Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 20, Number 20, October 2015 ISSN 1531-7714

RMP Evaluations, Course Easiness, and Grades: Are they Related?

Syed A. Rizvi, *College of Staten Island/CUNY*

This paper investigates the relationship between the student evaluations of the instructors at the RateMyProfessors.com (RMP) website and the average grades awarded by those instructors. As of Spring 2012, the RMP site included evaluations of 538 full-and part-time instructors at the College of Staten Island (CSI). We selected the evaluations of the 419 instructors who taught at CSI for at least two semesters from Fall 2009 to Spring 2011 and had at least ten evaluations. This research indicates that there is a strong correlation between RMP's overall evaluation and easiness scores. However, the perceived easiness of an instructor/course does not always result in higher grades for students. Furthermore, we found that the instructors who received high overall evaluation and easiness scores (4.0 to 5.0) at the RMP site do not necessarily award high grades. This is a very important finding as it disputes the argument that instructors receive high evaluations because they are easy or award high grades. On the other hand, instructors of the courses that are perceived to be difficult (RMP easiness score of 3.0 or less) are likely to be tough graders. However, instructors who received moderate overall evaluation and easiness scores (between 3.0 and 4.0) the RMP site had a high correlation between these scores and average grade awarded by those instructors. Finally, our research shows that the instructors in non-STEM disciplines award higher grades than the instructors in STEM disciplines. Non-STEM instructors also received higher overall evaluations than their STEM counterparts and non-STEM courses were perceived easier by the students than STEM courses.

Student evaluations of instructors have always been a controversial issue in academia, where student evaluations are routinely used in reappointments, tenure, and promotion decisions of faculty. Some consider them a valid measure of teaching effectiveness (Abrami & Mizener, 1985). Others, however, call them a *popularity contest* in which high grades mean high scores on student evaluations (Goldman, 1990). Regardless of the faculty perspective of student evaluations, students have no use of their evaluations because student evaluations are kept confidential at most institutions. Therefore, word of mouth had been the main source for students to decide whether an instructor is a good teacher or not, which may influence their choice of classes. Since the inception of RateMyProfessors.com (RMP), the website seems to provide for students an alternative means to make their judgment about an instructor. The RMP site allows students to rate their instructors in four areas: (1) easiness, (2) clarity, (3) helpfulness, and (4) overall evaluation. They used a scale of one to five with one being the least favorable and five being the most favorable scores in a category. Students' ratings in each category for all classes evaluated by students on the RMP site of an instructor are averaged and assigned to that instructor as the score in that particular category. The validity of RMP evaluations, however, has been in question largely because of the unscientific nature of the RMP survey. On the other hand, the evidence of the use of RMP evaluations by the students for choosing classes, largely

anecdotal though, has convinced many researchers to investigate the validity of the RMP evaluations.

The previous research on RMP has provided useful insight into a relationship between overall evaluations of the instructors at RMP website and the perceived easiness of the course by the students. Felton et al. reported a high correlation (r = 0.61) between the course easiness and the overall score (Felton, Mitchell, & Stinson, 2004). Coladarci et al. reported a little less strong correlation (r = 0.41) between the course easiness and the overall score (Coladarci & Kornfield, 2007). Davison et al. reported a strong correlation (r = 0.51) between the course easiness and the overall score, which affirms a significant relationship between these quantities (Davison & Price, 2009). Bleske-Rechek et al. reported on a discipline-specific analysis that showed that the students treat these two measures differently regardless of the correlation between them (Bleske-Rechek & Michels, 2010). In their more recent research, Bleske-Rechek et al. investigate the reliability of the RMP ratings based on students' consensus in their evaluation (Bleske-Rechek & Fritsch, 2011). They argue for the validity of the RMP ratings concluding that there is an exchange of useful information among the users of the RMP website about the quality of the instructions. Other research has been focused on the relationship between the RMP ratings and institutional student evaluations. Sanders et al. concluded that online ratings provide a reliable source of information for predicting institutional student ratings despite the dissimilarities between the two ratings (Sanders, Walia, Potter, & Linna, 2011). Coladarci et al. also investigated the relationship between the RMP ratings and the student evaluations of instructors at their institutions. They reported a strong correlation between many items in both ratings; the strongest correlation (r = 0.68) reported was between the overall evaluation scores of the RMP and overall evaluation scores of their institutional ratings (Coladarci & Kornfield, 2007).

