

# **Self and Peer Assessment of Group Work in Higher Education: A Game Theoretic Analysis**

Mehdi Chowdhury  
Bournemouth University

The paper demonstrates that the Self and Peer Assessment (SPA) method used in higher education, to map a group work mark to individual marks, can easily be modelled as a strategic form game. This modelling predicts students to report SPA following their dominant strategies. However, data of a real time SPA indicate that students may not report SPA by adopting dominant strategies. The findings are indicative of a mismatch between the game (SPA) designed by the tutor, and then played by students. The paper concludes that the interpretation of SPA is not possible, and statistics of SPA should not be relied upon to map a group mark to individual marks.

**Key Words:** Self and Peer Assessment, Dominant Strategy, Nash Equilibrium

**JEL Classification:** A23, I21

**Correspondence to:** Mehdi Chowdhury, Business School, Bournemouth University, D165, Dorset House, Talbot Campus, Fern Barrow, Poole, BH12 5BB, UK. E-mail: [mchowdhury@bournemouth.ac.uk](mailto:mchowdhury@bournemouth.ac.uk)  
Tel: +44 (0)1202 961 397

## **Acknowledgement:**

The author is thankful for the comments and suggestion received from Tim Lloyd, Anna Bennato, Peter Howard-Jones and Giuseppe Bova. All the remaining errors are the author's.

# **Self and Peer Assessment of Group Work in Higher Education: A Game Theoretic Analysis**

## **Abstract**

The paper demonstrates that the Self and Peer Assessment (SPA) method used in higher education, to map a group work mark to individual marks, can easily be modelled as a strategic form game. This modelling predicts students to report SPA following their dominant strategies. However, data of a real time SPA indicate that students may not report SPA by adopting dominant strategies. The findings are indicative of a mismatch between the game (SPA) designed by the tutor, and then played by students. The paper concludes that the interpretation of SPA is not possible, and statistics of SPA should not be relied upon to map a group mark to individual marks.

**Key Words:** Self and Peer Assessment, Dominant Strategy, Nash Equilibrium  
**JEL Classification:** A23, I21

## **1. Introduction:**

Group work based assessments are frequently used in institutions delivering higher education, specifically in Business and Economics disciplines. In a group work based assessment, a group of students jointly perform a task, i.e. essay, report or presentation. The task is assigned a group mark by an assessor (tutor) and is later mapped to individual marks. To facilitate this mapping, many assessors explicitly or implicitly depend on a mechanism known as the Self and Peer Assessment (SPA)<sup>1</sup> which essentially is comprised of students' reports about each other's contribution to the group work.

The usefulness of SPA has always been a highly debated topic in the education literature. This paper identifies that substantial insight about the SPA mechanism can be obtained utilising a game theoretic approach. Specifically, SPA can easily be modelled as a strategic (or normal) form game where students are the players, an individual's self and peer assessments are the strategies, and each combinations of strategies generate utilities through individual marks. However, to date no game theoretic paper has attempted to address this practice that impacts on the lives of so many students. This paper appears to be the first paper to demonstrate this aim.

The modelling of SPA as a strategic form game allows us to make a comparison between the predicted and actual outcomes of an SPA. The paper shows, in a generic SPA, the dominant strategy of a student is always to report the highest possible self and the lowest possible peer ratings. Therefore the predicted Nash equilibrium is characterised by all students reporting the maximum possible

---

<sup>1</sup> Also known as self and peer ratings.

self and the lowest possible peer ratings. This prediction is compared with the outcomes of a real time SPA. The comparison however reveals a mismatch between the predicted and actual outcomes, that is, students instead of playing dominant strategies reported each other as equal. This mismatch between predicted and actual outcomes indicates the presence of behavioural factors (e.g trust, reciprocity, altruism and guilt aversion) which have not been taken into consideration by the tutor. We infer that the game designed by the tutor and the game played by students differ, therefore the interpretation of the SPA is not possible in line with the design perceived by the tutor. This formulation of SPA, in this paper, as a strategic form game confirms the issues already known to the literature, however it is also indicative of the futility of SPA reliant mapping of group marks to individual marks.

