Insights and Lessons Learned from Analyzing ICSE 2016 Survey and Review Data

Version as of 16 May 2016

Willem Visser, Laurie Williams

1 Introduction

ICSE 2016 used a Program Board model, as first used by ICSE 2014. We believe this model allows us to scale to the large number of submissions ICSE now receives, without having a face-to-face meeting with an unmanageable number of people. The Program Board (PB) consisted of 28 people and the Program Committee (PC) consisted of 89 people. Each paper received 3 reviews by PC members after which the authors were given an opportunity to respond to the reviews. Authors were also given an option to rate the reviews they received at the rebuttal time. A 4 week online discussion period then followed that was overseen by the PB members. At the end of the this period, papers with no support were marked as rejected. The rest of the papers then received an extra review from a PB member. Again, an online discussion followed, whereafter some additional papers with no support were marked for rejection. Lastly, 100 papers were discussed in a face-to-face meeting that included all the PB members. Before the meeting, 36 papers were marked for acceptance without discussion (a.k.a. were not part of the 100 papers discussed) and during the meeting a further 65 papers were accepted to give a total of 101 accepted papers. Two months after the results were announced, surveys were conducted of the authors, the PC and the PB. This document reports on the results of these surveys, the author review rankings and some general stats during the reviewing process.

This is the third year in which the ICSE process data is analyzed and reported in this fashion. We believe this helps to improve the ICSE reviewing process. For example, the survey data from 2014 indicated that authors of papers that only received 2 reviews were more unhappy with the reviews than authors that received 3 reviews. For 2016, we thus decided all papers will get at least 3 reviews. Similarly, the need for authors to declare paper categories upon submission, as was introduced in 2014, was shown in the 2015 report not to make any meaningful difference, hence they were removed, thereby simplifying the submission process.

In 2016, we also made the complete process that was going to be followed public by releasing the same slides presenting the process (as the ones presented to the PB and PC). In addition, we also organized a Google Hangout session where we presented the

slides and encouraged people to ask questions. For more details ee https://plus.google.com/communities/110196757571521827561.

2 Some historical data

Year	Submitted	Accepted	%
2009	405	50	12
2010	380	52	14
2011	441	62	14
2012	408	87	21
2013	461	85	18
2014	495	99	20
2015	452	82	18
2016	530	101	19

As can be seen from the table above, there has been a growing trend in the number of submissions for ICSE. In fact, ICSE 2016 received a record number of submissions. The acceptance rate is however in line with the last 5 years, i.e. ICSE is receiving more submitted papers, but the quality of the papers seem to be roughly stable.

3 Basic Process data

In this section, the process flow will be described from the papers entering the reviewing process until they exit either as rejected or accepted. The process followed the following schedule with online discussions occurring any time after reviews began being submitted but especially beginning once all PC reviews were due and continuing through the time of the PB meeting:

- Paper submissions: 28 August

- PC/PB Assignments: 11 September
- PC Reviews Due: 25 October
- Author clarification period: 1-5 November
- Second PB member assigned: 20 November
- Second PB member review due: 5 December
- PB meeting: 9-10 December

As stated, 530 papers were submitted through the CyberChairPro system.

Desk Rejections. A desk rejects phase was conducted by the chairs and lasted about 4 days. Authors of 17 papers were notified that their paper was rejected, mostly for formatting reasons or being outside the scope of ICSE.

PC Reviewing and Online Discussion Each PC member got either 18 or 19 papers to review. Note that all papers got 3 reviews. Each paper was also assigned one PB member (each PB member oversaw roughly 19 papers). During the reviewing period, the PB members monitored the review and the online discussions. The first week of online discussion was used to refine the questions requiring an author response. Authors were then given 5 days to respond. Online discussions were lively, and the goal was to reach one of two outcomes: (1) reject; or (2) move on to the next phase with or without a consensus amongst the reviewers. During this phase, 308 papers were rejected and 193 papers were moved on. Additionally, 12 papers were withdrawn after the author response period, presumably because the authors could see that the paper would inevitably be rejected.

PB review and Online Discussion All 193 papers that moved on to the next phase got an additional review by a PB member. The rebuttal phase and an extensive online discussion post rebuttal were conducted prior to the second PB reviewer being assigned to reduce the number of papers that got a second PB review. Ultimately, each PB member got either 6 or 7 papers to review. The intent of the PB review was to ensure that 2 people from the PB would have good knowledge of a paper during the face-to-face meeting to allow a proper discussion. A short online discussion followed the PB review, involving all reviewers, including the PC members. Each paper was classified in one of three categories: (1) accept without discussion at the PB meeting; (2) reject; or (3) move-on to the next phase. Another 57 papers were rejected, leaving 136 to move on. When a paper moved on to the discussion phase, it was suggested but not required to see if there is consensus that it be accepted. Based upon this discussion, the chairs earmarked 36 papers that had a consensus to accept as clear accepts without discussion. The PC/PB was notified of this classification for a paper and were given the opportunity to call for an in-person discussion. If no such consensus could be reached on a paper, the decision would be reached during the meeting.

PB Meeting The remaining 100 papers were discussed, and 65 were accepted and 35 rejected during the meeting. Each paper, accepted and rejected, also received a summary of why the decision was reached which was sent with the notifications.

An important metric for the PB model is how often does the PB decision overturn the decision made by the PC. A criticism of the PB model is that it lessens the role of the PC and it is important to guard against any such perceptions. During the PB meeting, 5 papers, which had 3 positive PC scores (A or B), but a negative score by the PB member that reviewed the paper, ended up being rejected during the meeting. For 3 of these 5 cases, the criticism by the PB member was acknowledged by the PC members during the online discussion preceding the meeting (and was therefore in the no consensus category) and as such the rejection was not a surprise. However, 2 out of the 5 decisions was clearly overturned even though there was strong support from the PC. Notably, in 2014 the PB also overturned 2 accepts to rejects during the meeting, and we believe this number is small enough not to raise any major concern.

Lastly, the meeting acceptance rate of 65% is substantially higher than the last two years: 41% in 2014 and 46% in 2015.

3.1 Scoring

Below is the list of papers in the different scoring brackets, where the brackets are defined by the two letter grades, i.e. A-B means papers only got A scores or A and B scores, and A-D means the scores ranged from a high of A to a low of D (i.e. at least one A and one D).

