

## Using text data mining techniques for understanding free-style question answers in course evaluation forms

Amr Abd-Elrahman  
University of Florida

Michael Andreu  
University of Florida

Tiffany Abbott  
University of Florida

### ABSTRACT

Like many universities, University of Florida courses are evaluated by students using a standard form and set of questions including quantitative questions (rate on a scale of 1 – 5) as well as subjective short answer questions. Student answers to the short answer questions of the University of Florida standard course evaluation sheets were analyzed using text data mining techniques to identify unrevealed aspects affecting the teaching process and develop a quantification tool for these aspects. We analyzed student answers on 25 standard University of Florida (UF) course evaluations sheets representing 4 courses (5 sections – 2 instructors). The answers from these course evaluations were scored positively or negatively in two independent ways, manually using human interpretation and automatically based on keyword co-occurrence text mining algorithm. The number of positive and negative answers related to different teaching aspect categories was determined. We introduced the Teaching Evaluation Index (TEI), as an index to quantify students textual evaluations using the number of positive and negative comments interpreted from the text.

The TEI values computed using manually interpreted and computationally mined student short answers showed strong correlation ( $R^2=0.96$ ). A comparison of the TEI and overall course and instructor evaluation means extracted from the quantitative student responses were analyzed. This analysis showed strong correlation between the TEI values and the overall course and instructor evaluation means ( $R^2=0.86$  and  $R^2=0.92$ , respectively). The results of our experiment showed that text mining of student short answers, with its automation capacity, can provide efficient additional (or alternative) measure for the overall course evaluation process. More data is recommended to generalize these results.

Keywords: course evaluation, Text mining, short answer questions, teaching evaluation index, co-occurrence analysis

## INTRODUCTION

Student evaluation is an integral part of the education process, but it is often viewed with differing perspectives and purposes. Some experts view evaluation as a “test of effectiveness – of materials, teaching methods. Even further evaluation gives insight on how to improve current practices (Ramsden, 2003). In essence evaluation has been viewed to have two classic purposes: audit and development also referred to as accountability and improvement (Bowden & Marton, 1998), appraisal and developmental purpose (Kember et al., 2002), judgmental and developmental purpose (Hounsell, 2003), or quality assurance and quality enhancement (Biggs, 2003). Student evaluation of faculty at the collegiate level is seen as a means of accountability and aids in the efforts to define and measure teaching effectiveness (Chen & Hoshower, 2003). Evaluation is not a perfected practice but overall student ratings have been relatively well accepted by researchers and practitioners in the field because “student ratings are the single most valid source of data on teaching effectiveness - in fact there is little support for the validity of any other source of data” (Spencer & Schmelkin, 2002; McKeachie, 1997).

Even better students consistently do not show an opposition to answering the evaluations and typically answer the questions honestly and willingly (Douglas & Carroll, 1987; Hofman & Kremer, 1980; Marsh, 1984, 1987; Tom et al., 1990). In the end, students often view the evaluation ratings as a way to improve the faculties teaching methods. Also students perceived the current system of evaluating faculty to be effective and believed that faculty valued input from the evaluations. Student evaluation can be divided into Summative and Formative (Scriven, 1967). Abbott et al., (1990) found that students often preferred the use of mid-semester formative evaluations because they could see the feedback in practice rather than at the end of the semester. Generally, both types of evaluation contain numerical (quantitative) and textual (qualitative) questions. Quantitative questions are often considered by administration for overall evaluation of the faculty, while answers to qualitative essay-style questions are left to the faculty to examine and utilize. Although human comprehension of the text information on the evaluation sheet is important and optimal, quantitative analysis of the students’ narrative response can reveal hidden (or stress existing) aspects of the teaching process. Additionally, it can provide quality control measures for the evaluation sheet and another metric for the administration to assess faculty performance.

Text mining or text data mining is the process of deriving interesting information from text through discovering patterns and trends. Text mining algorithms are utilized in several applications such as summarizing and analyzing web content (Himmel, et al., 2009; Jackson & Moulinier, 2007) improving customer relations (Coussement & Vandenpoel, 2008) and managing scientific publications (Cohen & Hunter, 2008). Text mining generally starts with a text refining step, where free-style text is transformed into a structured form (e.g. relational database) (Delgado et al., 2002). Such data can be used for analysis that involves document clustering and categorization (Tan, 1999). The data can be used to deduce patterns and relationships among extracted data elements. The later analysis is domain dependent and requires conceptual knowledge about the theme of the extracted data.