The paper is therefore in opposition to the numerous attempts (see Spater et al. 2015 for a recent review) made by the education literature to use self and peer assessment to map group marks to individual marks. We argue that these mappings are inherently problematic as they are designed without any understanding of how marks and utilities of students are interrelated. It is already known to the psychologists, economists and educators (Magin 2001, Hanrahan and Isaacs 2001, Falchikov 2005, Fehr and Schmidt 2006) that the individuals/students not only care about their own satisfaction but also satisfaction of their peers. Hence, any mechanism not generated from a proper understanding of this interrelationship should not be relied upon, which is exactly the case with the utilisation of SPA in mapping the group work marks.

The paper is structured as follows. The second section demonstrates that a generic design of SPA can be easily expressed as a strategic form game, and identifies the dominant strategies and Nash equilibrium. In the third section, we look at the SPA of a case study and expressed that design as a strategic form game. We then identify the Nash equilibrium, dominant strategies and analyse the data. Section four discusses the result and section five concludes.

## **2. SPA as a Strategic Form Game:**

SPA asks individuals to rate self and peers according to their contribution to the group. The concern of the paper is the game playing between students when reporting these ratings. The SPA mechanism depends on the belief that the students will be truthful in their reports. However, it is not ensured in a generic SPA mechanism<sup>2</sup>.

As has been stated previously, a generic SPA can be easily viewed as a strategic form game as it contains (1) a set of players (2) strategies for each player and (3) utilities for each player for each combination of strategies. We are interested to know the characteristics, specifically of the Nash equilibrium of this SPA. It will allow us to identify if the Nash equilibrium constitutes truth telling of students in that specific design.

---

<sup>2</sup> For a generic design see Eberly Center (2018)

To elaborate further by developing a generic SPA, let us assume followings:

- I. There are  $n$  number of students in a group.
- II. The reported self-assessment of the student  $i$  ( $i = 1, 2, \dots, n$ ) is  $x_{ii} \in [0, x]$ .
- III. The reported peer-assessment of the student  $i$  about  $j$  ( $j \neq i$ ) is  $x_{ij} \in [0, x]$ .
- IV. The actual self and peer assessments are  $v_{ii} \in [0, x]$  and  $v_{ij} \in [0, x]$ .

(II) and (III) implies that the contributions reported are elements of the interval from 0 to  $x$ . This interval is predefined by the assessor; for example from 0 to 10. Note that the actual SPA and reported SPA are not necessarily the same, however, the assessor expects students to set  $x_{ii} = v_{ii}$  and  $x_{ij} = v_{ij}$ , that is to report truthfully. The design assumes that students are capable of assessing self and peer contributions appropriately.

Therefore, this design is explicit about the two elements of a strategic form game i.e. players and strategies. It surely has the third element i.e. the utility associated with the combinations of strategies. However, as stated earlier, no such assumption regarding the utility can be found in generic SPA designs.

To complete our description, let us assume that the utility of a student increases when his/her own mark increases and decreases when his/her own mark decreases. Additionally, a student is assumed to be indifferent to marks of other group members, that is, marks of peers do not affect their level of utility.

## 2.1. Self-assessment only:

If the assessor only uses self-assessment, truthful reporting i.e.  $x_{ii} = v_{ii}$  is not a dominant strategy when  $v_{ii} < x$ . To prove it assume :

- V. The award scheme is a function  $\phi_i : x_{ii} \rightarrow R_+$ . Assume it is increasing in  $x_{ii}$ .

(V) implies, as  $\phi_i$  is increasing in  $x_{ii}$ , the dominant strategy of a student is to set  $x_{ii} = x$  as the utility of a student is increasing in his/her own mark. Hence truthful reporting is not a dominant strategy when the assessor uses self-assessment only.