Score Range	Accept	Reject
A-B	26	0
A-C	29	17
A-D	8	14
B-B	8	0
B-C	27	82
B-D	3	59
C-D	0	227

The figure below shows the percentages of scores for the different categories over the last 3 years and it can be observed that there is not much difference. In 2016, there were slightly less papers with only A and B scores (5% versus 7% and 8%, respectively, for 2014 and 2015).
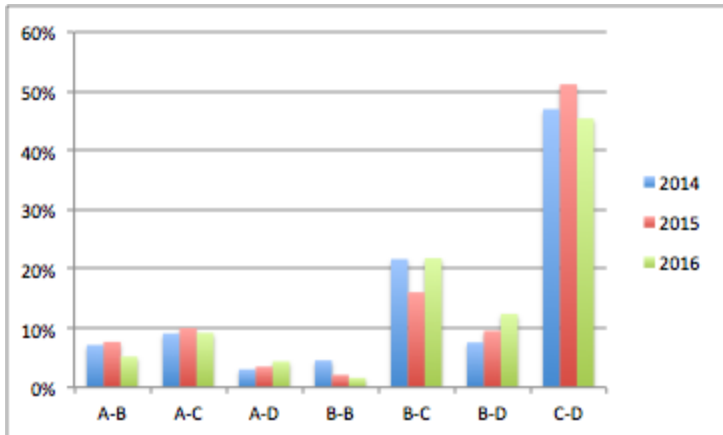
As in the past two years, the figure below also shows that the top (A-B) and bottom (C-D) brackets that define papers with only positive and only negative scores respectively are good indicators for being accepted and rejected. In 2014, there was 1 exception in the top bracket (that got rejected) whereas in 2016 there were three papers in the B-D category that were accepted (whereas there were none in 2015 and only 1 in 2014).
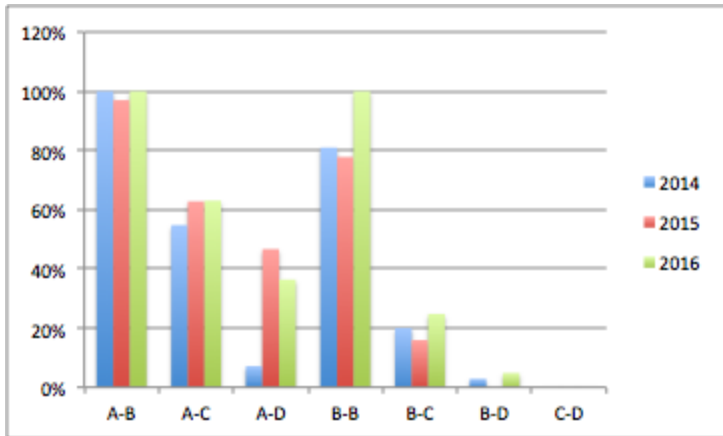
3.2 Expertise

The expertise as declared by the reviewers are shown below, note this includes both PC and PB reviews.

Expertise Declared	Number of Papers
XXXX	10
XXX	31
XXXY	30
XXXZ	4
XXY	103
XXYY	36
XXYYY	5
XXYYZ	1
XXYZ	13
XXYZZ	2
XXZ	9
XYY	83
XYYY	37
XYYY	1
XYYYZ	2
XYYZ	18
XYYZZ	1
XYZ	23
XYZZ	4
XZZZ	1
YYY	34

YYYY	22
YYYZ	13
YYZ	10
YYZZ	2
YZZ	4

In line with the reports from the previous two years, we also show comparative figures with respect to papers with at least one X, papers that have a minimum of 2 Ys, and papers with at least 1 Z expertise, across the last 4 years.

	2013	2014	2015	2016
At least 1 X	72%	77%	84%	83%
Minimum of 2Ys	95.5%	96%	97.7%	99.2%
Papers with at least 1 Z	23%	18%	7.2%	20.6%

The numbers for 2016 seems quite reasonable in comparison to the other years. With almost all papers being reviewed receiving a minimum of 2Ys. The PC+PB size might have played a positive role here as well as the fact that on average there were more reviews per paper in 2016. 2015 seems to be quite an outlier for papers with a Z on it.

Note that 1758 reviews were submitted for the 513 papers, i.e. on average 3.4 reviews per paper (compared with 2.6 in 2015).

3.3 Topics

The table below shows the topic options given to the authors during the submission process, with the number of papers submitted that picked each topic, how many that was from the total (513), how many was accepted, and lastly, the percentage accepted.

Topic	Submitted	% Subm	Accepted	%Accep
Agile software development	13	3%	1	8%
Autonomic and (self-) adaptive systems	20	4%	0	0%
Cloud computing	**14**	**3%**	**3**	**21%**

Component-based software engineering	12	2%	1	8%
Configuration management and deployment	15	3%	1	7%
Cooperative, distributed, and collaborative software eng.	**29**	**6%**	**8**	**28%**
Cyber physical systems	6	1%	1	17%
Debugging, fault localization, and repair	**62**	**12%**	**14**	**23%**
Dependability, safety, and reliability	27	5%	5	19%
Embedded software	4	1%	2	50%
Empirical software engineering	**121**	**24%**	**32**	**26%**
End-user software engineering	6	1%	1	17%
Formal methods	41	8%	7	17%
Green and sustainable technologies	3	1%	2	67%
Human factors and social aspects of software engineering	**42**	**8%**	**11**	**26%**
Human-computer interaction	18	4%	2	11%
Middleware, frameworks, and APIs	10	2%	2	20%
Mining software engineering repositories	**86**	**17%**	**21**	**24%**
Mobile applications	**35**	**7%**	**10**	**29%**
Model-driven engineering	18	4%	0	0%
Parallel, distributed, and concurrent systems	19	4%	2	11%
Performance	**34**	**7%**	**9**	**26%**
Probabilistic systems	4	1%	1	25%
Program analysis	**85**	**17%**	**24**	**28%**
Program comprehension	**33**	**6%**	**10**	**30%**
Program synthesis	**15**	**3%**	**5**	**33%**
Programming languages	**27**	**5%**	**10**	**37%**
Recommendation systems	26	5%	4	15%
Refactoring	13	3%	2	15%
Requirements engineering	23	4%	4	17%

Reverse engineering	18	4%	2	11%
Search-based software engineering	**25**	**5%**	**6**	**24%**
Security, privacy and trust	**38**	**7%**	**9**	**24%**
Software architecture	**24**	**5%**	**5**	**21%**
Software economics and metrics	12	2%	4	33%
Software evolution and maintenance	98	19%	18	18%
Software modeling and design	24	5%	3	13%
Software process	20	4%	3	15%
Software product lines	**19**	**4%**	**4**	**21%**
Software reuse	28	5%	4	14%
Software services	8	2%	1	13%
Software testing	**84**	**16%**	**20**	**24%**
Software visualization	13	3%	2	15%
Specification and modeling languages	16	3%	0	0%
Tools and environments	50	10%	9	18%
Traceability	9	2%	1	11%
Ubiquitous/pervasive software systems	2	0%	1	50%
Validation and verification	53	10%	10	19%

There were 3 topics that had no papers accepted: Autonomic and (self-) adaptive systems; Model-driven engineering and Specification; and modeling languages. Note that each of these topics had a fair number of submissions (4%, 4%, and 3% respectively). There are 17 topics that received at least 3% of the submissions and got an acceptance rate higher than the overall average of 19%.; these are indicated in bold in the table.

Below we compare the percentage submitted and accepted with those of 2015. There were some topic name changes between the years, so we only show the ones where a direct comparison is possible.