In this study we compare manual interpretation of short answer student response to course evaluation sheet questions with automated analysis using text mining algorithms. This prototype study demonstrates the potential for using student responses to extract information in a quantifiable manner through text mining techniques. We consider our results a proof of concept

that demonstrate the need for future analysis that utilizes larger data set and more sophisticated text mining analysis that is dedicated towards the teaching process domain.

## METHODOLOGY

In this study, we analyzed student answers on 25 standard University of Florida (UF) course evaluations sheets representing 4 courses (5 sections – 2 instructors). A typical UF course evaluation sheet contains 13 questions that are quantitative using a score or 1 – 5 (poor – excellent) at the front page of the sheet. On the other side of the evaluation sheet, 5 essay-style questions that allow the student to respond in a more qualitative open ended manner exist. Figure 1 shows the back of a standard UF course evaluation sheet.

The answers for the 5 free-style questions were analyzed using human interpretation and using text mining algorithms. As a pre-processing step, the answers were transcribed by an impartial person and checked for spelling errors. The data was organized in a database table that includes information about course number, instructor, section, number, semester, level, evaluation question, and transcribed student responses to the questions. Table 1 demonstrates few records of the used data table.

Manual interpretation of student answers was performed by identifying the five major elements (categories) of the teaching process: course; instructor; assessment; material; and delivery. Each of these categories was further broken into several subcategories to increase the analysis resolution. Each evaluation sheet was manually interpreted to identify the number of positive and negative comments for pre-identified categories (and their subcategories). For example, the number of positive and negative answers related to the course delivery method (e.g. live, video conferencing, asynchronous web-based, or synchronous via web) was determined. The numbers for these subcategories were summed together to form the number of positive and negative responses for the ‘Delivery’ main category.

The student response was analyzed using the Wordstat software to suggest a keywords list in addition to a list of excluded words (e.g. ‘the’, ‘about’, ‘can’). More words were manually added to the excluded words list due to their insignificant linguistic value in the data mining application. Two major groups of words indicating positive (e.g. ‘good’, ‘amazing’, and ‘challenging’) and negative (e.g. ‘poor’, ‘hard’, ‘confusing’) words were created. The remaining keywords were manually examined and inclusively divided into eight different categories pertaining to the quality of the teaching process. Most of these categories matched those identified in the manual analysis section.

Co-occurrence-based analysis was automatically performed on the data. The number of co-occurrences between positive and negative word groups and each of the teaching quality word categories in addition to other variables such as instructor number and section number within the same text unit (sentence) was automatically determined and summarized against data variables such as the instructor and course variables. Figure 2 shows a screen snapshot of the Wordstat software co-occurrence analysis. The figure shows how many co-occurrences of positive keywords (listed by sentence) with different variables (shown as columns). The figure also shows a bar chart diagram that illustrates how many co-occurrences of positive keywords and the instructor-number variable.

A newly introduced Teaching Evaluation Index (TEI) is computed based on the total positive (Pos\_cnt) and negative (Neg\_cnt) counts for each course section. The index can be computed for each variable (instructor, section, etc.)/teaching category (assessment, material,

instructor, etc.) influencing the teaching process based on the positive and negative occurrence count. The index also has a [-1,1] bound, where the bounds indicate totally negative and positive comments, respectively.

$$TEI = \frac{Pos\_cnt - Neg\_cnt}{Pos\_cnt + Neg\_cnt} \quad (1)$$

## RESULTS

Manual mining of student answers to identify strong and weak points in the teaching process was the original motivation for this research. The results of the human interpretation of student answers of individual course sections that counted the number of positive and negative response for each of five main categories affecting the teaching process is shown in table 2. The table also shows the overall Teaching Education Index for each of the analyzed course sections. The number of positive co-occurrence of negative/positive keywords with each of the 8 main categories affecting the teaching process (analyzed through the wordstat text mining software) and the computed TEI are listed in table 3 and summarized by section numbers.