## 2.2. Combination of self and peer assessment:

In the case of using of both self and peer assessment, again truthful reporting is not a dominant strategy, rather the dominant strategy is to report in a way that maximises own mark. To prove it assume:

VI. An award scheme  $g_i: R_+^n \rightarrow R_+$  and  $v_i: g_i \rightarrow R_+$ . Assume that  $g_i$  is increasing in  $x_{ii}$  and  $v_i$  is increasing in  $g_i$ .

(VI) implies that the assessor uses the ratings of self and peer assessment to create a representative number  $g_i$ . The representative number is then used to award the grade  $v_i$  to the student  $i$ . It is now obvious the dominant strategy of  $i$  is to report  $x_{ii} = x$  for any  $x_{ji}$ , i.e. the peer assessment by the other members of the group.

Additionally if  $g_i$  is decreasing in  $x_{ij}$ , the mark is maximised by the lowest possible peer rating. Hence dominant strategy of a student is to report the maximum possible self and the lowest possible peer ratings.

### **2.3. Peer-assessment only:**

From 2.2 it is now clear that peer-assessment only cannot also be truthful when associated to final individual marks. Actually, any association of self/peer assessment with the final marks incentivises non-truthful reporting.

### **2.4. Nash equilibrium:**

This section therefore demonstrates that in a generic SPA design, the dominant strategy of a student is report the highest possible self and the lowest possible peer ratings. These reports are not necessarily truthful. The Nash equilibrium of a generic SPA is therefore characterised by all players playing their dominant strategies aiming to receive the highest possible mark regardless of actual self and peer contributions.

Note that the assumption on the relationship between mark and utility has a significant role in identifying the Nash equilibrium. Without this assumption highest possible own mark may not be the aim of a student's report. This is actually the main reason behind of criticising the practice SPA in this paper. If the tutor, who designed the SPA, does not know about the relationship between utility and mark of a student, then the tutor does not in essence know which game the students have played. Therefore, interpretation of rating of SPA is questionable in accordance with the perceived game structure of the tutor.

## **3. A Case Study:**

The section evaluates a case study of a real time SPA. Our aim is to demonstrate that the case study resembles the generic design of section 2. We then identify the characteristics of the Nash equilibrium and compare that with actual ratings or outcomes.

### 3.1. The Basic Instructions<sup>3</sup>:

The students were asked to deliver a 3000 word group essay. Each group consists of 4 members. Students were asked to submit confidential self and peer-assessment on each other's contributions by emailing the tutor. The tutor instructed that individual marks will be adjusted upward or downward based on SPA, though the actual marking scheme was not made explicit. It was also stated that the unit tutor may conduct face to face interviews to ask for justification of the ratings. The ratings are as follows:

<b>Self-assessment</b>	<b><u>Scores</u></b>
Much more than all other group members:	25
More than other group members:	20
Same as other group members:	15
Less than other group members:	10
Much less than all other group members:	5
<b>Peer assessment</b>	
X contributed much more than all other group members:	25
X contributed more than other group members:	20
X contributed the same as other group members:	15
X contributed less than other group members:	10
X contributed much less than all other group members:	5

For marking, the tutor added the SPA received by a student and then divided by the group's total SPA. For example if the SPA of a student is 60 and group's total SPA is 150, the ratio is 0.40 implying that the student did 40% of the work and accordingly the mark of the student is adjusted upward. Note that this method is similar to the method stated in section 2.2.

### 3.2. Dominant Strategies and Nash Equilibrium:

We first establish that this design is similar to the generic design illustrated in section 2. This design has 4 players. Each player has a set of strategies. Therefore the two elements are explicit in the instructions. However no assertion has been made about the relationship between utility and marks. As in the section 2, let us assume the utility increases with marks and students are indifferent about their peers. Hence, this SPA is completely specified as a strategic form game.

---

<sup>3</sup> The instruction has been presented in a modified manner to preserve anonymity.