Much of the research done in the RMP evaluations has been focused on investigating the relationship between overall evaluation and easiness. However, little research has been done on the relationship between the overall evaluations of instructors and the average grades awarded by them. Furthermore, given the amount of data publicly available from the RMP website, more research is warranted to investigate the benefits of this site if any, to students. For example, do high evaluations mean easy grades? Or does taking a course with an instructor with low evaluations and low scores on easiness mean potentially risking the grade in that course? These and many similar questions cannot be answered by merely finding the correlation between these quantities. For example, many researchers have reported a strong relationship between overall evaluation scores and easiness in terms of correlation, which range from 0.4 to over 0.6. However, a measure represented by correlation tells us just that. It does not tell us what kind of relationship this is. If this strong relationship makes RMP ratings less reliable, is it true for the whole range of instructors (from very hard to very easy)? Do the students behave in a consistent manner while evaluating the whole range of instructors? Are the "easy" instructors rewarded with the same rigor as that with which the "hard" instructors are punished? Is there any value that RMP ratings bring to the institutions, instructors, or students if the easiness and overall evaluation scores are highly correlated? In this research, we attempt to answer some of these questions by investigating the relationship between these quantities through correlation as well as trend and multiple regression analyses. The rest of the paper is organized as follows:

Section II briefly describes the trend analysis. Section III explains the data set and methodology used in this research. Section IV presents the data analysis and observations. Section V concludes the paper indicating possible future research directions.

Trend Analysis

Time series analysis is used for forecasting, prediction, and identifying underlying trends in data in many areas such as the stock market, weather forecasting, prediction of seismic activity, etc. In a time series, the data are collected at different time intervals, which may or may not be equally spaced, over a period of time. However, the time series analysis can also be applied to data sequences, in which time is not a variable, to gain useful insight from the data, such as underlying trends, etc. This is done by mapping a non-time sequence (or even a non-numeric sequence) into a time series. For example, Kwan and Arniker propose several methods to map a DNA sequence to a numerical sequence for subsequent time series analysis (Kwan & Arniker, 2009). Trend analysis, though used extensively for time series, can also be used for non-time sequences to study the behavior of one quantity (for example overall evaluation) with respect to the

change in another quantity (for example, the average grade given by the instructors). Therefore, trend analysis can be more revealing than other measures that describe the relationship between two quantities using a single value such as correlation. In this research, we investigate the relationship between overall evaluation scores at RMP and the easiness of the courses as reported by the students on RMP, in terms of underlying trends. We also investigate the trends between overall evaluation scores at RMP, and the grades awarded by the instructors.

Trend analysis can potentially reveal more information about how two quantities are related than other measures such as correlation. For example, it can reveal whether the instructors who gave an average grade of C received poorer evaluations were hit harder in evaluations than the instructors who gave an average grade of B, or the students' evaluation of the instructors was consistent regardless of the average grade given by the instructors. Another aspect of this research was to investigate the relationship between overall evaluation scores at RMP, the easiness of the courses and grades received as a function of the number of evaluations received.

Dataset and Methodology

We obtained, from the RMP website, the overall evaluation and easiness scores of 538 full- and part-time instructors at the College of Staten Island. Of 538 instructors, we selected the instructors with at least ten evaluations at the RMP website. We also removed any duplicate records. We then included the average grades given by the instructors over a period of two academic years. We then removed instructors who did not teach in at least two semesters during the period of two academic years. This left us with a total of 419 instructors. The final data set included the instructors who taught courses in a wide range of disciplines at CSI in sciences, humanities, and social sciences as well as in some professional programs. Finally, we removed all identification information of the instructors from our data before further analysis of the data. To investigate the relationship between overall evaluation scores at RMP, the easiness of the courses, and grades received as a function of the number of evaluations received, we created two additional sets of data: (1) instructors with at least 20 evaluations; this data set included 194 instructors. (2) Instructors with at least 40 evaluations; this data set included 43 instructors. The mathematical foundation of our methodology is as follows:

The data are ordered with respect to overall evaluation scores of RMP to generate two sequences of length N each, an easiness sequence (E) and a grade sequence (G), where the sequences E and G are expressed as:

$$E = \{e_i\}_{i=1}^{N} \quad (1)$$

and

$$G = \{g_i\}_{i=1.}^{N} \quad (2)$$

Ordering the data with respect to the evaluation scores of RMP allows us to observe the behaviors (trends) of the easiness scores of RMP and the average grade given by the instructors as the overall evaluation score is changed from 1 to 5. Each data point on these curves corresponds to an instructor. There are many methods to generate a trend sequence from an arbitrary sequence. We chose the discrete convolution (moving average) method because of its simplicity and robustness. The trend sequences for easiness, E_t, is expressed as follows:

$$E_t = \{e_{ti}\}_{i=M}^{N-M} \quad (3)$$

where M is the length of the moving window, and the ith member of the trend sequence is given by

$$e_{ti} = \frac{1}{M}\sum_{j=i-M}^{i+M} e_{j.} \quad (4)$$

Similarly, the trend sequence for grades, G_t, is expressed as follows:

$$G_t = \{g_{ti}\}_{i=M}^{N-M} \quad (5)$$

Where M is the length of the moving window, and the ith member of the trend sequence is given by

$$g_{ti} = \frac{1}{M}\sum_{j=i-M}^{i+M} g_{j.} \quad (6)$$

Trend sequences for different data sets used in this research are analyzed in the next section.

Data Analysis and Observations

We selected the evaluations of the 419 instructors who taught at CSI for at least two semesters from Fall 2009 to Spring 2011 and had at least ten evaluations. The first set of experiments was performed for the instructors with at least ten evaluations. This data set included a total of 419 instructors. Figure 1 shows a plot of easiness scores vs. overall evaluations when raw data were used. As expected, the plot does not convey much information; however, a pattern can be seen that indicates an upwards movement in easiness scores with an increase in overall evaluations. A similar behavior can be seen in Figure 2, which shows a plot of average grades vs. overall evaluations. However, the upward movement in average grades with respect to overall evaluations is not as pronounced as that of easiness score shown in Figure 1.

All variables were correlated at the first level with a minimum of 10 evaluations. There was a significant positive correlation between overall evaluation and easiness: r=.68, $t(417)$=18.78, $p < .05$. There was a significant positive correlation between overall evaluation and average grade: r=.38, $t(417)$=8.49, $p < .05$. There was a significant positive correlation between easiness and average grade: r=.40, $t(417)$=8.79, $p < .05$. All variables were also correlated at the second level with a minimum of 20 evaluations. There was a significant positive correlation between overall evaluation and easiness: r=.73, $t(192)$=14.72, $p < .05$. There was a significant positive correlation between overall evaluation and average grade: r=.47, $t(192)$=7.31, $p < .05$. There was a significant positive correlation between easiness and average grade: r=.53, $t(192)$=8.66, $p < .05$. Finally, all variables were also correlated at the third level with a minimum of 40 evaluations. There was a significant positive correlation between overall evaluation and easiness: r=.80, $t(40)$=8.32, $p < .05$. There was a significant positive correlation between overall evaluation and average grade: r=.67, $t(40)$=5.65, $p < .05$. There was a significant positive correlation between easiness and average grade: r=.70, $t(40)$=6.17, $p < .05$.

Our results (Table I) show a strong correlation between the overall evaluation and easiness. The correlation between the overall evaluation and the grades is not as strong for the initial cohort of 419 instructors. Specifically, we found a correlation of .68 between the overall evaluation and easiness; whereas the correlation of .38 was found between overall evaluation and grades. However, this correlation grows stronger when only the instructors with at least 20 evaluations were considered. We observed a continued trend of this sort as the number of the minimum evaluations needed to be part of the cohort continued to increase. Specifically, we found a correlation of .73 between the overall evaluation and easiness; whereas the correlation of .47 was found between overall evaluation and grades.

We further studied the underlying trends in the overall evaluation, easiness, and grades of this cohort of 419 instructors. Figure 3 shows the trend curves for the easiness scores and average grades with respect to overall evaluations. Trend curves in Figure 3 not only show a clear pattern of upward movement in easiness scores and average grades with an increase in overall evaluations, but also they provide interesting insight regarding the relationship between these three measures. One observation can readily be made from Figure 3 that the behavior of the easiness scores and average grades with respect to overall evaluations is not consistent throughout the whole range of overall evaluation scores.

