

IS IT SAFE TO USE PEER ASSESSMENT OF INDIVIDUAL CONTRIBUTION LEVEL WHEN ASSESSING GROUP WORK?

M.P. Garcia-Souto

*Medical Physics and Biomedical Engineering Department, University College London
(UNITED KINGDOM)*

Abstract

The use of group work activities is increasing in Higher Education because of well-known educational benefits, e.g. students have to apply technical knowledge to the project at hand engaging them in deeper learning, as well as they develop team work and other professional skills. However, the assessment of the group work is challenging. Giving the same mark to all group members causes concerns among students and staff, both in terms of 'free riders' and mark fairness. One way of addressing both these issues is to use Peer Assessment of Individual Contribution to group work, namely IPAC. However, many academics are reluctant to use this methodology. Two main reasons are (i) that some feel nervous about giving "power" to the students to mark peers and how this would affect the final marks, (ii) that implementing this method can be administratively quite time consuming. This paper presents insights in both these concerns. This is of interest to anyone organizing and running assessed student group work activities, and that is using or might want to use in the future the IPAC methodology.

Keywords: Group Work, Individual Contribution, Assessment, peer assessment, IPAC.

1 INTRODUCTION

Higher Education is transforming as students of all disciplines are expected to have relevant technical or topic knowledge but also personal and professional skills by the time they graduate. Some of these skills are teamwork, ability to communicate and provide constructive feedback, ability to apply their knowledge within of the context of their field, etc. This is particularly of essence such that students' knowledge and abilities align with industry expectations [1]. Because of this, group work activities are increasingly used as they provide a great learning and training opportunity to students in both areas, i.e. knowledge and soft skills, as well as presenting students with projects that are more meaningful and closer to real life situations. However, the assessment of group work can be challenging, with problems particularly arising when all members of a group get the same marks at the end. Often there are "passengers" or "free riders" in the group, which cause general concerns among staff and students about the fairness of this method of assessment [2-4] and poor student experience. This also raises concerns with external examiners and professional accrediting bodies. A university-wide consortium (IPAC Consortium) was formed at University College London to look into this issue as it is a common problem across disciplines. Members of this consortium include academics, teaching fellows, educational experts, students and technologists.

Many of the challenges related to the engagement and assessment of group work can be addressed by the use of the IPAC methodology, i.e. incorporating an element of **I**ndividual **P**eer **A**ssessment of **C**ontribution to the group work. This method combines (a) a group mark given by the tutor for the outcome of the group with (b) an individual score for each student given by peers based on his/her contribution level (IPAC score). This is requested and welcome by students [4,5] and it improves student experience and team effectiveness [6]. The IPAC methodology is conceptually simple and its implementation can be easily customized to the staff preferences and activity requirements. The paper by Garcia-Souto et al. (2019) [7] shows the key points of the methodology, and the different options available, as well as it gives guidelines on how the IPAC practice should be implemented. However, there are two main reasons that discourage academics to some extent from implementing this method. On one hand, many academics feel nervous about giving "power" to the students to modify the marks of their peers, particularly because "mathematically and in theory" their effect can be very large. On the other hand, this method can be administratively quite time consuming if one does not use appropriate software. This paper presents insights in both these concerns.

This paper studies what the peer marks effect is in practice, and provides evidence based on several group work activities run in the academic year 2017/18 in various disciplines by IPAC Consortium members. It puts current and future practitioners at ease by showing that typically the IPAC scores are

in a narrow range, with only very large or very low scores in exceptional cases. The paper also investigates via statistical analysis the effect of several parameters into the IPAC scores observed, e.g. if the IPAC values are used as *added percentages* or *normalized factors* (these are the two main IPAC application approaches, and they are defined later), and the class size. Results are compared with others in the literature. We finally introduce the software that we are currently using at UCL to implement the IPAC methodology in a way that is easy and time efficient for the staff.

2 METHODOLOGY

This study aims to provide evidence-based understanding of the quantitative marks that students obtain as a result of the Individual Peer Assessment of contribution to group work (named as IPAC value) and how this affects the mark that they receive for the overall group project. It also looks at the effect of the class size, and how the IPAC value is computed and applied to the group mark.