In reporting SPA, each student selects self and peer ratings from the set of numbers 5, 10, 15, 20 and 25. In a group of 4 students, the number of n-tuples is therefore  $5^4=625$ . The strategies of students can be denoted by the set

$S_m = \{s_{m1}, s_{m2}, \dots, s_{m625}\}$  where,  $m=1,2,3,4$  are the members and  $i=1,2,\dots,625$  indicate index of the strategies.  $s_{mi}$  is a row vector with 4 elements,

$s_{mi} = \{a_{m1}^i \ a_{m2}^i \ a_{m3}^i \ a_{m4}^i\}$ . It implies that each time, a student  $m$  picks 4 numbers, one assigned to self and the others to the group members and a student can do it in 625 different ways. The vectors arrange the elements in order by indexing the members from 1 to 4. For example,  $a_{m2}^i$  implies the rating awarded by the student  $m$  to the student indexed as 2 in the  $i$ th strategy. In a group of 4, the final outcome of the SPA can be demonstrated by the following  $4 \times 4$  matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

We have avoided writing the upper subscript of elements as they are not essential in further analysis and make the matrix clumsy. The diagonal terms of the matrix are self-assessments and the off diagonal terms are peer-assessments. Each row of the matrix is now the reported ratings by a student. Note also that the set of outcomes of SPA with 4 students consists of a total of  $625^4$  possible matrices.

The SPA of the student indexed as 1 is the sum of the first column of  $A$ , i.e.  $\sum_{m=1}^4 a_{m1}$ . The total SPA of the group is,  $\sum_{m=1}^4 \sum_{m=1}^4 a_{mm}$ . The ‘contribution ratio

(CR)’ of a student (here indexed as 1) is,  $\frac{\sum_{m=1}^4 a_{m1}}{\sum_{m=1}^4 \sum_{m=1}^4 a_{mm}}$ .

This CR increases as the self-rating goes up and falls as the peer rating goes up. This ratio is used by the tutor to adjust the marks upward or downward, therefore the dominant strategy of a student is to report the maximum possible self (here 25) and the lowest possible peer ratings (here 5) as the utility is monotonically increasing with marks. The Nash equilibrium of the game has the following form:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} 25 & 5 & 5 & 5 \\ 5 & 25 & 5 & 5 \\ 5 & 5 & 25 & 5 \\ 5 & 5 & 5 & 25 \end{pmatrix}$$

The CR of an individual student is  $\frac{\sum_{m=1}^4 a_{m1}}{\sum_{m=1}^4 \sum_{m=1}^4 a_{mm}} = \frac{40}{160} = 0.25$ . If students

play their dominant strategies, the CR of a student is 25% implying that the mark of a student is the same as the other group members. Nevertheless, it is the only Nash equilibrium of this SPA design, as any deviation from the dominant strategy will decrease own marks and consequently utility. The above matrix is the benchmark for the subsequent data analysis. Based on identification of the Nash equilibrium, we predict to observe a matrix similar to the matrix A.

### 3.3. Analysis of Data:

This section analyses the data of the above SPA. The total number of students submitting the SPA was 60. Each group had 4 students with the exception of 4 groups, each of which had 3 students. We dropped the groups with 3 students to ensure uniformity of group size. The analysis therefore uses reports of 48 students in 12 groups.

**{Insert Figure 1 and 2 here}**

Figure 1 and Figure 2 show the histograms of self and average peer-assessment. Our prediction from the previous analysis implies that the self-assessment will be 25 and average peer-assessment will be 5. However the majority of the students (30 students) have reported 15 as the self-assessment implying that they stated their contributions are same as others. Interestingly, a majority of students (33 students) also reported the peer contribution equal to 15. This implies that students rated each other equal which is contradicting with our prediction.

**{Insert Table 1 here}**

The statistics have been further presented in the Table 1. As can be observed, there exists no significant difference in the means of self and average peer assessments. The average peer assessment has less spread than the self-assessment with should be case for an average. Interestingly one student assessed self only 5, that is the minimum, however none reported all peer contribution equal to 5 as reflected in the minimum of average peer assessment of 11.67.