Table I. Results in terms of correlation between overall evaluation and easiness, overall evaluation, and average grade, and easiness and average grade.

Cohort	Minimum Evaluations	Number of instructors	Correlation between overall evaluation and easiness	Correlation between overall evaluations and grades	Correlation between easiness and grades
1	10	419	.68*	.38*	.40*
2	20	194	.73*	.47*	.53*
3	40	42	.80*	.67*	.70*

*Indicates significant correlation, p <.05

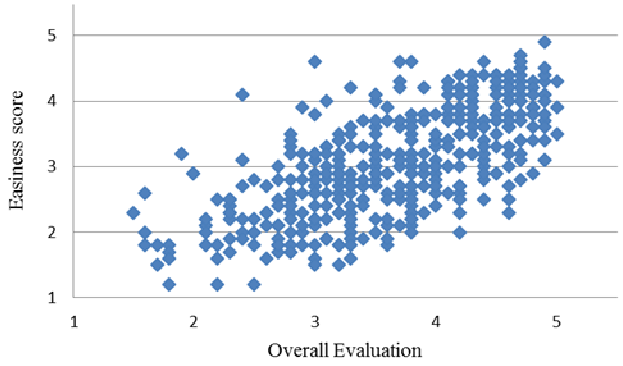

Figure 1. Overall evaluation vs. easiness score (raw data) with the minimum number of evaluations set at 10 (N=418).

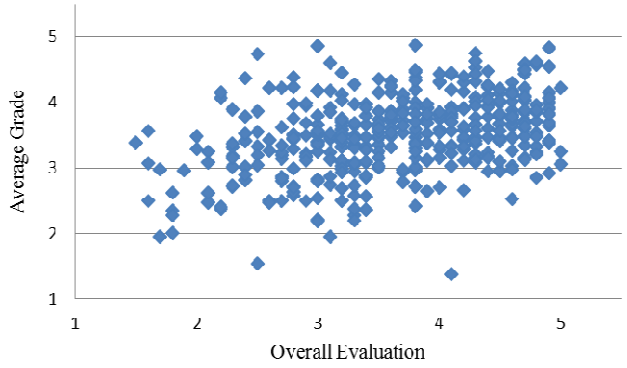

Figure 2. Overall evaluation vs. average grade (raw data) with the number of minimum number of evaluations set at 10 (N=418).

We will now investigate the validity of two common perceptions, one among faculty and the other among students:

1. Instructors receive high evaluation scores from the students because they are "easy graders." Or instructors receive lower evaluations if they are "tough graders" (among faculty).

2. Instructors receiving high evaluations are easy instructor and it will not be difficult to get a high grade in their classes. Or instructors receiving low evaluations are difficult instructors and it will not be easy to get a good grade in their classes (among students).

In order to test the validity of the above mentioned perceptions, we will examine the relationship between grades awarded the instructors and their perceived "easiness" for three cohorts of instructors. As noted earlier, RMP uses a rating scale of 1 (worst) to 5 (best) for its categories. Many institutions, however, use a scale of 1 (worst) to 4 (best). Therefore,

1. Cohort A: Instructors who are perceived to have poor evaluations (an RMP overall evaluation of less than 3.0, which corresponds to an evaluation score of 2.4 or less on the scale of 1thoughg 4).

2. Cohort B: Instructors who are perceived to have average evaluations (an RMP overall evaluation of 3.0 or higher but less than 4, which corresponds to an evaluation score of more than 2.4 but less than 3.2 on the scale of 1though 4).

3. Cohort C: Instructors who are perceived to have good evaluations (an RMP overall evaluation of 4.0 through 5.0, which corresponds to an evaluation score of more than 3.2 through 4.0 on the scale of 1 through 4).

A high correlation between the RMP variables and the average grade was found only in the Cohort B in which instructors received overall evaluation scores in the middle of the range. There was a significant positive correlation between easiness and the average grade in Cohort B: r=.60, $t(12)$=2.58, $p < .05$. There was also a significant positive correlation between overall evaluation and the average grade in Cohort B: r=.72, $t(12)$=3.62, $p < .05$.

