_D5	3.9	4.0	3.0	73.8	2.4	7.1	16.7	45.2	28.6

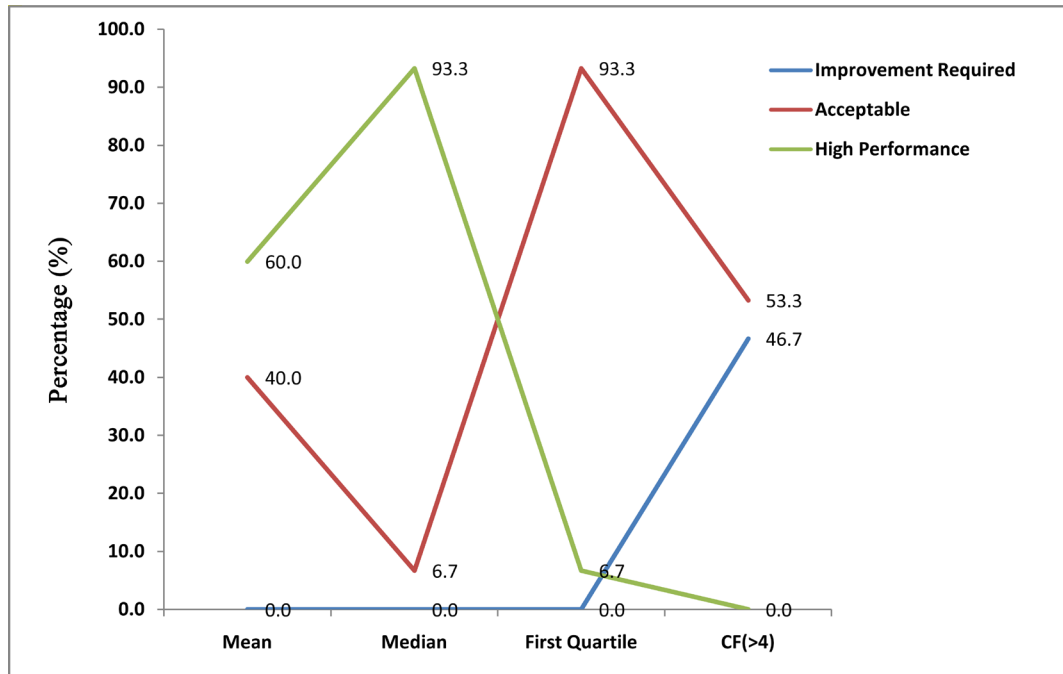


Figure 1. Results of course BC.

93% items showed only acceptable performance. Further, based on cumulative percentage, only 53% showed acceptable performance whereas 47% require further improvements. In other words, targeting satisfaction among at least 50% of the students (considering median results) showed high performance of 93% of the items. However, considering threshold of satisfaction among at least 75% students (first quartile result), there was higher performance of only 7% of the items. After enhancing the threshold of satisfaction further at least at 80%, none of the items showed high performance.

The analytical results of the course Math/Stat (Table 2 and Figure 2) suggest that based on the mean performance, high performance was reported for 33% items whereas 67% had acceptable performance. However, based on median performance, high performance was noticed regarding 40% of the items, acceptable performance was true for 27% of the items whereas remaining 33% of the items needed further improvement. In other words, even 50% of the students did not feel satisfied with 33% of the items. As obvious, this percentage doubled to 60% as evident from results under first quartile. In other words, 60% items did not reach to target of achieving satisfaction among at least 75% students. Further, 80% of the items require further improvement if we aim to gain satisfaction among at least 80% of the students.

As shown in Table 3 and Figure 3 for the course MHRM, based on average score, 93% of the items had acceptable performance. However, aiming to gain satisfaction among at least 50% of the students (based on median), 33% high performance and remaining 67% of the items showed acceptable performance. Further, increasing this threshold to at least among 75% students (first quartile), revealed that 67% items showed acceptable performance but remaining 33%

needed further improvement. However, increasing the threshold to satisfaction among at least 80% of the students, only 13% showed acceptable performance and a large number of items (87%) require further improvements.

For the course sociology, as evident from **Table 4** and **Figure 4**, based on

Table 2. Evaluation results of course Maths/Stats (n = 21).

Items_Domains	Mean	Median	First Quartile	CF (≥ 4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.6	4.0	3.0	57.1	0.0	19.0	23.8	33.3	23.8
item2_D1	2.8	2.0	2.0	38.1	23.8	33.3	4.8	14.3	23.8
item3_D1	2.9	2.0	1.0	47.6	33.3	19.0	0.0	23.8	23.8
item4_D2	2.9	3.0	1.0	38.1	33.3	9.5	19.0	9.5	28.6
item5_D2	3.6	4.0	2.0	61.9	0.0	33.3	4.8	28.6	33.3
item6_D2	3.1	3.0	2.0	38.1	19.0	14.3	28.3	14.3	23.8
item7_D2	2.7	2.0	1.0	42.9	33.3	19.0	4.8	28.6	14.3
item8_D2	2.8	2.0	1.0	38.1	33.3	19.0	9.5	9.5	28.6
item9_D2	3.4	4.0	3.0	57.4	14.3	9.5	23.8	28.6	28.8
item10_D2	2.9	3.0	2.0	42.9	23.8	19.0	14.3	28.6	14.3
item11_D3	3.0	3.0	2.0	38.1	23.8	9.5	28.6	23.8	14.3
item12_D3	3.8	4.0	3.0	66.6	14.3	9.5	9.5	19.0	47.6
item13_D3	3.1	4.0	2.0	52.4	19.0	23.8	4.8	28.6	23.8
item14_D4	4.1	4.0	4.0	76.2	0.0	14.3	9.5	28.6	47.6
item15_D5	3.7	4.0	3.0	57.1	4.8	0.0	38.1	33.3	23.8