**{Insert Figure 3 here}**

The final diagram we will look at is Figure 3. The figure shows the assessment of self, compared to the assessment of peers by the students. In the legend, 'assessed1Above' implies that the students have reported only 1 peer above self. From the figure, we see that 3 students have done so. Similarly, the



figure shows that 6 students have reported 2 peers above self and 2 students have reported all 3 peers above self. 37 students have assessed none above self and 22 students have reported all group members equal. It, therefore, shows that students have demonstrated a tendency to assess all equally with a slight tendency to report self over the peers. Interestingly, as discussed previously, when the students report an equal contribution of all the group members, the outcome yields the same marks as the Nash equilibrium of the SPA. With equal valuations, the matrix is,

$$A = \begin{pmatrix} 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 \end{pmatrix}$$

The CR of a student is  $\frac{60}{240} = 0.25$  implying that that the group members receive equal marks. However, the predicted and actual structures of the matrices differ greatly. In the next section we attempt to shed some lights on possible reasons of this difference.

#### 4. Discussion:

We predicted that students will not truthfully report the ratings, and it will be done in a manner that will demonstrate the willingness to maximise marks. However the matrix above does not demonstrate that predicted pattern. It shows that students in general reported all group members' contributions as equal, hence they definitely did not make untrue reports the way we predicted. Does it mean that they lied in a different manner or the reports we observed are indeed truthful? Our answer to this question will be that we simply do not know and the reasons are explained below.

Our predicted matrix crucially depended on the assumption that utility of student increases with own mark and independent of the marks of peers. However, a large number of papers through experiments (See Fehr and Schmidt 2006) have established that individuals do care about the utility of others. In addition behavioural factors such as trust, reciprocity altruism and guilt aversion influence human decision. Given the experimental findings, the assumption that students are indifferent about marks of peers seems an invalid assumption.

Note that we made this assumption to facilitate identification of the Nash equilibrium of our perceived game. SPA designs in practice would make no reference to the relationship between marks and utility.

Interestingly, the education literature has already identified that behavioural factor such as trust, reciprocity, altruism and guilt aversion may influence SPA. For example Magin (2001) and Falchikov (2005) mentioned about reciprocity bias that arises as a result of friendship and social interaction accompanying group task activities. Hanrahan and Isaacs (2001), identified guilt aversion as students expressed discomfort in criticising and assessing others'

performance as poor. However, the literature failed to recognise that without a proper understanding of the influence of these factors on reports of students, any design is deemed to be just an ad-hoc design.

The actual game structure perceived by the students is simply unknown to us. To see it further, note that the instruction set of the case study also resembles that of a trust game (Berg et al. 1995, Fehr and Schmidt 2006), where an individual would trust another individual to return a monetary sum in the first period and the other individual would reciprocate by returning the money in the second period. In SPA, a tutor may trust his/her students and expects them to reciprocate by reporting the true contribution. However, students may also be playing a trust game with each other where the trust is achieved through pre-communication and friendship. In this regard, SPA can be viewed as a part of a large repeated game, where threats and punishments act as enforcement mechanisms to ensure equality of reporting.

All the above is indicative of the existence of a game structure significantly more complicated than as it appears at the outset. The problem is the lack of understanding of the relationship between marks and utility, and of the game structure. Therefore the actual meaning of the reporting of students in relation to their contributions is also unknown. As the meanings are unknown, the ratings should not be used to map group marks to individual marks.

The discussion in this section is therefore also indicating that it is not possible to develop a reliable method to map group marks to individual marks using SPA. A robust design would require a proper understanding of the underlying game and unless such an understanding is attained, any attempt to develop a method of mapping using SPA is futile.

## **5. Conclusion:**

SPA is utilised in higher education to map group marks to individual marks. The paper demonstrated that the SPA can be modelled as a strategic form game. If a student is indifferent about the marks of peers and the utility is increasing in own mark, the dominant strategy is to report the maximum possible self and the lowest possible peer ratings. However analysis of a case study demonstrated that in a real time SPA, students did not play their dominant strategies. The result indicated that students are likely to be not indifferent about the marks of peers and there exists a different underlying game relative to what has been perceived by the tutor. We therefore conclude that in a group work based assessment, the meaning of SPA is unknown, and as such SPA should not be used to map groups marks to individual marks.