Topic	2016 % Sub	2016 % Acc	2015 % Sub	2015 % Acc
Agile software development	3%	8%	4%	19%
Autonomic and (self-) adaptive systems	4%	0%	5%	13%
Component-based software engineering	2%	8%	4%	0%
Configuration management and deployment	3%	7%	3%	14%
Cooperative, distributed, and collaborative software engineering	6%	28%	5%	30%
Debugging, fault localization, and repair	12%	23%	11%	25%
Dependability, safety, and reliability	5%	19%	5%	24%
End-user software engineering	1%	17%	4%	6%
Green and sustainable technologies	1%	67%	1%	25%
Human factors and social aspects of software engineering	8%	26%	10%	25%
Human-computer interaction	4%	11%	5%	20%
Middleware, frameworks, and APIs	2%	20%	4%	6%
Mining software engineering repositories	17%	24%	17%	15%
Mobile applications	7%	29%	9%	16%
Model-driven engineering	4%	0%	5%	0%
Parallel, distributed, and concurrent systems	4%	11%	5%	18%
Performance	7%	26%	3%	13%
Program analysis	17%	28%	16%	32%
Recommendation systems	5%	15%	4%	6%
Requirements engineering	4%	17%	4%	5%
Reverse engineering	4%	11%	4%	16%
Search-based software engineering	5%	24%	4%	18%
Security, privacy and trust	7%	24%	6%	23%
Software architecture	5%	21%	5%	9%

Software economics and metrics	2%	33%	5%	10%
Software evolution and maintenance	19%	18%	20%	19%
Software modeling and design	5%	13%	6%	8%
Software process	4%	15%	5%	14%
Software product lines	4%	21%	3%	29%
Software reuse	5%	14%	4%	5%
Software testing	16%	24%	15%	31%
Tools and environments	10%	18%	13%	22%
Traceability	2%	11%	3%	18%
Ubiquitous/pervasive software systems	0%	50%	3%	14%

What is rather remarkable is that there is hardly any difference between the submitted percentages; they are mostly within 2%. The only exceptions are end-user software engineering (down 3%), performance (up 3%), tools and environments (down 3%) and Ubiquitous/pervasive software systems (down 3%). Even for the acceptance rates, if we again only consider topics that got at least 3% of the papers then there are only 6 where we have a difference of acceptance rate of more than 10%: Agile software development (down 11%), Autonomic and (self-) adaptive systems (down 13%), Mobile applications (up 12%), Performance (up 13%), Requirements engineering (up 12%) and Software architecture (up 12%). However, what might need some attention is the fact that Model-driven engineering for the 2nd year in a row got zero papers accepted (out of 40 submitted the last 2 years).

This data seems to show that the process is fairly stable with no major fluctuations in research focus or outcomes the last 2 years. I will be interesting to see how this develops over a longer period.

3.4 Authors

Roughly 1500 authors submitted papers to ICSE 2016. The figure below shows the distribution of the number of papers to authors as well as the accept (blue) and reject (red) ratio.
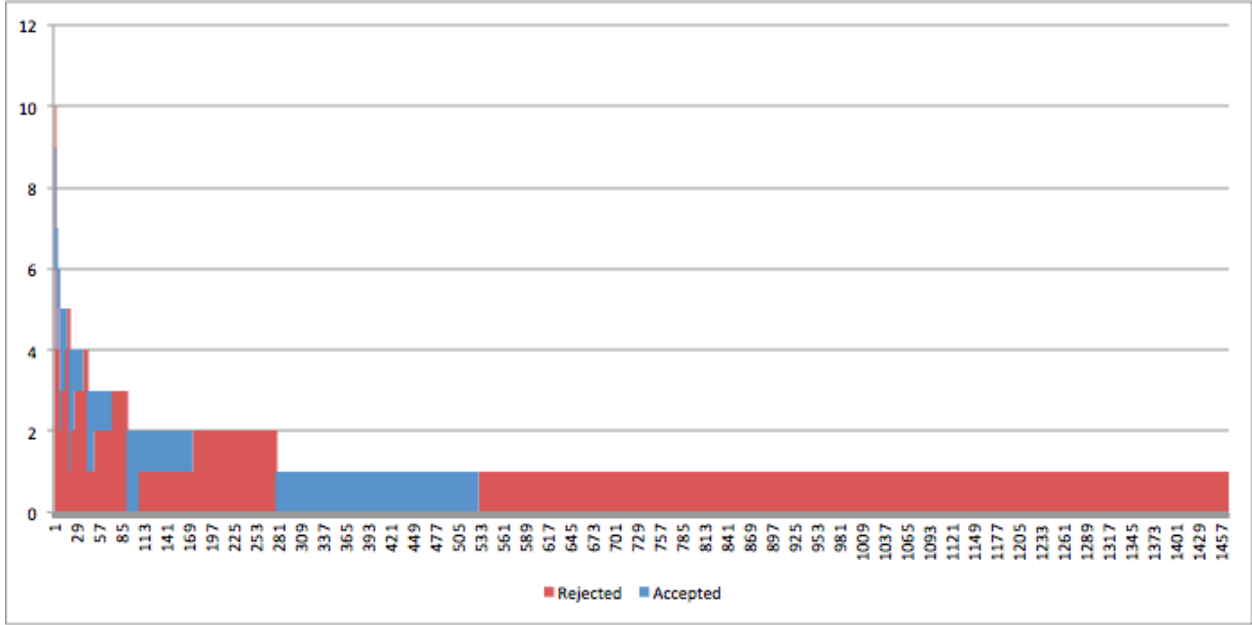
278 authors submitted 2 or more papers (184 in 2015), 91 had 3 or more, 42 had 4 or more and 19 had 5 or more (14 in 2015). See below the figure for 4 or more papers
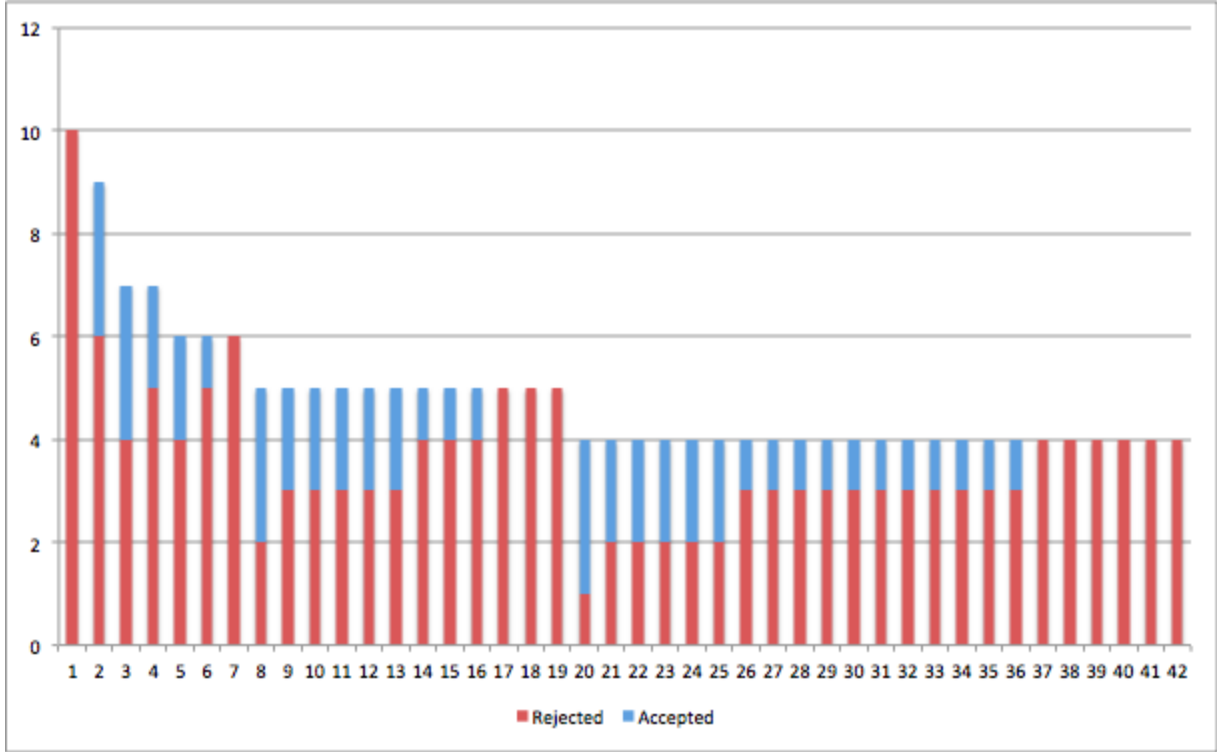
In 2015, it was found that 1% of the authors contributed 19% of the submissions. In 2016, we found 1% of the authors contributed approximately 17% of the submissions. Also, the data that