Table II shows the correlation between the overall evaluations and grade as well as between easiness scores and grade for, Cohort "A," Cohort "B," and Cohort "C." It can be seen that the correlation between the grades and overall evaluation scores are lowest among Cohorts A and C. That suggests that instructors with high evaluation scores are not necessarily high graders. Also, low evaluation scores for an instructor do not mean that the instructor is a stringent grader.

Table II. Correlation between easiness and grades and overall evaluation and grades for three Cohorts of instructors. Each instructor had at least 40 evaluations.

Correlation	Cohort A N = 11	Cohort B N = 14	Cohort C N = 17
Easiness and Grades	.55	.60*	.28
Overall Evaluation and Grades	.33	.72*	.15

*Indicates significant correlation, p < .05.

However, the instructors with evaluation scores between 3.0 and 4.0 (mid-range) have a very high correlation with grades. Similarly, it can be seen that the correlation between the grades and easiness scores is lowest for Cohort C. That suggests that instructors with high easiness scores are not necessarily high graders. Unlike the overall evaluation scores, there is a high correlation between the grades and easiness scores in Cohort A. That implies that if an instructor has a low easiness score it is likely that students would have to work a lot harder to get a high grade in that class. However, the instructors with easiness scores between 3.0 and 4.0 (mid-range) have a very high correlation with grades.

A multiple regression analysis was also conducted on the sample with a minimum of 40 evaluations demonstrating that overall, the two RMP variables still predicted the average class grade. Together, both RMP variables (easiness and overall evaluation) significantly predicted average grade: $F(2,39) = 21.19$, $p < .001$, $R^2 = .52$.

The trend analysis in Figure 3 also confirms the above analysis in that the easiness scores and grades do not seem to correlate well with the overall evaluations of instructors having high overall evaluations (roughly 4.0 or higher score on overall evaluations). This is an important result from the faculty perspective, which demonstrates that high scores on overall evaluations are not necessarily a result of high grades. This result also provides useful insight from students' perspectives that high scores on overall evaluations (or on easiness) do not mean easy "A" grades.

Finally, we also investigated whether disciplines have any impact on how students evaluate the courses. Specifically, we compared average scores of instructors teaching STEM courses with the average scores of instructors teaching non-STEM courses. We also compared the average grade awarded by instructors in both disciplines. Table III shows mean overall evaluations, easiness scores, and grades awarded by the instructors in STEM and non-STEM disciplines. It can be seen from the Table III that the average grade awarded by the instructors in non-STEM disciplines is higher than the average grade awarded by their STEM counterparts. This is true for all three cohorts. The instructors in non-STEM disciplines also received higher average overall evaluations and easiness scores than their counterparts in STEM disciplines. This leads to the conclusion that STEM courses are perceived harder and students give lower evaluations to STEM instructors. The STEM instructors also award lower grades than their non-STEM counterparts.

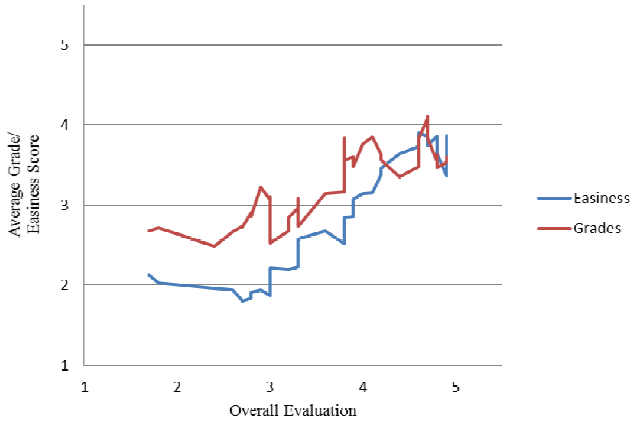

Figure 3. Trends in easiness scores and average grades with respect to overall evaluations (for instructors with at least 40 evaluations; M = 2, N = 43).