## ANALYSIS AND DISCUSSION

Table 2 shows that mining the free-style text at the back of the evaluation sheets revealed some aspects affecting the teaching process that were hidden in the student answers. The table shows high number of negative responses for distance courses (section 7258:1/7 and section 6241:0/5). It also shows potential difference for student evaluation standards from graduate and undergraduate students. Investigating the number of positive and negative keyword co-occurrences with the teaching process categories summarized by variables such as instructor, course or section number could reveal important information or pattern. For example, figure 3 shows the number of positive and negative keyword co-occurrences for section 7258. The figure reveals some delivery and schedule related problems associated with this section.

The results shown in figure 4 indicates a strong correlation ( $R^2=0.96$ ) between TEI values computed using manual and automated (text mining) analysis considering all teaching evaluation categories combined. However, the results of individual categories did not show such correlation as shown in tables 2 and 3. This may be attributed to the accurate sentence-level comprehension of positive and negative results belonging to each category in the manual text analysis case. In contrast, in the automated text mining case, general keywords were interpreted and classified into different categories regardless of sentence semantics. Figure 5 shows that the TEI values and overall course and instructor evaluation means computed from the front page of the evaluation sheet questions are strongly correlated ( $R^2=0.86$  and  $R^2=0.92$ , respectively). This indicates the potential of using the TEI index as an extra measure for overall course performance.

## CONCLUSION

We utilized a small data set of student course evaluation answers to provide preliminary analysis on the feasibility of text mining techniques in analyzing the students' narrative answers. Although, only small dataset was used in this study, our results proved that text mining is a promising technique to analyze short answer textual information in the students' course evaluation sheets more efficiently than by simply having to read each comment individually. By analyzing these responses and calculating the Teaching Evaluation Index (TEI) can transform

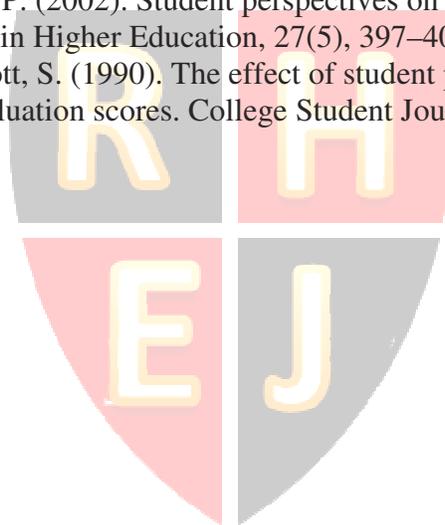
qualitative responses into quantitative information so that one can gain additional insights to evaluate the value of the course from the students perspective.

Examining the TEI computed from manual interpretation of student results showed significant correlation with student answers to the overall course and instructor evaluation questions located at the front page of the sheet. This result suggests potential use of analyzed student narrative answers as alternative (or quality control measure) to student answers of the quantitative questions at the front of the evaluation sheet. The strong correlation between TEI values computed through human text interpretation and text mining algorithm suggest the potential for automating the process, which may be necessary for large scale implementation. However, significant linguistic and psychological research is needed to fine tune keyword selection and to better understand word semantics in a teaching evaluation domain.