was used to produce the figure above indicates that the authors with 4 or more papers (42 authors i.e. about 3% of the authors) contributed almost 28% of the papers (143 papers). However, they also contributed 37 accepted papers which is roughly 37% of the total accepted papers and an acceptance rate of 26%.

As is customary at ICSE, PC and PB members can also submit papers, and below one can see the breakdown of submitted and accepted papers for the PC and PB.

	Submitted	Accepted	Acceptance Rate
PB	44	12	27%
PC	109	25	23%

PB members are typically more established members of the community with a track record of papers at previous ICSEs, so it is not unexpected that they have a slightly higher acceptance rate than the PC members and the author base as a whole. Although not directly comparable, in 2015 the PC had an acceptance rate of 29% whereas the Review Committee (RC) had a rate of 22%. However the trend that the reviewing bodies have a slightly higher acceptance rate than the authors in general seems to be holding steady at about the same rate as before.

4. Survey Analysis

As was the case in 2014 and 2015, an anonymous survey was conducted asking the PB, the PC, and the submitting authors for their opinions on a number of topics. We collected data on a Likert scale of 1 (Strongly disagree) to 5 (Strongly agree). Additionally, we asked authors to rate the reviews they received during the author response period. The authors were not provided with the score for the reviews nor author expertises. The goal was to obtain an unbiased as possible view from the authors.

Note that one of the questions on the author survey was whether their paper was accepted or rejected. Since the survey was anonymous, we used this answer to partition the responses between accepted and rejected outcomes. Note further that we asked that only one response (a.k.a one author) be entered per paper, but we had no way of checking that.

We got exceptionally high response rates, especially from the authors. For the PB we got 25 out of 28 responses (comparable to 2014), for the PC we got 59 out of 89 responses (comparable to 2014 and 2015). For the authors, we got 359 out of 513 responses (70%, compared to 40-45% in 2014 and 2015).

We will structure the following discussion around the general theme of the questions and try and compare the responses of the various groups. We will also compare against previous year's but mostly only in the positive categories (4 and 5 in the scale), to make comparisons reasonably

obvious. Where we do refer to negative scores it is 1 and 2 in the scale. Lastly the 2014 and 2015 scale were using a 7-point likert scale and there 1,2 and 3 are negative and 5, 6 and 7 positive.

4.1 Reviewer Expertise

Both the PC and the PB was asked the following question

"I mostly reviewed papers within my area(s) of expertise"
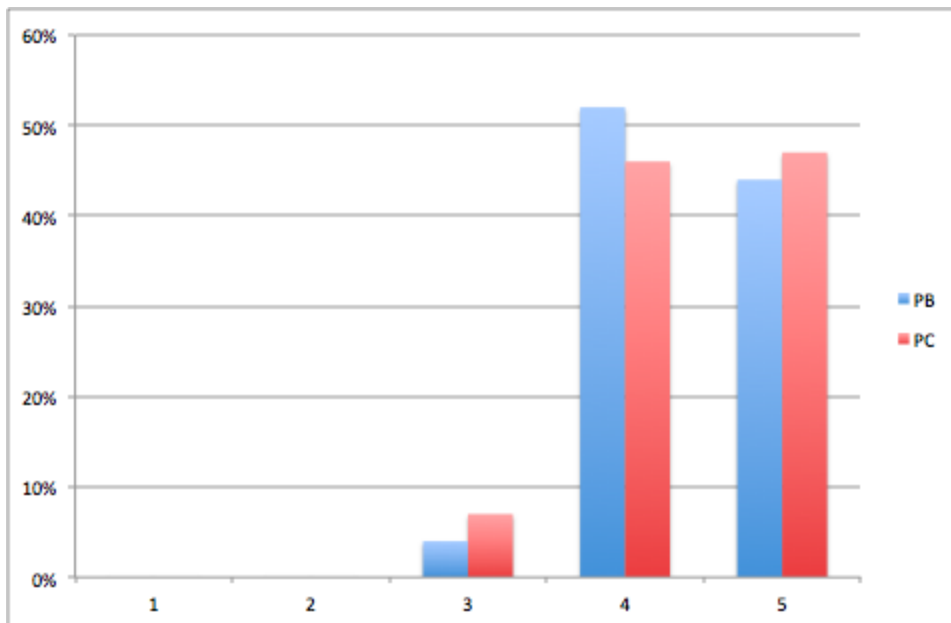
As can be seen from this data the two groups considered the mapping of papers to reviewer's expertise quite good. With regards to previous years, the 2014 PB was 100% positive, compared to 96% in 2016. The PC was 88% (2014), 94% (2015) and 93% positive in 2016. These numbers all seem quite good.

The picture however is slightly different when asking the author's a similar question

"Reviews were performed by reviewers knowledgeable in the area"
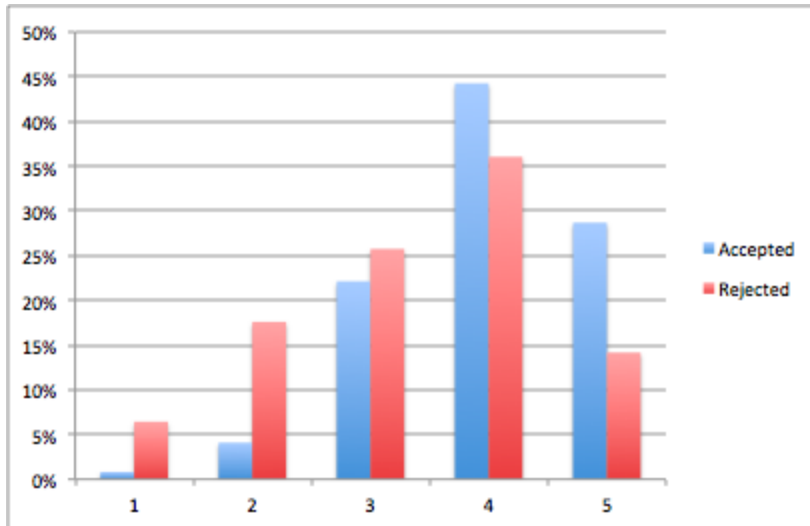
Positives now go down to 71% for authors of accepted papers and 50% for authors of rejected papers. Even if we take the top end of that scale (i.e. only look at responses from authors with accepted papers), there is still a wide gap between what the reviewers thoughts and what the authors thought.