2.1 IPAC methodology: Main calculation and application approaches

There are two main approaches in which the IPAC value can be computed and combined with the group mark, these are as (i) an *added percentage*, and (ii) a *normalized factor* [7]. In either case, students are asked to assess their peers (and in cases also themselves) using a rating scale on one or several attributes related typically to contribution level and professional behaviour. In the first case, the scores received by a student from peers are averaged and the percentage calculated with respect to the maximum number of scores that a student could obtain. In the second case, the scores received by a student are averaged and then divided by the average score obtained by students in that group, i.e. normalized. The IPAC value is then combined with the group work as shown in equation 1 and 2 for the options of *added percentage* and *normalized factor* respectively. The main difference is that and IPAC mark is seen as always increasing the mark when using the first method, i.e. *added percentage*, while when used as *normalized factor* it will either increase or decrease the mark of the individual depending on whether his/her contribution was more or less than the average of the group members [7]. Other formulas and moderations can be applied, but they are based on either of these approaches [7].

$$\text{Individual mark} = a \times \text{group mark} + b \times \text{IPAC} \quad (\text{Eq 1})$$

where $a = \text{cte}$; $b = \text{cte}$; $a + b = 1$; IPAC is given in percentage

$$\text{Individual mark} = \text{group mark} \times \text{IPAC} \quad (\text{Eq 2})$$

where IPAC is normalized and given around 1, this representing the average contribution

2.2 Collection and analysis of case studies at University College London

Data was collected from 11 group work activities that took place across the University College London in 2017/18, and where the IPAC assessment method was used. 10 of the activities were run at departmental level with cases from 6 different departments, while the 11th activity involved all the engineering students in year 1 from 7 different engineering departments. A list of the activities is given in Table 1 with basic information about the activity, class and group size and assessment method (if the IPAC value was used as an *added percentage* or as a *normalized multiplying factor*).

The data used for the analysis were the calculated IPAC values assigned to each student before any tutor moderation, i.e. one single IPAC value per student obtained as the average of the raw individual peer scores received. Data was grouped according to the type of IPAC value used, i.e. normalized or percentage, and statistically characterized. In addition, the effect of the class size was studied in the particular case where a normalized value of IPAC was used. For that, data from relevant case studies was grouped according to the class size and statistically compared, with cohorts being defined as (a) small class, with less than 40 students; and (b) medium class, with number of students between 40 and 100.

Table 1. IPAC assessment case studies included in analysis.

#	Department	N students	Group size	Year	Project length (weeks)	IPAC method
1	Biomedical Eng	22	3-4	2	18	Normalized
2	Biomedical Eng	13	4-5	3	20	Normalized
3	Biomedical Eng	21	4-5	2	1	Normalized
4	Civil, Environ.&Geomatic Eng	80	8-9	1	1	Normalized
5	Civil, Environ.&Geomatic Eng	79	8-9	1	1	Normalized
6	Civil, Environ.&Geomatic Eng	79	3-4	2	1	Normalized
7	Biochemical Eng	20	3-4	1	6	Normalized
8	UCL Culture	41	5-6	2	5	Normalized
9	Economics	67	5-4	3	4	Normalized
10	Economics	59	5-4	3	8	Normalized
11	Engineering Fac	714	4-6	1	5	Out of 100

2.3 Effect of the IPAC values onto individual marks

The effect of the IPAC methodology onto the final student marks was investigated based on the statistics obtained from the study described in section 2.2. Examples are given for a poor, an average and an outstanding group. Again, the IPAC values used have not been tutor moderated.

2.4 IPAC software

All UCL case studies reported in this paper used the UCL-developed IPAC software. This software (depicted in Fig. 1) was designed based on the staff, student and administration requirements identified following the conversations maintained by the IPAC Consortium members at UCL. The software and design requirements are presented in the conference paper by Garcia-Souto (2019) titled “Making assessment of group work fairer and more insightful for students and time-efficient for staff with the new IPAC software” [8]. The software key features are that it allows for full customization, and implements the IPAC methodology in a time-efficient way for staff, and insightful for students by facilitating quick and personalized feedback to students. Anyone interested on knowing more or potentially using it can contact the corresponding author.