## REFERENCES

- Abbott, R.D., Wulff, D.H., Nyquist, J.D., Ropp, V.A. & Hess, C.W. (1990). Satisfaction with processes of collecting student opinions about instruction: The student perspective. *Journal of Educational Psychology*, 82, 201–206.
- Ah-Hwee Tan, 1999. Text mining: The state of the art and the challenges. In *Proceedings PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases (KDAD'99)*, pages 71–76, 1999.
- Biggs, J. (2003). *Teaching for quality learning at university: What the student does*. Buckingham, UK: SRHE and Open University Press.
- Bowden, J., & Marton, F. (1998). *The university of learning: Beyond quality and competence in higher education*. London: Kogan Page.
- Chen, Y., & Hoshower, L.B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71–88.
- Cohen KB, Hunter L (2008) Getting started in text mining. *PLoS Comput Biol* 4(1): e20. doi:10.1371/journal.pcbi.0040020
- Coussement K., Vandenpoel ,D. (2008) Improving Customer Complaint Management by Automatic Email Classification Using Linguistic Style Features as Predictors, *Decision Support Systems*, 44(4), 870-882
- Delgado, M., M.J. Martín-Bautista, D. Sánchez, M.A. Vila, “Mining Text Data: Special Features and Patterns”. In *Proc. of EPS Exploratory Workshop on Pattern Detection and Discovery in Data Mining*, London, September 2002. *Lecture Notes in Computer Science* 2447, D. Hand et.al., Eds., pp. 140–153. Springer-Verlag.
- Douglas, P.D., & Carroll, S.R. (1987). Faculty evaluations: Are college students influenced by differential purposes? *College Student Journal*, 21(4), 360–365.
- Himmel W., Reincke, U., Michelmann H. W. (2009) Text mining and Natural language Processing Approaches for automatic categorization of lay requests to web-based expert forums. *Journal of Medical Internet Research*, 11(3):e25.
- Hofman, J.E., & Kremer, L. (1980). Attitudes toward higher education and course evaluation. *Journal of Educational Psychology*, 72, 610–617.
- Hounsell, D. (2003). The evaluation of teaching. In H. Fry, S. Ketteridge & S. Marshall (Eds.), *A handbook for teaching and learning in higher education: Enhancing academic practice* (pp. 200–212). London: Kogan Page.

- Jackson P, Moulinier I (2007) Natural language processing for online applications: Text retrieval, extraction, and classification. 2nd edition. Herndon (Virginia): John Benjamins Publishing Company.
- Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education*, 27(5), 411–425.
- Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76(5), 707–754.
- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research*, 11(2), 253–388.
- McKeachie, W.J. (1997). Student ratings the validity of use. *American Psychologist*, 52(11), 1218–1225.
- Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London: Routledge Falmer, p. 223.
- Scriven M. (1967) The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.
- Spencer, K.J., & Schmelkin, L.P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, 27(5), 397–409.
- Tom, G., Swanson, S., & Abbott, S. (1990). The effect of student perception of instructor evaluations on faculty evaluation scores. *College Student Journal*, 24(3), 268–273.



APPENDIX

**SIDE 2**

Instructors can improve their classes through thoughtful student reactions. Please comment on any of the following about which you have an opinion—and be frank. Instructors do not have access to course evaluations until after final grades have been submitted.

**PLEASE COMMENT ABOUT THE FOLLOWING:**

1. What personal qualities or teaching practices of the instructor contributed to the success of the course?  
 Dr. Amr ~~was~~ lent tremendous depth to an introductory class. I especially enjoyed learning about map projections & satellites. He is also funny.
2. What personal qualities or teaching practices of the instructor hindered the success of the course?  
 In regards to teaching practices, I feel that ~~the~~ an instructor in the flesh is always better than listening to someone through a screen. This is no reflection on the professor. It's doable & not all that bad. I understand having to do this to save money by having one instructor for 2 campuses. That way  
 3. What is your opinion of the course, including the text(s) and materials?  
 the president, provost & other administrators can maintain their ridiculously high salaries.
4. Comment on the adequacy of the materials/equipment provided to conduct class/laboratory activities.  
 Good. I would like to have learned how to use a GPS.
5. Add any other comments you wish.  
 To improve this course, I highly recommend the following: using too time for the tutorial is not very helpful. Assistance isn't really needed for the tutorial. what we began to do towards the end of the semester was to meet in a group with Emily Foster so we did our assignments collaboratively & with guidance. This is when my learning curve took an upward swing. I found having someone there in person, rather than e-illuminate

Figure 1. The back page of the back of a standard UF College of Agricultural and Life Sciences course evaluation sheet