Comparing overall trends (not broken out into accept/reject) with 2014 and 2015 we get: 58% positive (2014), 66% (2015) and 58% (2016). Seems to be close enough to what has gone before.

4.2 Review Quality

As stated above, we had one more data point this year in that we also collected opinions of review quality from the authors BEFORE the result was known, during the rebuttal phase. Specifically, we asked them on a Likert scale whether a review was not useful at all (1), not very useful (2), neutral (3), useful (4) or very useful (5). We got very high response rates with 1399 out of the 1539 reviews being rated (91%). This lead to the following aggregated result.
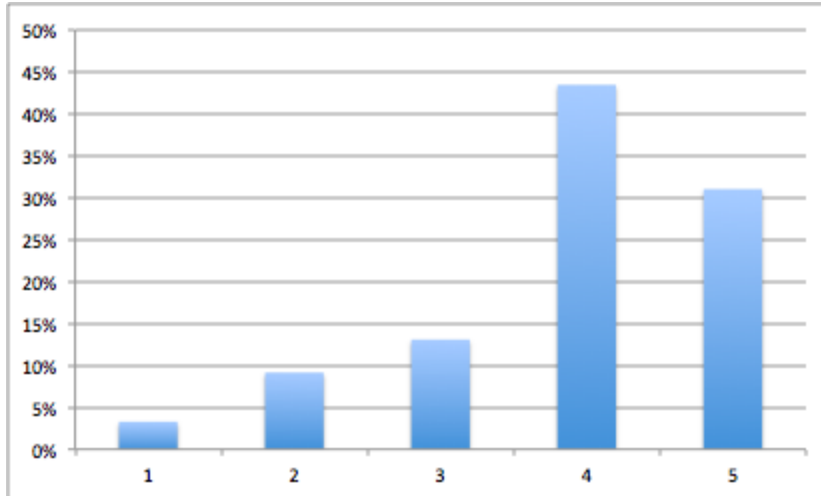
From this it seems the authors were in fact very happy with the reviews they received (74% being positive and only 12% being negative). However comparing this with their views afterwards paints a very different picture. The question that was asked during the survey afterwards was

"Reviews were thorough?"

The before and after question was not identical, but we believe more than close enough to warrant a direct comparison. The overall response, compared to the before responses were:
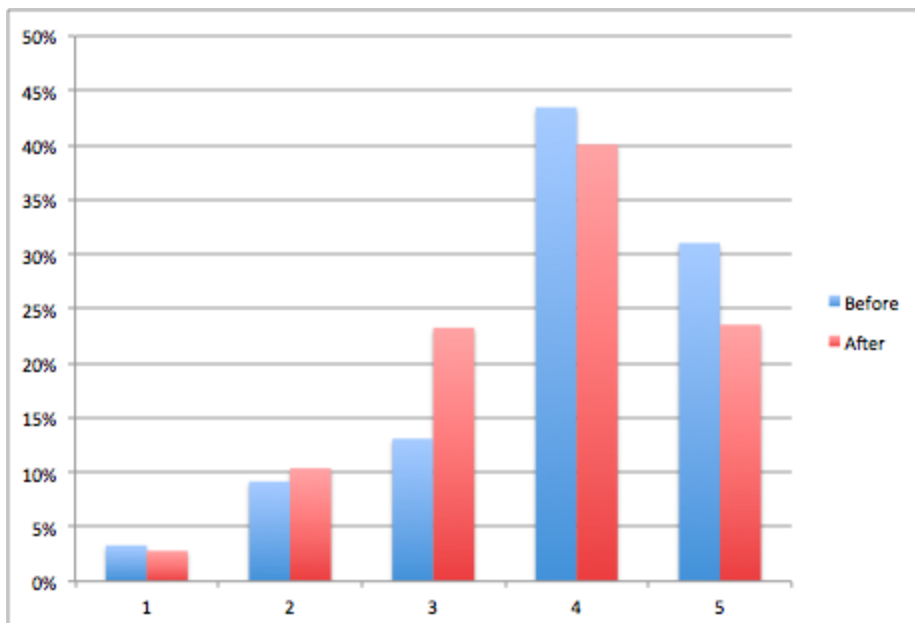
We believe these results look quite similar with slightly more authors shifting to neutral from a positive stance before. The big difference however comes when you consider the breakout of accept versus reject results (below).
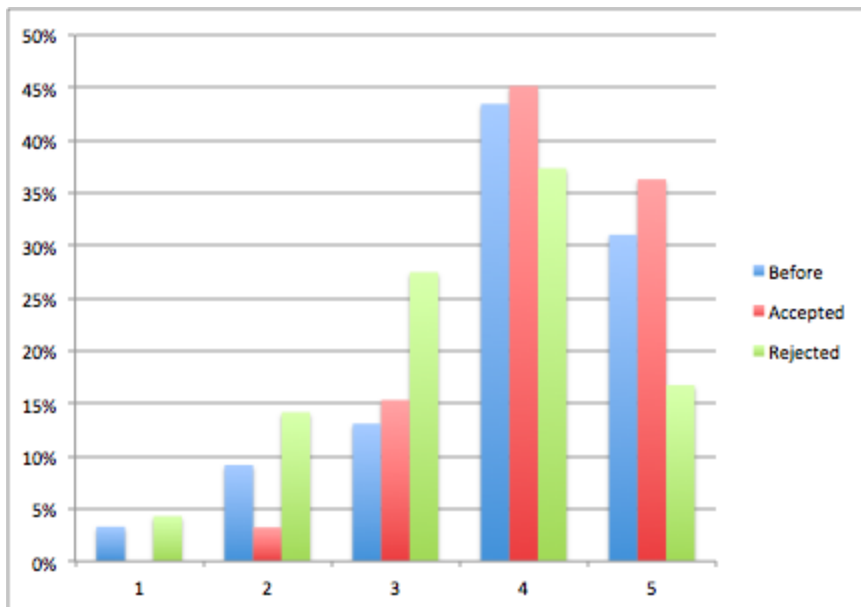
The author's view before and the view of the authors that had their papers accepted track quite closely, but the authors that had their papers rejected became much more negative. It is obvious that knowing the result biases the author view of the quality of the reviews. This is hardly surprising, but now there is clear evidence. Based upon these results, we recommend that this question about review quality should be asked of authors during the process rather than after the accept/reject decisions have been communicated.

If we compare with the positive categories from the previous years we see that results are very similar: 66% (2014), 68% (2015) and 66% (2016).