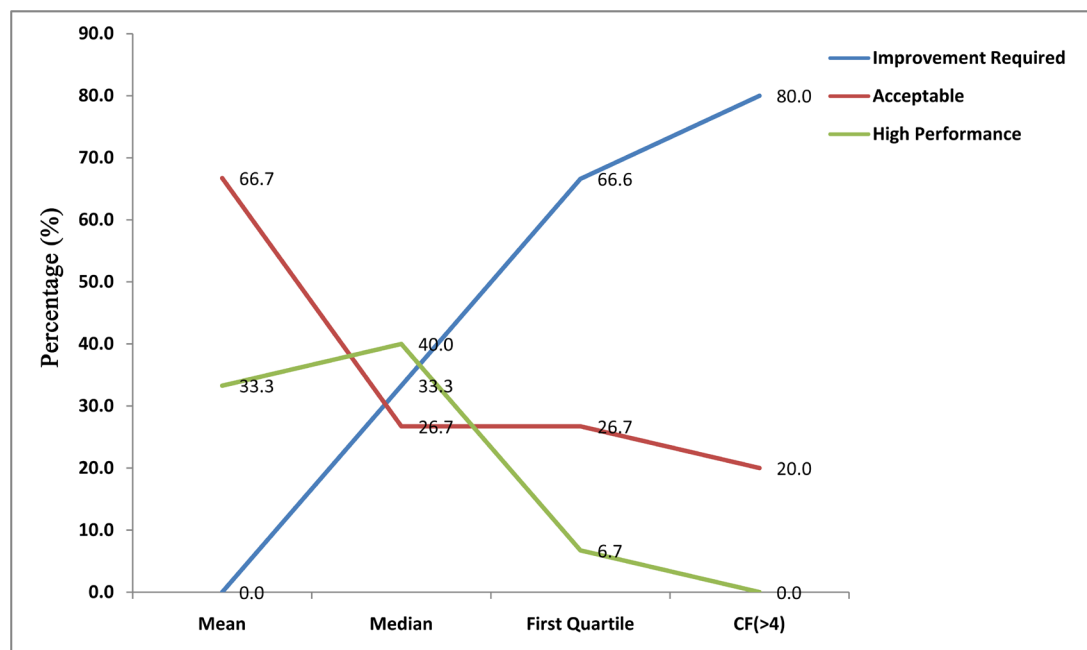


Figure 2. Results of course Maths/Stats.

Table 3. Evaluation results of course MHRM (n = 29).

Items_Domains	Mean	Median	First Quartile	CF (≥4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.5	4.0	3.0	51.7	10.3	6.9	31.0	27.6	24.1
item2_D1	3.4	4.0	3.0	51.7	10.3	10.3	27.6	34.5	17.2
item3_D1	2.7	3.0	2.0	27.6	13.8	34.5	24.1	27.6	0.0
item4_D2	3.6	4.0	3.0	62.1	6.9	13.8	17.2	41.4	20.7
item5_D2	3.4	4.0	3.0	55.1	6.9	17.2	20.7	37.9	17.2
item6_D2	2.9	3.0	2.0	58.4	24.1	30.8	20.7	27.6	30.8
item7_D2	3.3	4.0	2.0	51.7	3.4	27.6	17.2	34.5	17.2
item8_D2	3.0	3.0	2.0	34.4	3.4	41.4	20.7	24.1	10.3
item9_D2	3.2	3.0	2.0	48.2	13.8	17.2	20.7	31.0	17.2
item10_D2	3.1	3.0	2.0	41.4	6.9	34.5	17.2	20.7	20.7
item11_D3	2.9	3.0	2.0	41.3	20.7	17.2	20.7	31.0	10.3
item12_D3	2.7	3.0	2.0	31.0	20.7	27.6	20.7	20.7	10.3
item13_D3	2.4	2.0	1.0	27.5	34.5	24.1	13.8	17.2	10.3
item14_D4	3.1	3.0	2.0	44.8	17.2	17.2	20.7	27.6	17.2
item15_D5	3.3	3.0	3.0	62.8	6.9	17.2	31.0	31.0	31.8