The analysis of the paper indicated that any attempt to utilise SPA is futile. However many (including the author of this paper) maintain the view that group work enhances the learning experience and interpersonal communication skill of students. This practice of group work therefore needs to be maintained, however there needs to be a method to map group marks to individual marks. In this regard

the usefulness of self and peer assessment needs to be further analysed, as it is influential in determining the degree classification of students in higher education. We anticipate that this paper will attract the attention of other researchers to conduct further work on this vital issue, combining game theory and behavioural economics with the higher education literature.

## References

- Berg, J., J. Dickhaut, and K. McCabe. 1995. "Trust, reciprocity and social history." *Games and Economic Behavior* 10(1): 122-142.
- Eberly Center. 2018. "Assess Learning & Teaching." *Eberly Center for Teaching Excellence and Educational Innovation*, Carnegie Mellon University, Available on <http://www.cmu.edu/teaching/assessment/index.html>. Date accessed 16 March, 2018.
- Falchikov, N. 2005. "Improving assessment through student involvement: practical solutions for aiding learning in higher and further education." Routledge Falmer.
- Fehr, E. and K. Schmidt. 2006. "The economics of fairness, reciprocity and altruism: experimental evidence and new theories." In S. C. Kolm and J. M. Ythier (Eds.), *Handbook of Economics of giving, altruism and reciprocity*, pp. 615-691. North-Holland.
- Hanrahan, S. and G. Isaacs. 2001. "Assessing Self-and Peer-assessment: the students' views." *Higher education research and development* 20(1): 53-70.
- Magin, D. 2001. "Reciprocity as a source of bias in multiple peer assessment of group work." *Studies in Higher Education* 26(1): 53-63, DOI: 10.1080/03075070020030715.
- Spatar, C., N. Penna, H. Mills, V. Kutija, and M. Cooke. 2015. "A robust approach for mapping group marks to individual marks using peer assessment." *Assessment & Evaluation in Higher Education* 40(3): 371-389. DOI:10.1080/02602938.2014.917270