The following, related, questions were asked to both the PC and the PB

"Reviews by the other program committee members were (mostly) thorough " (PC) and "Reviews were (mostly) thorough" (PB)
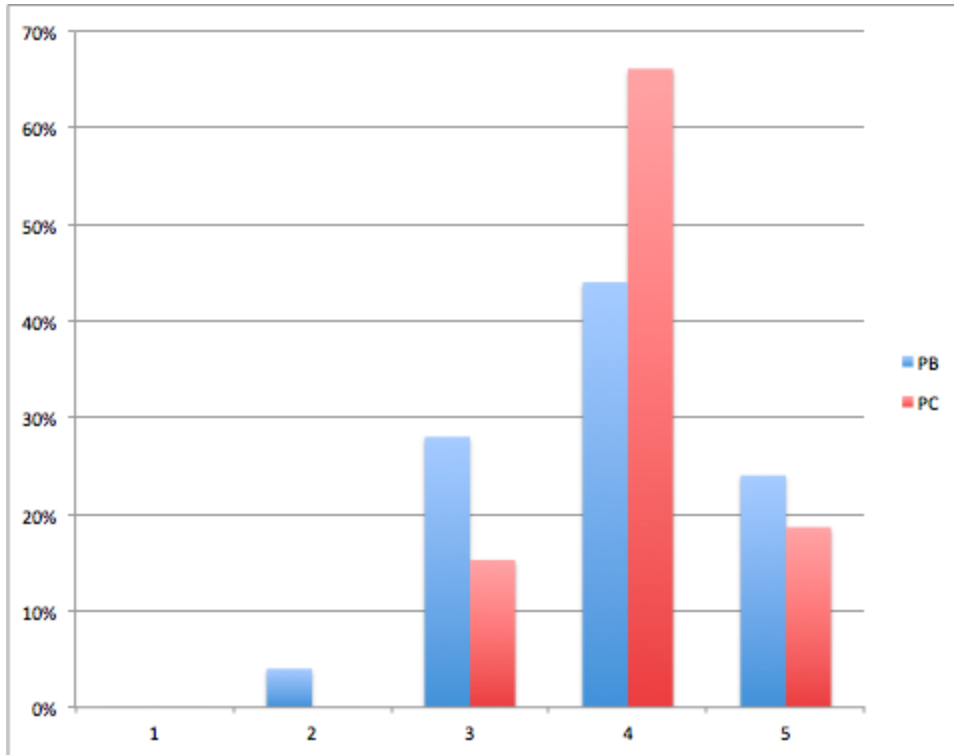
Not too surprisingly, the PC (i.e. the people who did the reviews) felt most strongly positive about the quality of the reviews. The PC's positive view stayed about the same between 2014 (83%), 2015 (88%) and 2016 (85%). The PB in 2014 considered the reviews more thorough (95%) than in 2016 (68%). However, this likely due to the 7 point scale used in 2014 versus the 5 point scale of 2016, since if we consider the PB's negative views, they are about the same 5% in 2014 and 4% in 2016.

The last set of questions related to review quality was a set of questions related to whether the reviews were constructive. The following three versions of this question was asked to the three groups:

"Reviews were constructive" (Authors)
"Reviews by the other program committee members were (mostly) constructive" (PC)
"Reviews were (mostly) constructive." (PB)
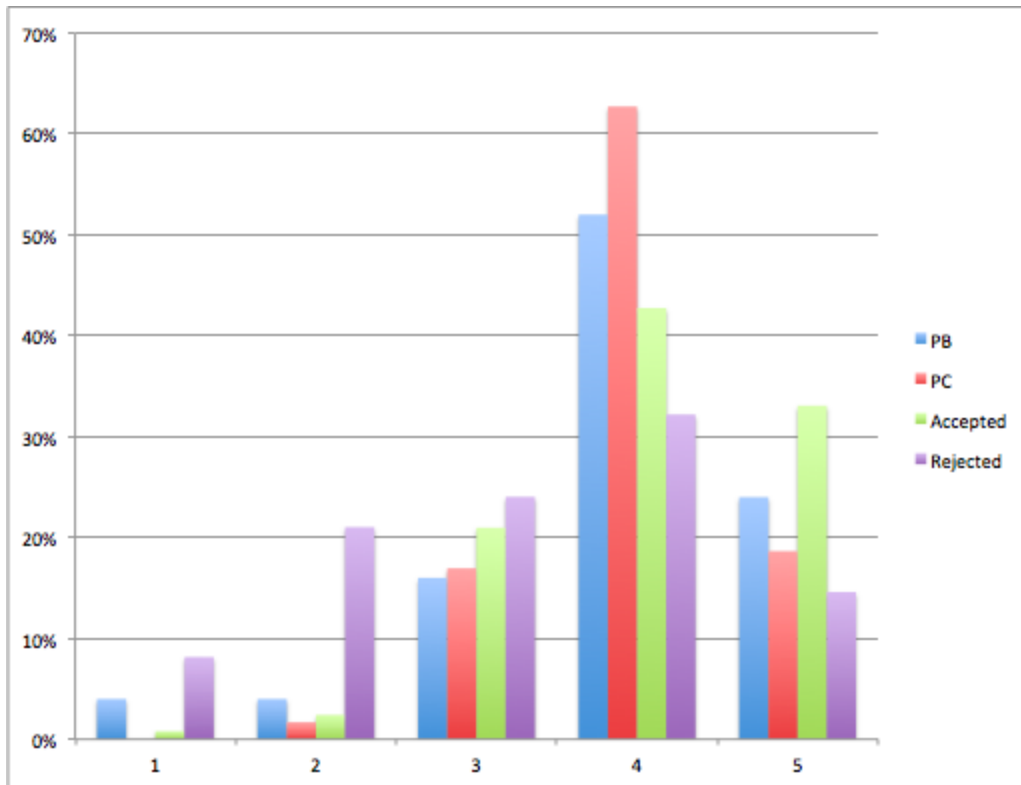
As before, the authors of papers that got rejected did not consider the reviews as constructive as the ones where the papers were accepted. The reviewers (PB and PC) views align quite well with the authors of accepted papers (especially for non-neutral cases). In terms of comparison with 2014 we get: PB positives in 2014 (80%) and 2016 (76%); PC positive 2014 (89%), 2015 (89%) and 2016 (82%). The author views compare as follows: 2014 (64%), 2015 (64%) and 2016 (57%). The slight dip in 2016 is again likely due to the Likert scale of 5 used in 2016, when considering just the negative views we get 2014(20%), 2015 (19%) and 2016 (21%),

4.3 Reviewing Process

A number of the survey questions were focused on the reviewing process and those can again be subdivided into ones that relates directly to the authors and those that deal with the process related to the Program Board model.

4.3.1 Author Process

This year the authors had the chance to write a response to the reviews they received and the following question was asked about this aspect of the process

"I appreciated the opportunity to provide a rebuttal to reviewers"

With the following response
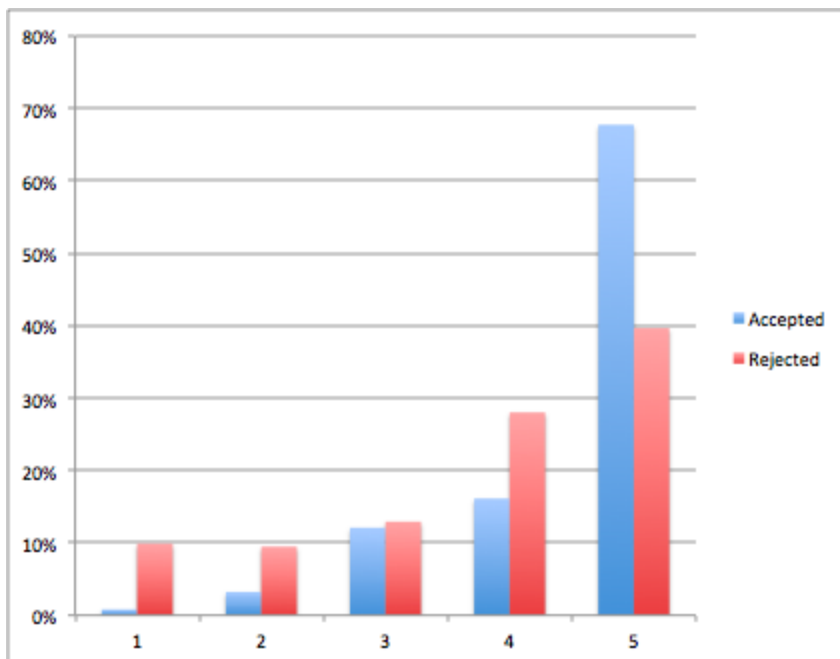
The opportunity to provide a response we believed to be an important addition to the process, and the authors overwhelmingly agreed. Although authors of accepted papers were again more happy with the feature, this time though even the authors of rejected papers liked the feature by

a decent margin: overall 74% positive, accepted 84% positive and rejected 68% positive. As a result, we recommend rebuttals should always be used.