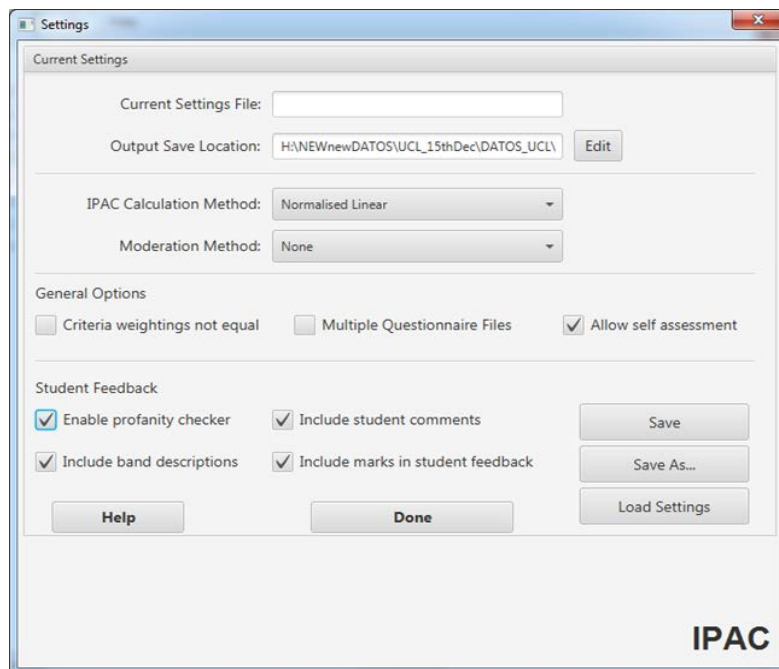


Figure 1. Settings screen of the UCL IPAC software. Copyright © 2018 UCL [8].

2.5 Comparison with other studies in the literature

Some studies were found in the literature that report statistical data of IPAC values in respective case studies. However, the information available is in different formats and for different cohorts of students (e.g. some separate female from male students). In an attempt to compare results from this and other studies, some basic parameters were selected as required descriptors of the IPAC value distribution, i.e. number of students, average and SD of the IPAC score, and minimum and maximum values of IPAC scores when available. These values were obtained from our data sample, and collected and/or estimated when necessary from the information available in other studies.

The IPAC values reported are either in percentage or in normalized factor form, depending on how the tutors set up the assessment in their group work activity.

3 RESULTS

3.1 Analysis of case studies at University College London

For each activity, the IPAC value per student was calculated either in normalized or percentage form, depending on what the activity lead advertised to the students at the start of the project. The distribution of marks in each case is shown in Fig. 2, and the corresponding percentiles given in Table 2. The distribution of normalized IPAC values for two different class sizes (small class meaning <40 students, and medium class meaning 40-100 students) is presented in Fig. 3. Statistical analysis shown the distributions are not significantly different at $p < 0.01$ level.

All values presented in this analysis are the raw IPAC values, i.e. the values calculated using the raw scores that all students gave to each other, without applying any tutor moderation or removing any cases such as those that might have been dealt with via extenuating circumstances protocol.