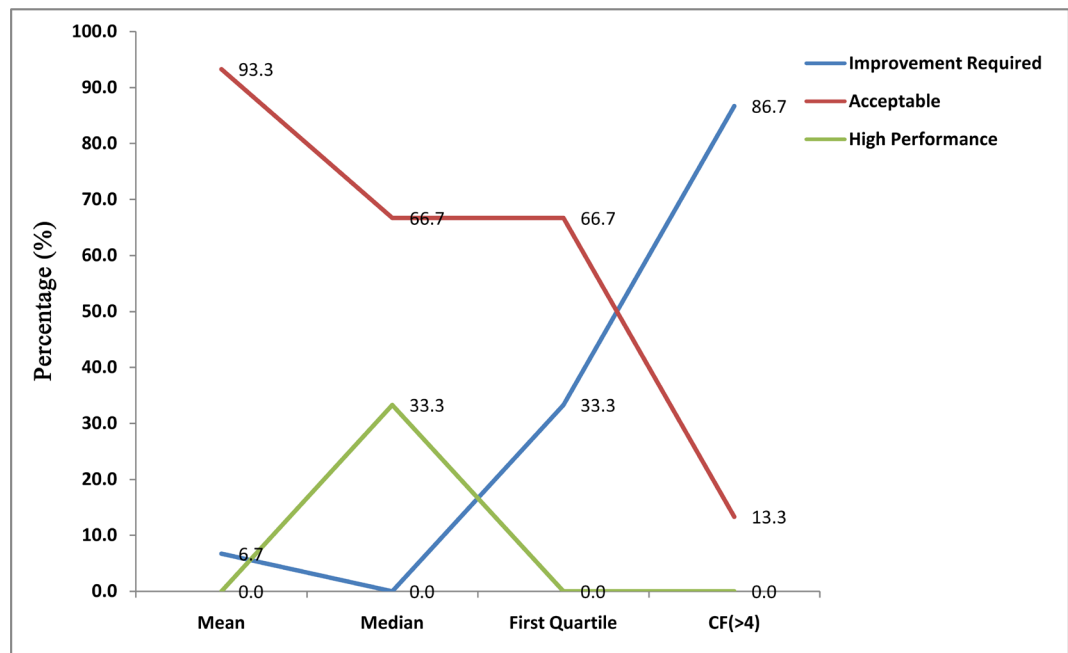
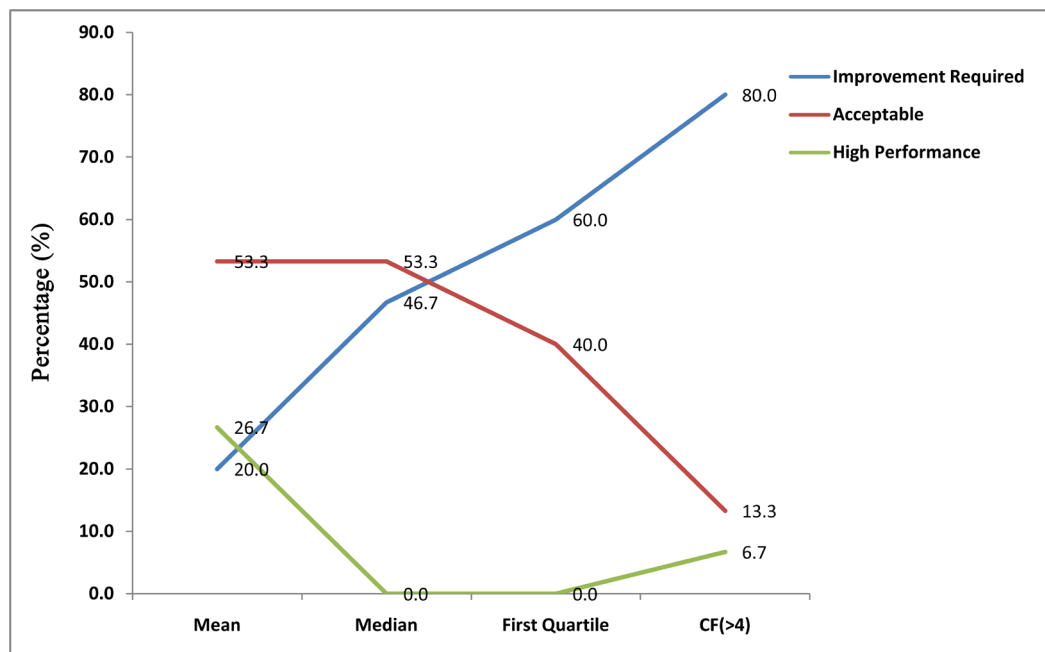


Figure 3. Results of course MHRM.

average score 53% of the items showed acceptable performance followed by 27% high performance and remaining 20% showing improvement required. Based on median score, 33% items showed high performance, 53% of the items showed acceptable performance whereas remaining 13% showed improvement required. Further, based on the threshold of gaining satisfaction among at least 75% students,

Table 4. Evaluation results of course Sociology (n = 18).

Items_Domains	Mean	Median	First Quartile	CF (≥ 4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.7	4.0	3.0	66.6	5.6	5.6	22.2	44.4	22.2
item2_D1	3.4	3.0	3.0	44.4	5.6	5.6	44.4	33.3	11.1
item3_D1	3.1	3.0	3.0	22.2	0.0	22.2	55.6	11.1	11.1
item4_D2	3.8	4.0	3.0	55.6	5.6	5.6	33.0	16.7	38.9
item5_D2	3.4	3.5	3.0	50.0	11.1	11.1	27.8	27.8	22.2
item6_D2	4.1	4.0	4.0	83.3	5.6	0.0	11.1	44.4	38.9
item7_D2	3.1	3.0	2.0	38.9	16.7	16.7	27.8	22.2	16.7
item8_D2	2.4	2.5	2.0	11.2	22.2	27.8	38.9	5.6	5.6
item9_D2	3.2	3.0	2.0	38.9	5.6	27.8	27.8	22.2	16.7
item10_D2	3.1	3.0	2.0	38.9	16.7	16.7	27.8	16.7	22.2
item11_D3	3.5	3.5	2.0	50.0	16.7	11.1	22.2	5.6	44.4
item12_D3	2.2	2.0	1.0	16.7	38.9	27.8	16.7	11.1	5.6
item13_D3	2.1	2.0	1.0	11.2	38.9	33.3	16.7	5.6	5.6
item14_D4	3.3	4.0	2.0	55.5	5.6	22.2	16.7	44.4	11.1
item15_D5	3.8	4.0	3.0	72.3	0.0	5.6	22.2	55.6	16.7

**Figure 4.** Results of course Sociology.

53% items require further improvement and 40% showed acceptable performance, while only 7% showed high performance. An additional threshold of 5% revealed that 80% items require further improvements and 13% had acceptable performance. Under this threshold also, % of items with high performance remained to be 7%.

As evident from analytical results (Table 5 and Figure 5) for the course history, considering average criterion, 20% had high performance whereas remaining 80% of the items had acceptable performance. The consideration of median threshold (satisfaction among at least 50% of students), 27% of the items had higher performance whereas 73% had acceptable performance. An additional

Table 5. Evaluation results of course History (n = 18).