Figure 1: Distribution of Self Assessment scores

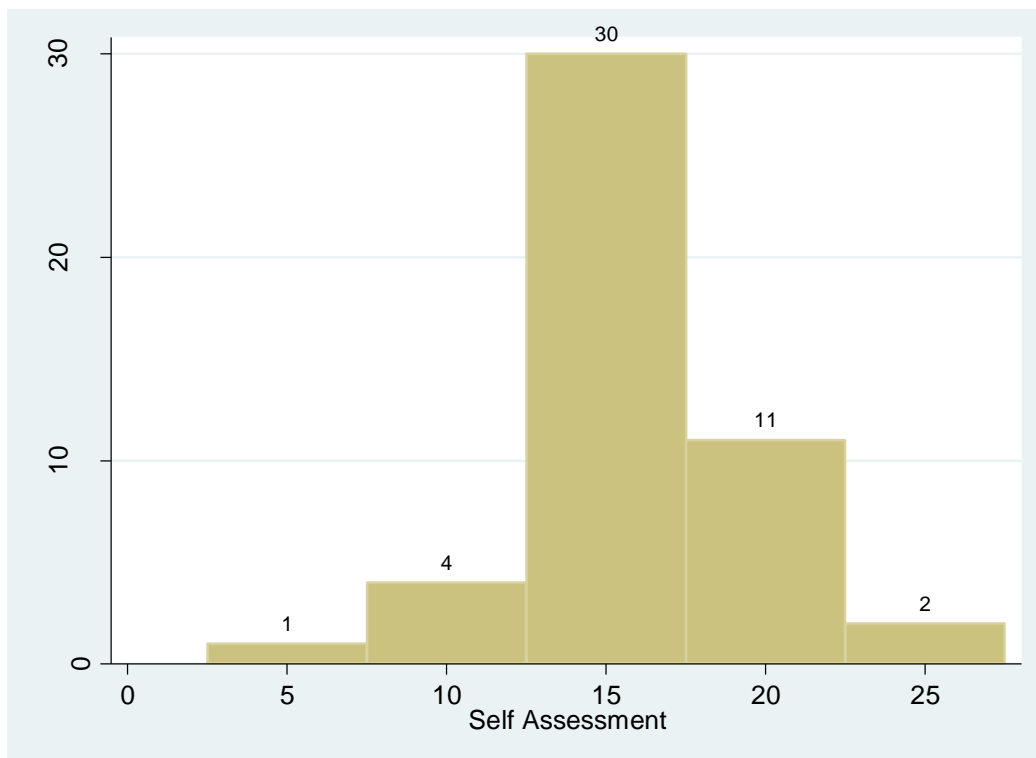


Figure 2: Distribution of Average Peer Assessment scores

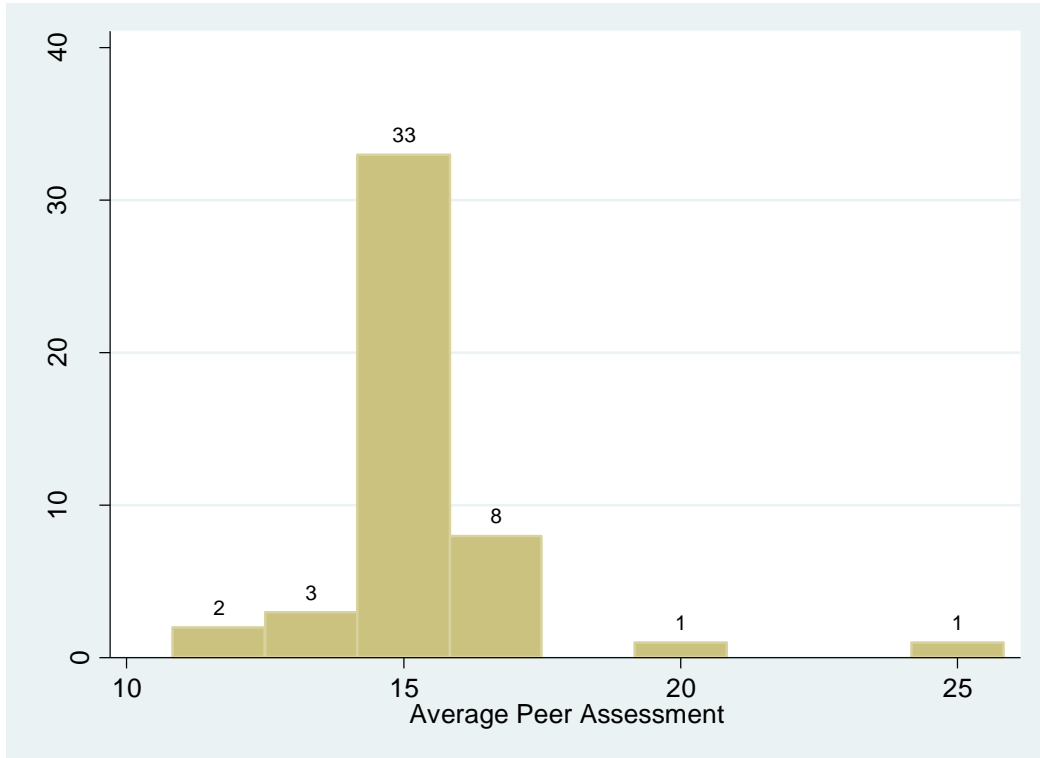


Table 1: Summary statistics of self and peer assessment scores

	Observations	Mean	Standard Dev.	Minimum	Maximum
Self Assessment	48	15.94	3.67	5	25
Peer Assessment	48	15.35	1.91	11.67	25

Figure 3: Assessment of self compared to the peers