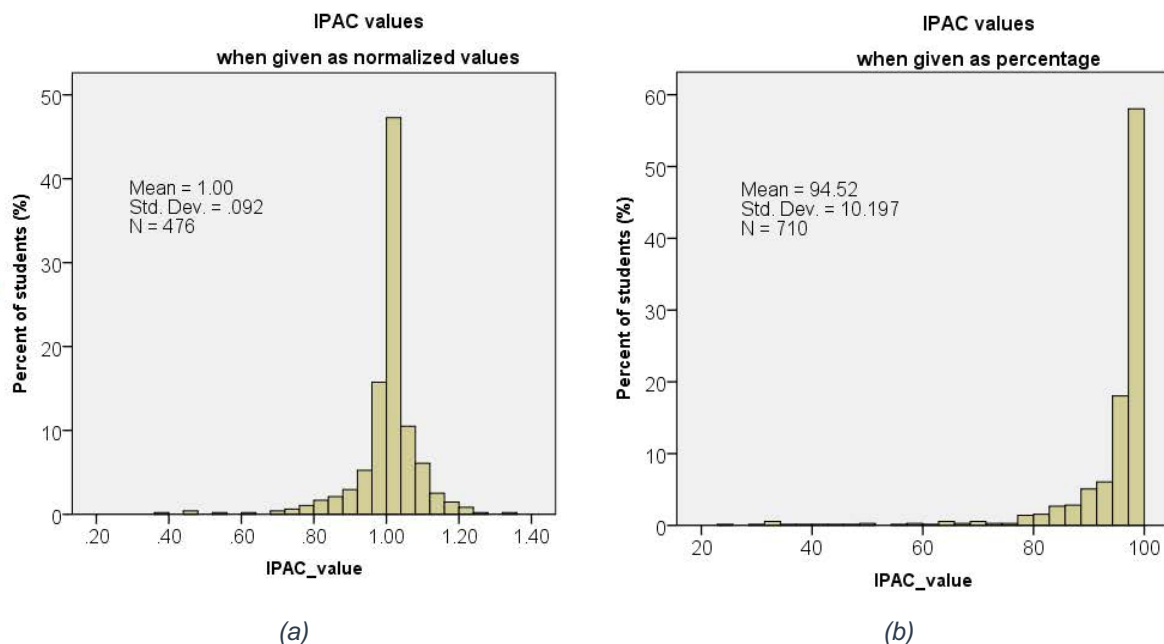


Figure 2. Distribution of calculated IPAC values for all activities that used (a) normalized IPAC value; or (b) percentage IPAC value.

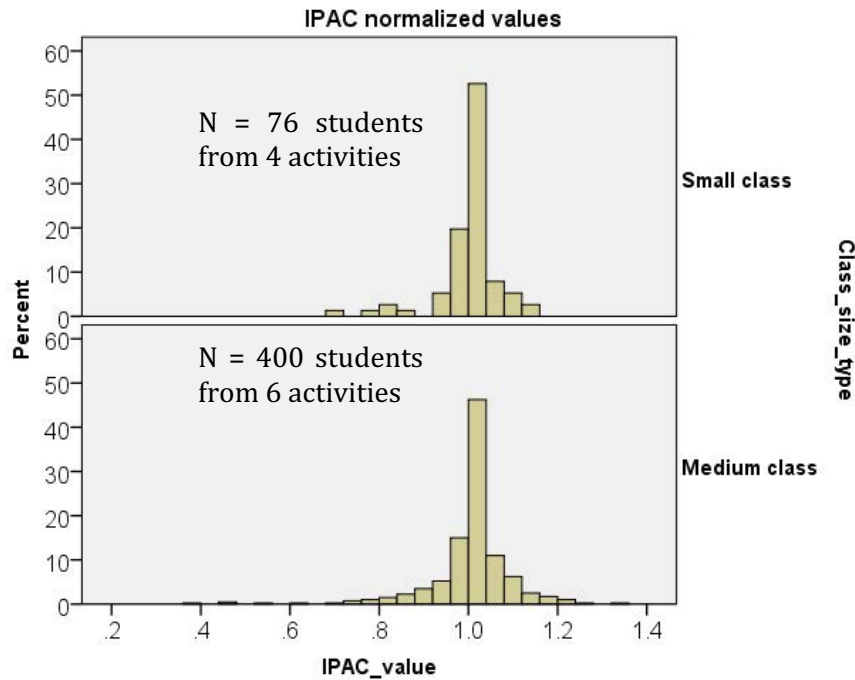


Figure 3. Distribution of IPAC values in normalized form observed in activities run in (a) small classes <40 students, and (b) medium classes 40-100 students.

Table 2. Percentiles corresponding to the IPAC values obtained per student.

Percentile	IPAC value in normalized form	IPAC value in percentage form
2	0.74	56.5
5	0.84	78.2
10	0.92	85.1
20	0.98	92.4
50	1.00	97.9
80	1.05	100
90	1.09	100
95	1.14	100
98	1.19	100

3.2 Effect of the IPAC values onto individual marks

The effect of IPAC values on final individual student marks can be seen in table 3 for three different case scenarios, these are when the baseline group mark is 40%, 60% and 80%. Differentiation was made between the two main trends when applying the IPAC value, i.e. IPAC value as an *added percentage* or as a *normalized multiplying factor*. The percentiles correspond to the IPAC values found in the case studies of this paper, reported in Table 2.

Table 3. Spectrum of the effect of combining IPAC values with group marks.