Items_Domains	Mean	Median	First Quartile	CF (≥ 4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.3	3.0	3.0	38.9	5.6	5.6	50.0	27.8	11.1
item2_D1	3.3	3.5	3.0	50.0	5.6	11.1	33.3	50.0	0.0
item3_D1	3.0	3.0	3.0	16.7	5.6	5.6	72.2	16.7	0.0
item4_D2	3.8	4.0	4.0	77.6	16.7	0.0	5.6	44.3	33.3
item5_D2	3.1	3.0	2.0	38.9	22.2	11.1	27.8	16.7	22.2
item6_D2	3.8	4.0	3.0	66.7	0.0	11.1	22.2	38.9	27.8
item7_D2	4.0	4.0	4.0	83.4	0.0	0.0	16.7	66.7	16.7
item8_D2	2.9	3.0	2.0	33.3	11.1	27.8	27.8	22.2	11.1
item9_D2	3.3	3.0	2.0	44.5	11.1	16.7	27.8	16.7	27.8
item10_D2	3.2	4.0	1.0	55.6	27.8	5.6	11.1	27.8	27.8
item11_D3	3.0	3.0	2.0	38.9	16.7	22.2	22.2	22.2	16.7
item12_D3	2.7	2.5	1.0	44.5	33.3	16.7	5.6	38.9	5.6
item13_D3	2.4	2.0	1.0	16.7	33.3	22.2	27.8	0.0	16.7
item14_D4	3.2	3.0	2.0	44.5	11.1	16.7	27.8	27.8	16.7
item15_D5	3.3	3.0	3.0	44.5	5.6	16.7	33.3	27.8	16.7

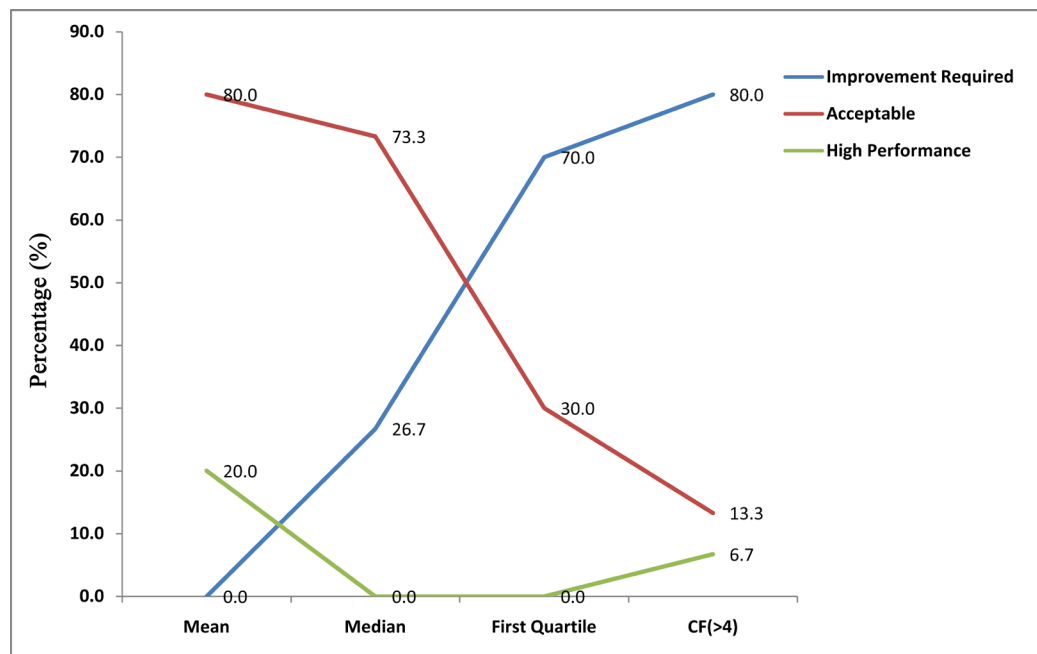


Figure 5. Results of course History.

consideration of 25% threshold (that is, satisfaction among at least 75% students), revealed high performance only for 13% of items, average performance of 33% items and remaining 54% of the items required further improvement. Likewise, the increase in threshold of satisfaction among at least 80% students, only 7% items showed high performance, 13% with acceptance performance and remaining 80% requiring improvement.

For the course Financial Accounts (**Table 6** and **Figure 6**) under postgraduate academic program MCOM, criterion used regarding average performance revealed high performance of 73% of items, and remaining 27% had acceptable performance. Interestingly, based on median score, high performance went upper side as 80% of the items. Further, none of the items showed the need for further improvement. Further, 25% increase in threshold about satisfaction among students (based on first quartile), only 13% of items remained to have high performance and remaining all other items had acceptable performance. Surprisingly, another increase in threshold by 5% resulted in high performance of only 7% items and acceptable performance of 20% items. Remaining 67% of the items required further improvement.

For the course Economics (**Table 7** and **Figure 7**), average score considerations revealed 80% items showing high performance and remaining with acceptable performance. A target of satisfaction amongst at least 50% of students interestingly showed high performance for all the items (100%). However, aiming higher target of satisfaction among at least 75% students resulted in drastic reduction in proportion of items showing high performance (6.7%). Further, out

Table 6. Evaluation results of course Financial Accounts (n = 29).