As in previous years, we asked whether the decision summaries provided with the notifications were useful. Specifically we asked:

" The decision summary was helpful "
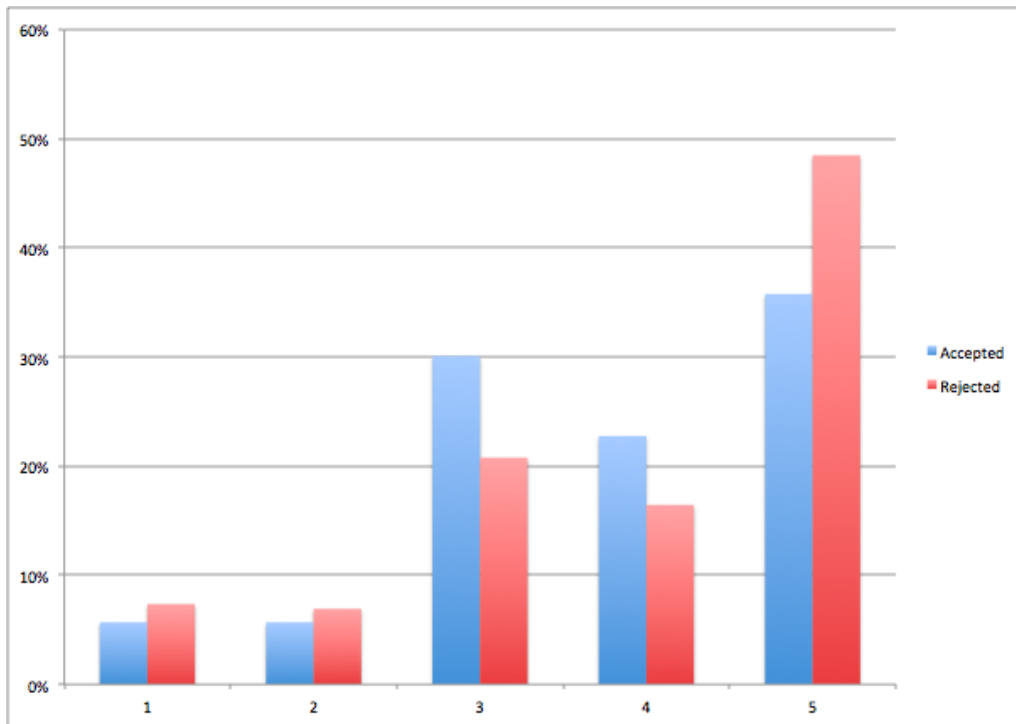
Very interestingly, the rejected paper's authors were slightly more positive than the accepted paper's authors. This is great news, since for the rejected papers the information in the rejection summaries are, of course, much more valuable because they are intended to summarize the discussion and the essence of the rationale behind the decision. In 2014, decision summaries were also provided and interestingly 63% of the authors were positive about it, the exact same ratio as in 2016. However in 2014 the authors of rejected papers were more negative than the ones for accepted papers. It is likely that the fact that all papers got at least 3 reviews contributed here, since one of the takeaways from 2014 was that authors with only 2 reviews (and thus rejected) was more negative than authors with 3 reviews that was also rejected.

4.3.2 Reviewing Process

4.3.2.1 Online Discussions

Both the PC and PB were asked the following question:

"Most online discussions were insightful and useful"
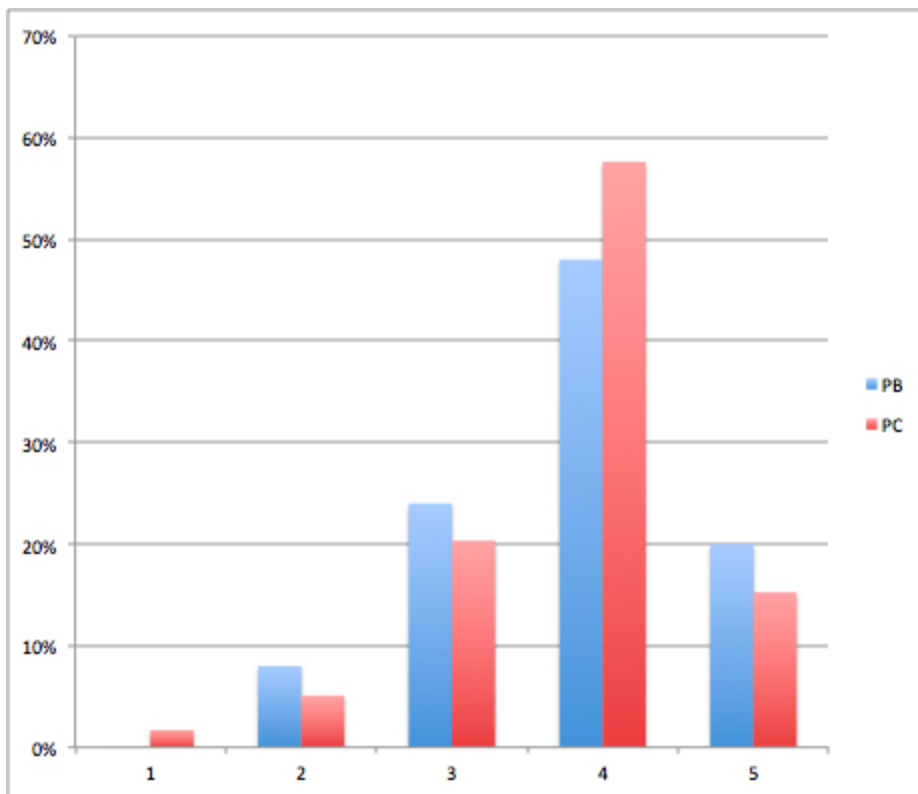
Both the PC and PB was quite positive about the online discussions, with the PC being slightly more positive at 73% versus 68% for the PB. In 2014, the PB was 60% positive and the PC 76% positive. In 2015, the PC was 80% positive (note there was no PB in 2015).