Percentile	Final individual mark					
	IPAC value * group mark			0.3*IPAC + 0.7*group mark		
	Baseline group mark			Baseline group mark		
	40%	60%	80%	40%	60%	80%
2	29.6	44.4	59.2	44.95	58.95	72.95
5	33.6	50.4	67.2	51.46	65.46	79.46
10	36.8	55.2	73.6	53.53	67.53	81.53
20	39.2	58.8	78.4	55.72	69.72	83.72
50	40	60	80	57.37	71.37	85.37
80	42	63	84	58	72	86
90	43.6	65.4	87.2	58	72	86
95	45.6	68.4	91.2	58	72	86
98	47.6	71.4	95.2	58	72	86

3.3 Comparison with the literature

Some relevant papers found in the literature are listed in Table 4, along with the relevant data extracted describing the IPAC value distributions that they reported. The rows labelled as *Garcia-Souto (2019)* refer to the results in the present study. Studies were clustered according to the type of IPAC value used, namely *percentage* or *normalized*.

Table 4. IPAC values distributions: comparison with other studies.

Study	Student sample size	Type of IPAC value	IPAC value			
			Average	SD	Min	Max
Garcia-Souto (2019)	476	normalized	1	0.092	0.36	1.32
Cheng (2000)	53	normalized	1	0.044	0.857	1.1
Conway (1993)	-	normalized	1	0.094	0.76	1.19
Kilic (2006)*	95	normalized	1	0.15	0.68	1.34
	95	normalized and scaled down***	1	0.07	0.84	1.17
	95	normalized	1	0.12	0.63	1.32
Garcia-Souto (2019)	710	percentage	94.52	10.20	25	100
Das (1998)	64	percentage	79.67	17.33		
Falchikov (1997)*,**	38 F	percentage	76.5	15		
	29 M	percentage	76.1	13.5		
	32 F	percentage	67.35	12.5		
	38 M	percentage	63.95	13.5		
	68 F	percentage	76.33	14.0		
	101 M	percentage	75.27	16.0		
Northrup (2006)	43	percentage	89.4	8.0	67.5	99.5
Tucker (2014)**	603 F	percentage	74.96	10.48		
	902 M	percentage	73.42	12.57		

* Data provided for different case studies and **separated into (F) female and (M) male students.

*** Kiric (2006) reported results for a case study before and after the IPAC values were scaled down, i.e. bringing the values closer to 1 with a constant factor of 0.5. Both are included in this table.

4 DISCUSSION

4.1 Typical IPAC values

4.1.1 IPAC values computed and used as an added percentage

Typical IPAC values are quite high when students assess each other under the understanding that this IPAC value will be used as an added percentage. We observe a one-tail distribution with a median of 97.9%, hence almost half the students getting the maximum mark possible (100%). The 5 percentile is still very healthy (78.2%) with only 2% of the students scoring below 56%. Compare with other studies in the literature, these values are quite high. The present study has an average of 94.5% and SD of 10.2%, while the averages reported in the literature (Table 4) range between 64 and 80% with SD mainly around 12-16%. Still, the average per student across the reported studies from literature is 74.5% which places most of the students in the top grade.

We believe that the reason for these high marks is that this way of implementing the IPAC assessment gives students an individual mark that is independent of that of their peers. As such, students are more “forgiving” and “generous” with their peers as it carries no consequence to them, and some groups might even strategically decide to score everyone high. Despite that, the practice still identifies students that underperformed more significantly, provides individual marks for each team member and therefore is appreciated by the students.

4.1.2 IPAC values computed and used as a normalized multiplying factor

When students assess each other under the understanding that the IPAC value will be used as a normalized multiplying factor, the IPAC values seem to be more meaningful. The IPAC value distribution is fairly symmetric around 1, which is the average mark per group, established by the way the normalization of the IPAC value takes place. This practice identifies both cases, students who underperformed and overperformed in the group with values below and over 1 respectively. In addition, now the IPAC value of one individual partially depends on the IPAC values received by the others, as values are given in reference to the average group mark. We believe that this makes students to mark more honestly. There is a good spread of marks as shown in this study with a SD of 0.09, but with 80% of the students achieving a “safe” IPAC value between 0.92 and 1.09. Even the outliers have sensible IPAC values, with the IPAC value for the 2 percentile and 98 percentile being 0.74 and 1.19 respectively. Studies found in the literature show similar results, with a similar SD (ranging between 0.07 and 0.12) and less extreme outliers. This is very positive, and might indicate that the IPAC value in normalized form is less influenced by external factors.