Items_Domains	Mean	Median	First Quartile	CF (≥ 4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	3.9	4.0	3.0	69.0	6.9	6.9	17.2	27.6	41.4
item2_D1	3.6	4.0	3.0	51.7	3.4	6.9	37.9	34.5	17.2
item3_D1	3.6	4.0	3.0	58.6	3.4	13.8	24.1	41.4	17.2
item4_D2	4.0	4.0	4.0	82.8	10.3	6.9	0.0	41.4	41.4
item5_D2	3.9	4.0	3.0	68.9	3.4	6.9	20.7	37.9	31.0
item6_D2	3.3	4.0	3.0	55.2	13.8	10.3	20.7	41.4	13.8
item7_D2	3.9	4.0	4.0	75.9	6.9	10.3	6.9	34.5	41.4
item8_D2	3.5	4.0	3.0	51.7	13.8	6.9	27.6	20.7	31.0
item9_D2	3.3	4.0	3.0	51.7	6.9	13.8	27.6	41.4	10.3
item10_D2	3.3	3.0	3.0	48.3	13.8	10.3	27.6	27.6	20.7
item11_D3	3.4	4.0	2.0	62.0	13.8	20.7	3.4	31.0	31.0
item12_D3	3.3	4.0	2.0	55.1	20.7	17.2	6.9	24.1	31.0
item13_D3	2.8	3.0	2.0	31.0	17.2	27.6	24.1	20.7	10.3
item14_D4	3.5	4.0	3.0	58.6	6.9	17.2	17.2	34.5	24.1
item15_D5	3.6	3.0	3.0	44.8	0.0	6.9	48.3	24.1	20.7

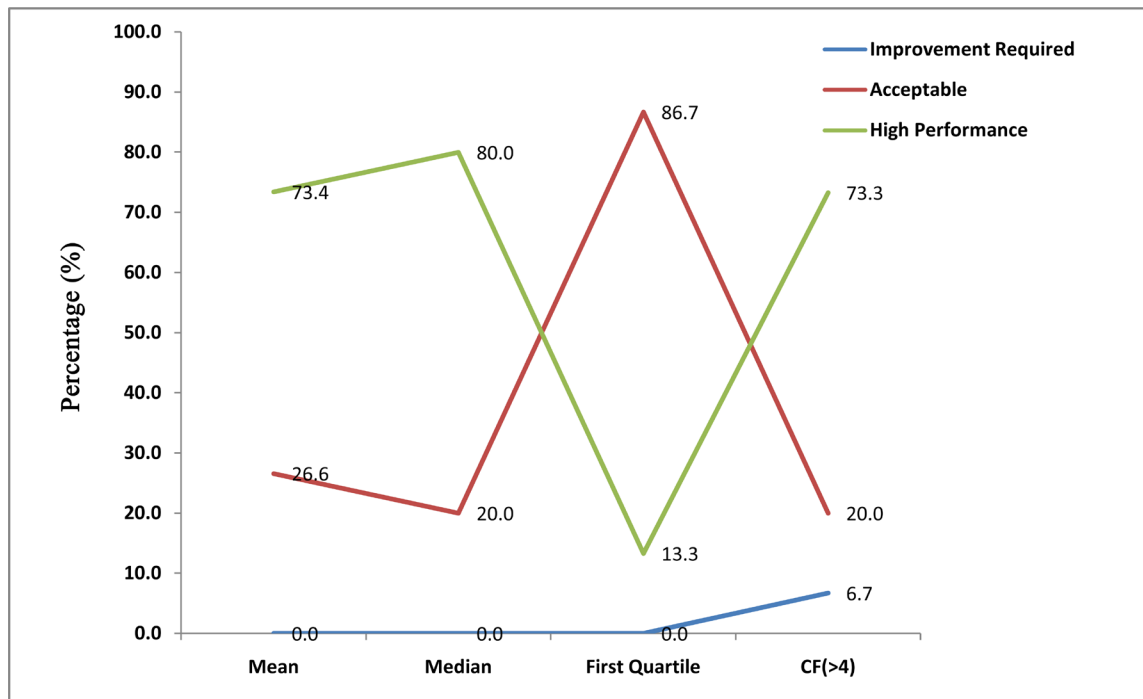


Figure 6. Results of course Financial Accounts.

Table 7. Evaluation results of course Economics (n = 28).

Items_Domains	Mean	Median	First Quartile	CF (≥4)	Strongly Disagree	Disagree	Neither nor	Agree	Strongly Agree
item1_D1	4.0	4.0	3.0	67.8	3.6	3.6	25.0	21.4	46.4
item2_D1	3.6	4.0	3.0	40.7	3.6	7.1	28.6	26.4	14.3
item3_D1	3.8	4.0	3.5	75.0	0.0	10.7	14.3	57.1	17.9
item4_D2	3.9	4.0	3.5	95.0	14.3	3.6	7.1	48.6	46.4
item5_D2	3.8	4.0	3.0	60.7	7.1	3.6	28.6	21.4	39.3
item6_D2	3.7	4.0	3.0	64.3	7.1	7.1	21.4	39.3	25.0
item7_D2	3.9	4.0	4.0	78.6	10.7	7.1	3.6	35.7	42.9
item8_D2	3.8	4.0	3.5	75.0	17.9	3.6	3.6	28.6	46.4
item9_D2	3.8	4.0	3.0	71.4	10.7	7.1	10.7	39.3	32.1
item10_D2	3.5	4.0	3.0	57.1	7.1	14.3	21.4	35.7	21.4
item11_D3	3.8	4.0	3.0	71.4	10.7	7.1	10.7	35.7	35.7
item12_D3	3.5	4.0	2.0	60.7	10.7	25.0	3.6	28.6	32.1
item13_D3	3.4	4.0	2.5	53.6	14.3	10.7	21.4	28.6	25.0
item14_D4	3.8	4.0	3.0	67.8	10.7	7.1	14.3	32.1	35.7
item15_D5	3.7	4.0	3.0	57.1	3.6	3.6	35.7	32.1	25.0

of remaining items, only one item (6.7%) required further improvements. Another increase in threshold by 5% resulted in required further improvements for 27% of the items.