Conclusion

This research has several interesting results, which have significant implications for students who rely on RMP rating to select classes. The results suggest that:

- If you take a class with a group "C" instructor (who

Table III. Average RMP scores received and grades awarded by the instructors teaching in STEM and non-STEM disciplines

			Overall evaluation		Easiness scores		Mean grades (GPA)	
Cohort	Minimum Evaluations	Number of instructors	Non-STEM disciplines	STEM disciplines	Non-STEM disciplines	STEM disciplines	Non-STEM disciplines	STEM disciplines
1	10	419	3.76	3.42	3.19	2.83	2.89	2.6
2	20	194	3.78	3.44	3.19	2.79	2.85	2.62
3	40	42	4.02	3.35	3.11	2.63	2.86	2.29

has high RMP overall evaluation scores) hoping you would get an "easy grade," or if you avoid taking a class with a group "A" instructor (who has low overall RMP evaluation scores) assuming that it would be a difficult class in which to get a high grade, you might be in for a surprise.

- Similarly, a high easiness score might not result in a high grade in that class. However, in a class that is perceived to be very difficult (with a low easiness score), the instructor is likely to be a more difficult grader.
- Group C instructors (the instructors who receive very high overall evaluations and/or are perceived to be very easy) are not "high graders."
- Group A instructors who receive low overall evaluations are not "hard graders."
- Group A instructors who receive low easiness scores are likely to be "hard graders."

Students should also keep in mind that the data at RMP is not a result of a scientific survey. They will get much more reliable information about courses and instructors that they are interested in through course evaluations at their institution. Many institutions have made that information public and students have access to that data. If your institution had not done so you can request the Chief Academic Officer at your institution to see the course evaluations.

Our future research direction focuses on extending this research to analyzing the course evaluation data at our institution using the same set of instructors included in this research to further assess the validity of the RMP evaluations or the lack thereof.

References

Abrami, P. C., & Mizener, D. A. (1985). Student/instructor attitude similarity, student ratings, and course performance. *Journal of Educational Psychology*, 693-702.

Bleske-Rechek, A., & Fritsch, A. (2011). Student Consensus on RateMyProfessors.com. *Practical Assessment, Research and Evaluation*, Vol. 16, No.18, pp. 1-12.

Bleske-Rechek, A., & Michels, K. (2010). RateMyProfessors.com: Testing Assumptions about Student Use and Misuse. *Practical Assessment, Research and Evaluation*, Vol. 15, No. 5, pp. 1-12.

Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching. *Practical Assessment, Research and Evaluation*, Vol.12, No. 6, pp. 1-15.

Davison, E., & Price, J. (2009). How do we rate? An evaluation of online student evaluations. *Assessment & Evaluation in Higher Education*, Vol. 34, No. 1, 51–65.

Felton, J., Mitchell, J., & Stinson, M. (2004). Web-based student evaluations of professors: the relations. *Assessment and Evaluation in Higher Education*, 29: 91–108.

Goldman, L. (1990). Student evaluations of their professors rarely provide a fair measure of teaching. *Chronicle of Higher Education*, August 8, B2.

Kwan, H. K., & Arniker, S. B. (2009). Numerical Representation of DNA Sequences. IEEE Transactions on Biomedical Engineering, 307-310.

Sanders, S., Walia, B., Potter, J., & Linna, K. W. (2011). Do more online instructional ratings? *Practical Assessment, Research and Evaluation*, Vol. 16, No. 2, pp. 1-6.

Stone, C. A., & Yang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. *Practical Assessment, Research & Evaluation*, 18(13), 1-12.

Thoemmes, F., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, 46(1), 90-118.

Xu, S., Ross, C., Raebel, M. A., Shetterly, S., Blanchette, C., & Smith, D. (2010). Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value in Health*, 13(2), 273-277.

Acknowledgment:

The author thanks Kevin Strubler, Institutional Research Specialist at the College of Staten Island for assistance in data analysis.

Citation:

Rizvi, Syed A. (2015). RMP Evaluations, Course Easiness, and Grades: Are they Related?. *Practical Assessment, Research & Evaluation*, 20(20). Available online: http://pareonline.net/getvn.asp?v=20&n=20.

Corresponding Author

Syed A. Rizvi
Associate Provost for Institutional Effectiveness
Office of Academic Affairs, Room 1A-305
College of Staten Island/CUNY
2800 Victory Blvd., Staten Island, NY 10314

email: Syed.rizvi [at] csi.cuny.edu