For the data in this study, the class size was shown to have no significant effect on the IPAC values distribution, against the general believe that peer marking might be less accurate in small classes because the students closely know each other. However, the results seem to agree with the students claims that they will assess peers honestly even if friendship is involved [5].

4.1.3 Is it safe to use the IPAC values to calculate the individual marks?

We believe it is safe and beneficial to use the IPAC values to calculate the individual marks, using either of the approaches, i.e. *added percentage* or *normalized factor*. In both cases the mere fact of informing students that the IPAC methodology is to be used is an effective way of encouraging students to engage and behave more professionally during the group work, improving group dynamics and efficiency. It also manages to provide individual marks that are fairer, and easier to defend in exam boards and to professional recognition bodies.

In the case of using the IPAC value as an added percentage, mathematically the final individual marks will always be within the 0-100% range, which gives a sense of reassurance to the academics. It is true that students will be more generous when marking in this way, and that therefore the effect on the final mark is more limited. For instance, when combining the marks as $0.3*IPAC + 0.7*group\ mark$, the final individual mark of all students in a group will be statistically within a 2.3% mark range in 80% of the cases (see Table 3). Also, students will only receive less than the group mark in 2% of the cases when the group mark is ~60% and above, meaning that in most cases the IPAC value will benefit the students by raising their mark - this is fine as far as the tutor is happy. Despite this small differentiating effect, tutors still gain a better distribution of marks within the class.

In case of using the IPAC value as a normalized multiplying factor, the individual marks have been proven to remain within the standard 0-100% range even when they can mathematically go beyond it.

We can see from the examples in Table 3 that the IPAC effect is sensible and within control even before tutor moderations take place, with 80% of the students having final individual marks within a 6.8%, 10.2% and 13.6% marks when their group scores 40%, 60% or 80% respectively. The effect of the IPAC value can be positive or negative respect to the group mark. Therefore, this method rewards contribution and penalizes lack of engagement, moving students across bands such that marks are fairer and less dependent of the group to which an individual was assigned (e.g. a good student in a “bad performing” group and vice versa). The average mark of the students in the class remains unchanged.

Finally, the application of the IPAC value can move students across qualification bands in either of the IPAC forms, and it is a fairer representation of the individual student performance. Values given in this study correspond to unmoderated IPAC scores. However, there are several moderation approaches that tutors can use to smooth or limit the effect of the IPAC values onto the individual marks; this is part of the practice customization and it gives the reassurance to staff that IPAC is safe to use.

5 CONCLUSIONS

This study has proven that the use of the IPAC methodology is safe and actually recommended. It allows academics to give a fairer mark to the individual students that better reflect their performance during the group work activity, mitigating to an extend the choice of putting him/her in one group or another. Depending on whether the IPAC value is used as an added percentage or a normalized multiplying factor, the students will be more lenient or honest respectively when scoring their peers. However, in any of the cases, it works as “the carrot in front of the horse”, i.e. encouraging engagement and professional behaviour among students, which is an excellent by-product. The final individual marks are sensible and with good spread, though this can be controlled by tutor moderation at will. IPAC values in normalized form do not appear to be dependent on the class size. Typical IPAC values were given from a current case study and from literature that support these claims.

The implementation of the IPAC methodology can be very time consuming in practice if appropriate tools are not used, and it increases exponentially with the class size. Practitioners are advised to use a dedicated software that can save them from spending time on the administrative side, and gives them the chance to have more time to improve the activity and the student experience. The IPAC software has been developed at UCL that allows academics to run this practice time-efficiently, and generates a range of feedback for students. Anyone interested on knowing more or potentially using it can contact the corresponding author.

Further work: There might be factors affecting the distribution of IPAC values and that were not considered in this study, and that might account for some of the variability. Some of these factors might be discipline, previous student experience with the IPAC methodology, length of the group work, gender, cultural background, year of study, number of credits associated with the activity, etc. A larger and more complete database needs to be collected in order to do relevant statistical analysis to test the effect of such parameters.