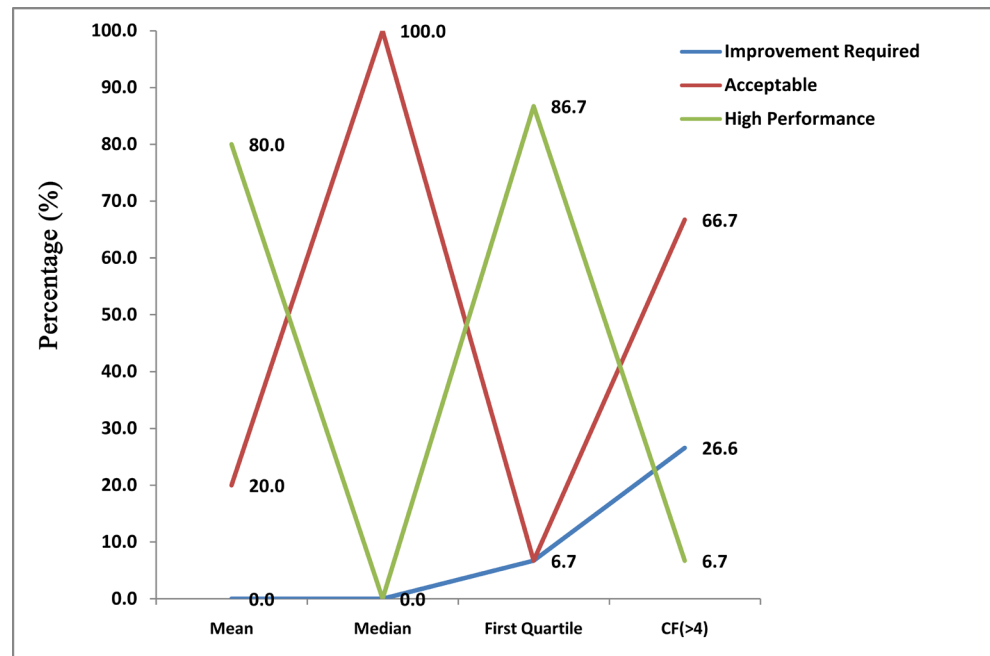


Figure 7. Results of course Economics.

Comparative Results

Based on average measure (Table 8), out of the seven subjects only economics emerged to be performing well, followed by BC. Another three subjects Math/Stat, Sociology and Financial Accounts showed moderate level performance. Surprisingly, for the subject MHRM each item (other than 13th) could show acceptable performance. Further, as evident from Table 8, the overall satisfaction (global item 15) did not reveal consistency in terms of the grading for the remaining items. As evident from the table, average measure did not provide any concrete clues useful for educational policy planners.

Like in case of average score, based on median summary measures presented in Table 9, three subjects BC, Economics and Financial Accounts showed high performance whereas remaining subjects in relation to all items had mixed performance in terms of high and acceptable performances. In summary, at least 50% of the students were found to be satisfied with every item under two subjects, namely BC and Economics. As a contradictory finding under subject Financial Accounts, in spite of satisfaction by at least 50% students regarding most of the items, overall satisfaction did not emerge to be high performing subject. Such findings might advocate the need of further exploration exclusively on global item (15th item).

Keeping in view of the fact that first quartile as 4.0 or 4.5 (Table 10) reveals that at least 75% of the students expressed satisfaction to the related items, none of the subjects emerged to be high performer. Overall each of the seven subjects showed acceptable performance. Three subjects, BC, Economics and Financial Accounts showed acceptable performance in relation to the majority of the items. Four subjects namely Math/Stat, MHRM, History and Sociology require

Table 8. Average summary measures for all courses.

Items	Mean						
	BC	Maths/Stat	MHRM	History	Sociology	Economics	Financial Accounts
1	3.9	3.6	3.5	3.3	3.7	4.0	3.9
2	3.5	2.8	3.4	3.3	3.4	3.6	3.6
3	3.8	2.9	2.7	3.0	3.1	3.8	3.6
4	4.0	2.9	3.6	3.8	3.8	3.9	4.0
5	4.2	3.6	3.4	3.1	3.4	3.8	3.9
6	3.6	3.1	2.9	3.8	4.1	3.7	3.3
7	3.9	2.7	3.3	4.0	3.1	3.9	3.9
8	3.7	2.8	3.0	2.9	2.4	3.8	3.5
9	3.6	3.4	3.2	3.3	3.2	3.8	3.3
10	3.5	2.9	3.1	3.2	3.1	3.5	3.3
11	3.4	3.0	2.9	3.0	3.5	3.8	3.4
12	3.4	3.8	2.7	2.7	2.2	3.5	3.3
13	3.1	3.1	2.4	2.4	2.1	3.4	2.8
14	3.8	4.1	3.1	3.2	3.3	3.8	3.5
15	3.9	3.7	3.3	3.3	3.8	3.7	3.6

Table 9. Median summary measures for all courses.

Items	Median						
	BC	Maths/Stat	MHRM	History	Sociology	Economics	Financial Accounts
1	4.0	4.0	4.0	3.0	4.0	4.0	4.0
2	4.0	2.0	4.0	3.5	3.0	4.0	4.0
3	4.0	2.0	3.0	3.0	3.0	4.0	4.0
4	4.0	3.0	4.0	4.0	4.0	4.0	4.0
5	5.0	4.0	4.0	3.0	3.5	4.0	4.0
6	4.0	3.0	3.0	4.0	4.0	4.0	4.0
7	4.0	2.0	4.0	4.0	3.0	4.0	4.0
8	4.0	2.0	3.0	3.0	2.5	4.0	4.0
9	4.0	4.0	3.0	3.0	3.0	4.0	4.0
10	4.0	3.0	3.0	4.0	3.0	4.0	3.0
11	4.0	3.0	3.0	3.0	3.5	4.0	4.0
12	4.0	4.0	3.0	2.5	2.0	4.0	4.0
13	3.0	4.0	2.0	2.0	2.0	4.0	3.0
14	4.0	4.0	3.0	3.0	4.0	4.0	4.0
15	4.0	4.0	3.0	3.0	4.0	4.0	3.0

Table 10. Summary measures based on first quartile for all courses.