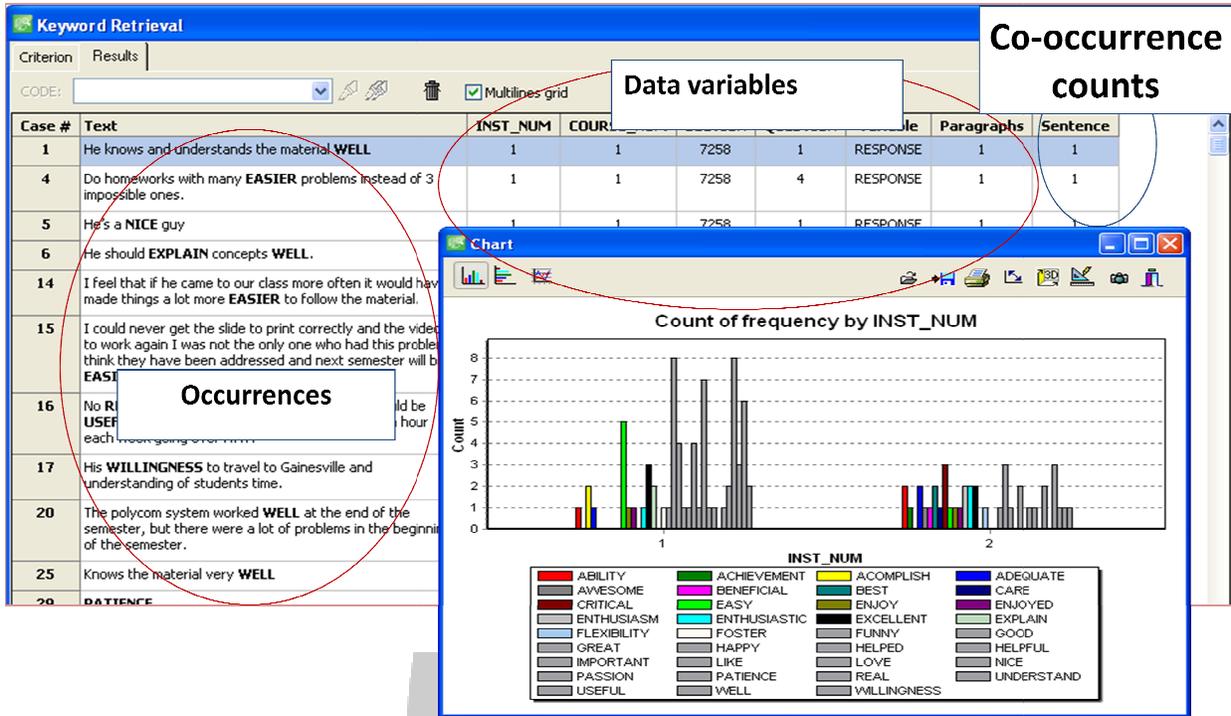


Figure 2. Screen snapshot of the Wordstat software co-occurrence analysis

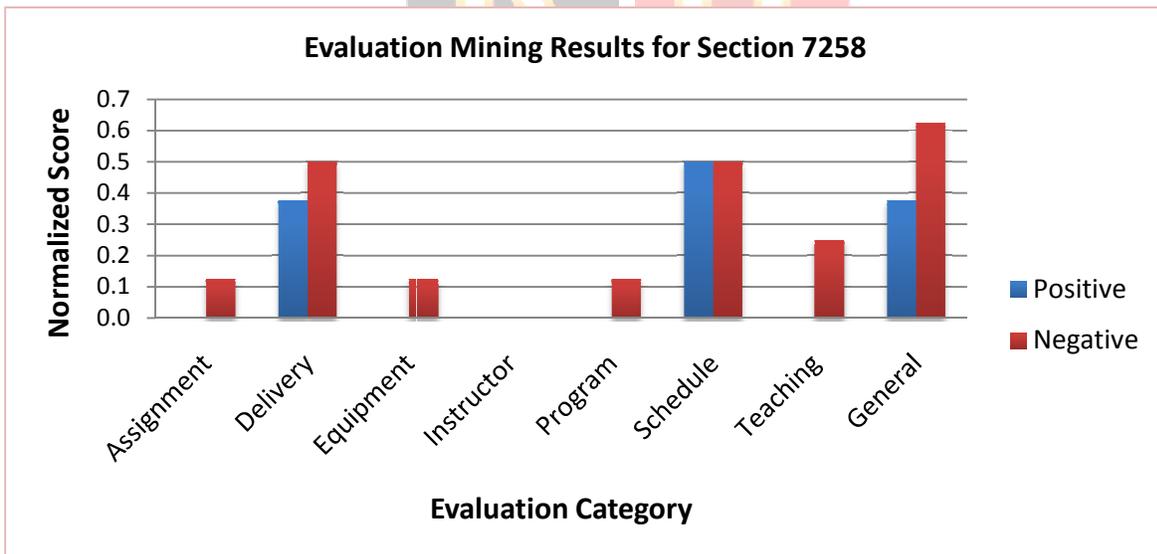


Figure 3. Number of positive and negative keyword co-occurrences with different teaching categories for section 7258