ACKNOWLEDGEMENTS

The author would like to give great thanks to the IPAC Consortium members and IPAC users that kindly provided data for the analysis presented in this paper, those were Dr. Cloda Jenkins, Dr. Tristan Robinson, Dr. Yuhong Zhou, Dr. Thomas Kador, Dr. Dean Barratt and Ms Kate Roach. Our thanks also to the UCL Vice-Provost office (Education & Student Affairs), UCL ELDG, UCL CC Collab, UCL Medical Physics and Biomedical Engineering Department, UCL Chemical Engineering Department, and UCL Change Makers for their financial support to the IPAC Consortium; this has been key to gain a good understanding of the IPAC methodology. Thanks to UCL Knowledge and Innovation Fund for supporting the further development and dissemination of the developed software and IPAC practice. Thanks to my department, UCL Medical Physics and Biomedical Engineering, Engineering Faculty, for their support in conducting this study.

REFERENCES

- [1] D.Q. Nguyen, “The essential skills and attributes of an engineer: A comparative study of academics, industry personnel and engineering students,” *Global Journal of Engineering Education*, vol. 2, no. 1, pp. 65-76, 1998.

- [2] R. Conway, D. Kember, A. Sivan, and M. Wu, "Peer assessment of an individual's contribution to a group project," *Assessment & Evaluation in higher Education*, vol. 18, no. 1, pp. 45-54, 1993.
- [3] M.D.P Garcia Souto, T. Kane, G. Hughes, S. Searles-Bryant, A. Gibson, "Moderated peer assessment of individual contribution to group work," *UCL Teaching and Learning Conference*, London, UK, 19 April 2016.
- [4] M.D.P. Garcia Souto, C.N. Striolo, M. Vogel, R. Grammenos, M. Whyndham, G. Hughes, T. Kador, J. Britton, E. Mambetisaeva, M. Richardson, L. Albelda-Gimeno, I. Chester, F. Akinmolayan, M. Pollock, E. Bele, D. Wright, and A. Gibson, "Peer assessing individual contributions in a group project." *50th Society for Research into Higher Education (SRHE) Conference*, Celtic Manor, Newport, Wales, UK, 6-8 Dec 2017.
- [5] R. Grammenos, M.d.P. Garcia-Souto, I. Chester, L. Albelda-Gimeno, "Peer assessment of individual contribution in group work: a student perspective", *European Society for Engineering Education Conference (SEFI)*, Budapest, Hungary, 16-19 Sept 2019.
- [6] D. Kelley, "Peer Evaluation within a Team Design Project," *Journal of Engineering Technology*, vol. 32, no. 1, pp. 44-50, 2015.
- [7] M.D.P. Garcia Souto, Y. Azma, R. Grammenos, T. Kador, C. Striolo, M. Whyndham, M. Vogel, M. Richardson, A. Gibson, J. Britton, T. Robinson, G. Hughes, "Individual peer assessment of contribution to group work (IPAC): Key points and recommendations", *European Society for Engineering Education Conference (SEFI)*, Budapest, Hungary, 16-19 Sept 2019.
- [8] M.D.P. Garcia-Souto, "Making assessment of group work fairer and more insightful for students and time-efficient for staff with the new IPAC software", *13th International Technology, Education and Development Conference (INTED)*, Valencia, Spain, pp. 8636-8641, 2019. DOI: 10.21125/inted.2019.2154
- [9] W. Cheng, M. Warren, "Making a difference: Using peers to assess individual students' contributions to a group project," *Teaching in Higher Education*, vol. 5, no. 2, pp. 243-255, 2000.
- [10] G.B. Kilic, and M. Cakan, "The analysis of the impact of individual weighting factors on individual scores," *Assignment & Evaluation in Higher Education*, vol. 31, no. 6, pp.639-654, 2006.
- [11] M. Das, "Self and Tutor Evaluations in Problem Based Learning Tutorials: Is There a Relationship?" *Medical Education* vol. 32, no. 4, pp. 411-418, 1998.
- [12] N. Falchikov, and D. Magin, "Detecting Gender Bias in Peer Marking of Students' Group Process Work." *Assessment & Evaluation in Higher Education*, vol. 22, no. 4, pp. 385-396, 1997.
- [13] S.G. Northrup, D.A. Northrup, "Multidisciplinary Teamwork Assessment: Individual Contributions and Interdisciplinary Interaction," *Frontiers in Education 36th Annual Conference Proceedings*, San Diego, CA, USA, 27-31 Oct 2006.
- [14] R. Tucker, "Sex does not matter: gender bias and gender differences in peer assessments of contributions to group work," *Assessment & Evaluation in Higher Education*, vol. 39, no. 3, pp. 293-309, 2014.