Items	Q1						
	BC	Maths/Stat	MHRM	History	Sociology	Economics	Financial Accounts
1	3.0	3.0	3.0	3.0	3.0	3.0	3.0
2	3.0	2.0	3.0	3.0	3.0	3.0	3.0
3	3.0	1.0	2.0	3.0	3.0	3.5	3.0
4	3.0	1.0	3.0	4.0	3.0	3.5	4.0
5	4.0	2.0	3.0	2.0	3.0	3.0	3.0
6	3.0	2.0	2.0	3.0	4.0	3.0	3.0
7	3.0	1.0	2.0	4.0	2.0	4.0	4.0
8	3.0	1.0	2.0	2.0	2.0	3.5	3.0
9	3.0	3.0	2.0	2.0	2.0	3.0	3.0
10	3.0	2.0	2.0	1.0	2.0	3.0	3.0
11	3.0	2.0	2.0	2.0	2.0	3.0	2.0
12	3.0	3.0	2.0	1.0	1.0	2.0	2.0
13	2.0	2.0	1.0	1.0	1.0	2.5	2.0
14	3.0	4.0	2.0	2.0	2.0	3.0	3.0
15	3.0	3.0	3.0	3.0	3.0	3.0	3.0

further improvements in relation to the majority of the items. As true in case of median measures, inconsistency of overall grading with those on other items also remains true.

As per global practice, any academic administrator would like to achieve satisfaction among at least 80% students not only in relation to global item but also in relation to each of the remaining items. However, as evident from **Table 11**, four of the seven subjects Math/Stat, History, Economics and Financial Accounts need further improvements. Remaining three courses BC, Sociology and MHRM need improvement from acceptable performance to high performance. It may be worthwhile to mention here that inconsistency regarding overall grading of the subject remains true in this case also. Keeping in view of the points mentioned above, almost every subject needs attention regarding every item listed in the questionnaire so that the desired target by the academic administrator might be achieved. Further, the issue of inconsistency in reporting an overall satisfaction in comparison to those on remaining items may be investigated to find out the preferred items by the students.

To begin with, the academic administrator of any institution might focus on overall satisfaction by the students while evaluating the quality of a specific course. This becomes more useful specifically to the institutions established little earlier.

Under the present study, out of seven courses four namely BC, Math/Stat,

Table 11. Summary measures based on cumulative percentage for all courses.

Items	CF (>4)						
	BC	Maths/Stat	MHRM	History	Sociology	Economics	Financial Accounts
1	64.3	57.1	51.7	38.9	66.6	67.8	69.0
2	59.5	38.1	51.7	50.0	44.4	40.7	51.7
3	66.7	47.6	27.6	16.7	22.2	75.0	58.6
4	71.4	38.1	62.1	77.6	55.6	95.0	82.8
5	78.7	61.9	55.1	38.9	50.0	60.7	68.9
6	61.9	38.1	58.4	66.7	83.3	64.3	55.2
7	69.1	42.9	51.7	83.4	38.9	78.6	75.9
8	52.4	38.1	34.4	33.3	11.2	75.0	51.7
9	54.8	57.4	48.2	44.5	38.9	71.4	51.7
10	54.7	42.9	41.4	55.6	38.9	57.1	48.3
11	54.8	38.1	41.3	38.9	50.0	71.4	62.0
12	54.8	66.6	31.0	44.5	16.7	60.7	55.1
13	45.2	52.4	27.5	16.7	11.2	53.6	31.0
14	62.0	76.2	44.8	44.5	55.5	67.8	58.6
15	73.8	57.1	62.8	44.5	72.3	57.1	44.8

Sociology and Economics had high performance even in terms of the satisfaction among at least 50% students. However, remaining three subjects MHRM, History and Financial Accounts remained to be acceptable performers which obviously require support by the item specific analysis presented earlier.

Surprisingly, as evident from **Table 10**, every subject remained under acceptable performance category if we aim to achieve satisfaction among at least 75% of the students. As obvious, after increasing the threshold to achieve satisfaction among at least 80% of the students (**Table 11**), only three subjects namely BC, MHPM and Sociology remained under acceptable performing category. Remaining four subjects Math/Stat, History, Economics and Financial Accounts were pushed down to the category meant for further improvement required.

2.4. Conclusions

The average agreement score did not provide any legible and accurate clues regarding grading of courses towards high satisfaction, average satisfaction and further improvement required. If we take the minimum threshold of satisfaction among at least 50% of the students, the high agreement was lowest (0%) for the course MHRM and highest performance (80%) for the course Economics. The next highest performance was observed under the course Financial Accounts. Interestingly results on global item showed consistently high performance for the courses BC, Math/Stat and Economics. On contrary, larger proportion of in-

dividual items under Financial Accounts showed high performance, however, its overall satisfaction grading was merely at average level. For the courses MHRM and History overall grading was consistent at average level. On contrary, in the case of Sociology also global item emerged to be with high performance whereas less than one-third of individual item only were graded with high performance [13] [14] [15] [16].