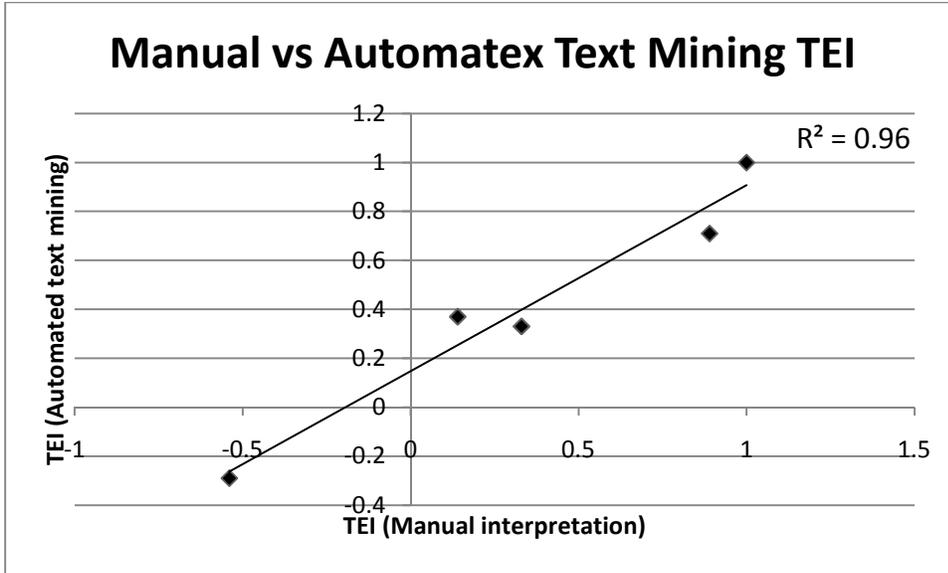


Figure 4. Plot of TEI values computed using manual interpretation and text mining techniques

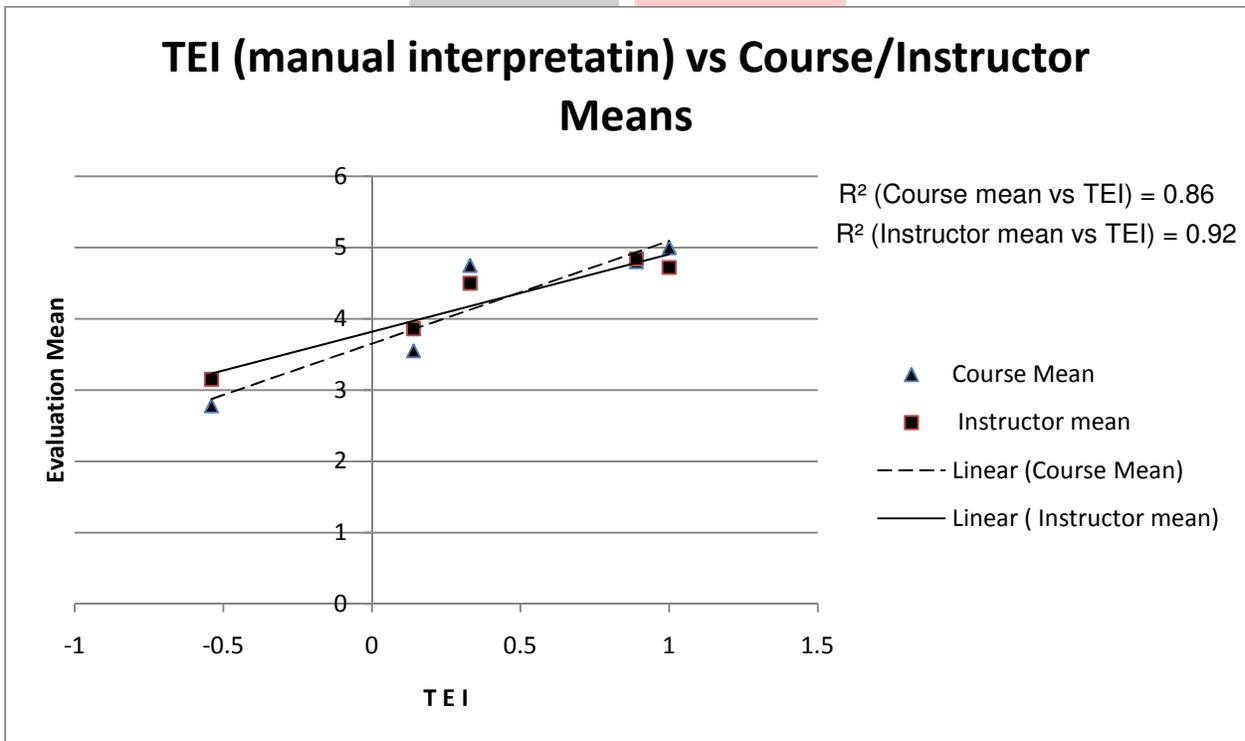


Figure 5. Plot of TEI values computed through manual interpretation and the answer to the overall course and instructor evaluation at the front of the evaluation sheet.

Table 1. Sample of database table containing course evaluation information including student textual response

ID	Instructor	course	Course_Num	Semester_Yr	level	section	Student	Question	Response*
1	1	SUR3641	1	fall2008	UG	7258	1	1	e.g. He knows and understands the material well
2	1	SUR3641	1	fall2008	UG	7258	1	2	e.g. Polycom is not the same as face-to-face
3	1	SUR3641	1	fall2008	UG	7258	1	3	e.g. 3 hr course is too long, and will be IMPOSSIBLE to maintain discipline/attention spans
4	1	SUR3641	1	fall2008	UG	7258	1	4	Response Masked
5	1	SUR3641	1	fall2008	UG	7258	2	1	Response Masked
6	1	SUR3641	1	fall2008	UG	7258	2	2	Response Masked
7	1	SUR3641	1	fall2008	UG	7258	2	3	Response Masked
8	1	SUR3641	1	fall2008	UG	7258	2	4	Response Masked

UG: undergraduate \* response masked for privacy reasons

Table 2. Manual Interpretation results of student response for different teaching categories

section (partici/enrol)	7258 (9/9)	8546 (2/2)	8624 (2/4)	6241 (11/18)	7371&736 (5/8)					
Delivery/level	Dist/UG	Dist/G	live/UG	Dist/UG	live/UG					
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
General/Course	0	1	1	0	0	0	1	0	9	0
Instructor	5	5	2	0	3	0	5	3	14	0
Assessment	0	4	0	0	0	0	0	1	0	0
Material	0	3	5	0	1	0	10	3	14	2
Delivery	1	7	0	0	0	2	0	5	0	0
sum	6	20	8	0	4	2	16	12	37	2
TEI	-0.54		1.00		0.33		0.14		0.89	

UG: ndergraduate Dist: Distance Education (videoconference) and virtual classroom

Table 3. Count of positive/negative co-occurrences for each of different teaching categories summarized by the course section variable.

section	7258		8546		8624		6241		7371&7362	
(partici/enrol)	(9/9)		(2/2)		(2/4)		(11/18)		(5/8)	
loc/level	Dist/UG		Dist/G		live/UG		Dist/UG		live/UG	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
General/Course	3	5	5	0	0	0	4	4	11	3
Instructor	0	0	4	0	4	1	1	1	0	0
Assessment	0	1	0	0	0	0	0	4	0	0
Material	2	4	5	0	1	4	11	1	18	3
Delivery	3	4	0	0	0	0	5	1	4	0
Equipment	0	1	0	0	4	0	2	0	0	0
Program	0	1	2	0	4	1	2	0	1	0
Schedule	4	4	0	0	1	0	1	1	3	1
Teaching	0	2	0	0	0	1	2	1	4	0
sum	12	22	16	0	14	7	28	13	41	7
TEI	-0.29		1.00		0.33		0.37		0.71	

UG: undergraduate Dist: Distance Education (videoconference) and virtual classroom