After considering the increased threshold as satisfaction among at least 75% of the students, surprisingly none of the subjects could emerge with high performance globally. Each of the subject remained at average level suggesting further improvements. In addition, Math/Stat, MHR, History and Sociology have high proportion of individual items merely at further improvements required. This simply indicates that every course needs to be monitored almost on each of the item regarding further improvement.

A further movement of threshold to 80% changed four subjects namely Math/Stat, History, Economics and Financial Accounts in the level of further improvements required. Also, remaining three courses BC, Sociology and MHRM remained at acceptable performance level. The overall grading remained consistent in relation to subjects Math/Stat, History and Financial Accounts. On contrary, there was inconsistency regarding global grading compared to individual items in relation to subjects BC, MHRM, Sociology and Economics. In summary, to achieve this target the academic administrator of the courses need to work hard sincerely regarding improvement in every aspect from first item to fourteenth item.

Regardless of the threshold regarding satisfaction among students (from 50% to 80%), overall performance of almost every subject has been unsatisfactory. There is need to make focused approach to take every course at the level of high performance [17] [18].

2.5. Recommendation

The inconsistency noticed under every threshold revealed that under such poorly performing subjects globally, one needs to analyze merely at global level. Once the global level analysis reveals high performance of a course, then only item specific analysis may be required to find out the items requiring further improvements.

Acknowledgements

The authors are thankful to the concerned authorities to grant permission to collect the data. Also, thanks to every student who participated in the study. The editorial board members as well as reviewers are also duly acknowledged who could help in strengthening the article further.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Al Rubaish, A., Wosornu, L. and Dwivedi, S.N. (2011) Using Deduction Form Assessment Studies towards Furtherance of the Academic Program: An Empirical Appraisal of Institutional Student Course Evaluation. *iBusiness*, **3**, 220-228. <https://doi.org/10.4236/ib.2011.32029>
- [2] Al Rubaish, A., Wosornu, L. and Dwivedi, S.N. (2012) Item Reduction in “Course Evaluation Survey” Questionnaire: Some Exploratory Analysis and Empirical Evidence. *IJBNSI*, **2**, 1-11.
- [3] Franklin, J. (2001) Interpreting the Numbers: Using a Narrative to Help Others Read Student Evaluations of Your Teaching Accurately. In: Lewis, K.G., Ed., *Techniques and Strategies for Interpreting Student Evaluations. New Directions for Teaching and Learning Series, No. 87*, Jossey-Bass, San Francisco, 85-100. <https://doi.org/10.1002/tl.10001>
- [4] Theall, M. and Franklin, J. (2001) Looking for Bias in All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction? In: Theall, M., Abrami, P.C. and Mets, L.A., Eds., *The Student Ratings Debate. Are They Valid? How Can We Best Use Them? New Directions for Institutional Research, Series, No. 109*, Jossey-Bass, San Francisco, 45-46. <https://doi.org/10.1002/ir.3>
- [5] Zableta, F. (2007) The Use and Misuse of Student Evaluations of Teaching. *Teaching in Higher Education*, **12**, 55-76. <https://doi.org/10.1080/13562510601102131>
- [6] Johnson, R.L. and Morgan, G.B. (2016) Survey Scales: A Guide to Development Analysis and Reporting. The Guilford Press, New York, 70-89.
- [7] Carifio, J. and Perla, R.J. (2007) Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and Their Antidotes. *Journal of Social Sciences*, **3**, 106-116. <https://doi.org/10.3844/jssp.2007.106.116>
- [8] Sundaram, K.R., Dwivedi, S.N. and Sreenivas, V. (2015) Medical Statistics: Principles & Methods. Second Edition, Wolters and Kluwer (Health) Pvt. Ltd., New Delhi.
- [9] Gob, R., Mc Collin, C. and Rmalhoto, M.F. (2007) Ordinal Methodology in the Analysis of Likert Scales. *Quality & Quantity*, **41**, 601-626. <https://doi.org/10.1007/s11135-007-9089-z>
- [10] Cashin, W.E. (1990) Students Do Rate Different Academic Fields Differently. In: Theall, M. and Franklin, J., Eds., *Student Ratings of Instruction: Issues for Improving Practice [Special Issues]*, *New Directions for Teaching and Learning, Series, No. 43*, Jossey-Bass, San Francisco, 113-121. <https://doi.org/10.1002/tl.37219904310>
- [11] Gravestock, P. and Gregor-Greenleaf, E. (2008) Student Course Evaluations: Research, Models and Trends. Higher Education Quality Council of Ontario, Toronto.
- [12] Cohen, R.J. and Swerdlik, M.E. (2005) Psychological Testing and Assessment. 6th Edition, McGraw Hill, New York.
- [13] Field, A.P. (2005) Discovering Statistics Using SPSS. 2nd Edition, Sage Publications, London.
- [14] Dufy, K.A. and O’Neill, P.A. (2003) Involving Students in Staff Development Activities. *Medical Teacher*, **25**, 191-194. <https://doi.org/10.1080/0142159031000092616>
- [15] Fuller, M., Georgeson, J., Healey, M., Hurst, A., Ridell, S., Roberts, H. and Weedon, E. (2009) Enhancing the Quality and Outcomes of Disabled Students’ Learning in Higher Education. Routledge, London.

- [16] Healey, M., O'Connor, K.M. and Broadfoot, P. (2010) Reflection on Engaging Student in the Process and Product of Strategy Development for Learning, Teaching, and Assessment: An Institutional Case Study. *International Journal for Academic Development*, **15**, 19-32. <https://doi.org/10.1080/13601440903529877>
- [17] Marsh, H.W. and Roche, L.A. (1997) Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility. *American Psychologist*, **52**, 1187-1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- [18] Wright, R.E. (2006) Student Evaluations of Faculty: Concerns Raised in the Literature, and Possible Solutions. *College Student Journal*, **40**, 417-422.