The PC and PB were asked about the length of the online discussions, which was about 4 weeks. Specifically:

**"The length of the online discussion was too short…too long"**
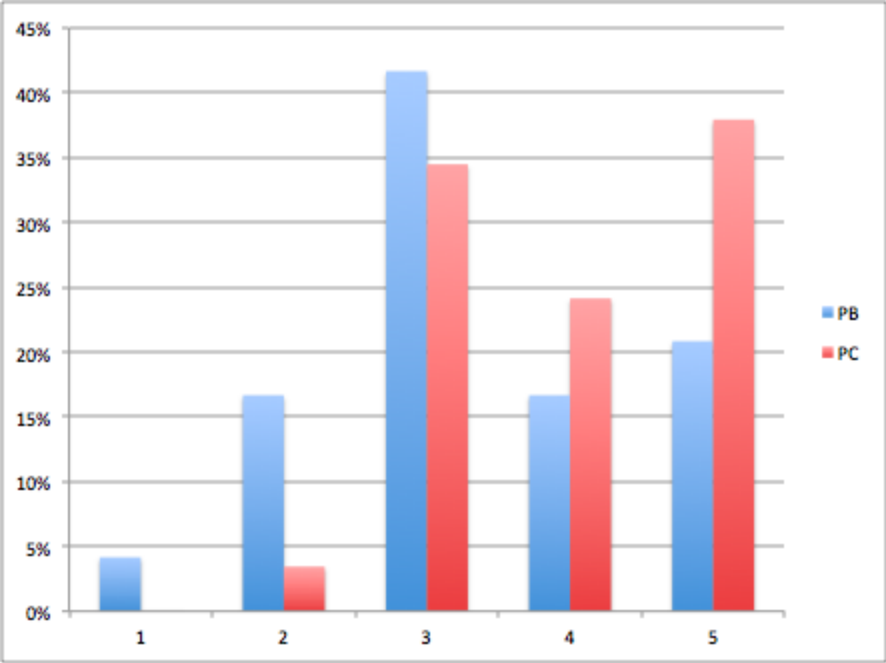
Note that here the scale is between too short and too long. It is clear that the PC thought the process was too long, whereas the PB was more evenly split. The PB view in 2014 was 10% thought it was too short and 30% too long (the rest were neutral), whereas in 2016 it was 21% and 38%, respectively. In 2014 the PC said 7% too short (3% in 2016) and 57% too long (62% in 2016). In both 2014 and 2016 the online discussions lasted for about 4 weeks and although the PC is slightly more negative now than they were in 2014, but not by much.

Next the question of engagement in the online discussions were asked about with the following three similar questions:

"I participated fully in the online discussion" (PC)
"Most other program committee and program board members participated fully in the online discussion" (PC)
**"Program committee members participated fully in the online discussion." (PB)**
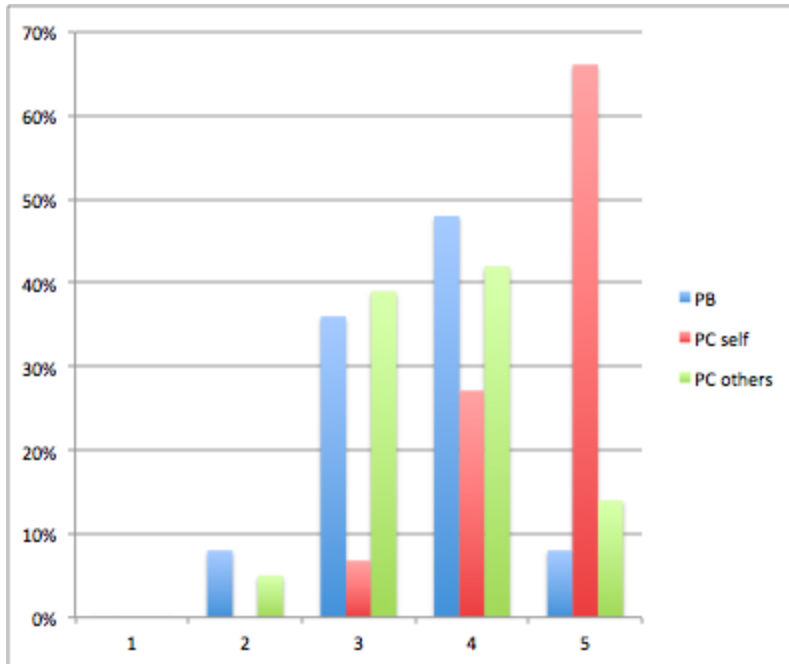
As can be appreciated, the PC thought their own participation level was better than what the PB thought or what the other PC members thought. The PB and the PC's opinion of the other reviewers are much more closely aligned. The PC self-assessment numbers (93% positive) look similar to 2014 (94% positive), but the PB (56% positive) and the PC's view of other PC members (56% positive)) are more negative this time round: PB 2014 (95% positive) and PC 2014 (83% positive). This is likely again to do with the Likert scales being different since the negative numbers in each case matches almost exactly (8% PB vs 5% in 2014 and 5% PC vs 5% in 2014).

ICSE is historically a conference that makes decisions at the face-to-face meeting, but the PB model relies on the fact that some decisions are made online and they may, or may not be revisited at the face-to-face PB meeting. Both the PC and the PB was asked:

"I was comfortable making some accept/reject decisions online."
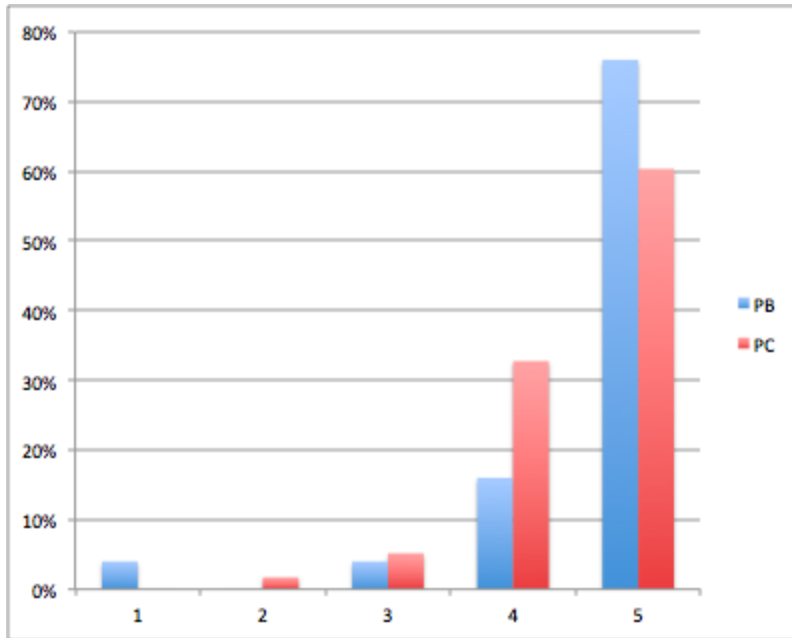
Both the PC and PB were very comfortable making decisions online (93% and 92% positive respectively). In 2014 the PC was 90% positive (the PB was not asked this in 2014).

4.3.2.2 Program Board Model

With regards to the process the most important questions related to the workings of the program board model and the interactions between the PB and PC. A change made in 2016 was to have a PB member do a full review for papers that passed the online discussion phase. In 2014 this was called the 2nd Reader for a paper and they did not do a full review. It was thought it was better to have them do a full review and have the PC see the review and allow discussion before the PB meeting. This also meant there were two PB members familiar with the paper in the room at the PB meeting, the one who oversaw the discussion and the one that did the full review.

The first questions to the PC and PB were whether they found this extra review helpful:

"The program board review brought a helpful perspective to our discussion" (PC)
**"Having a program board member perform a full review was a good idea" (PB)**
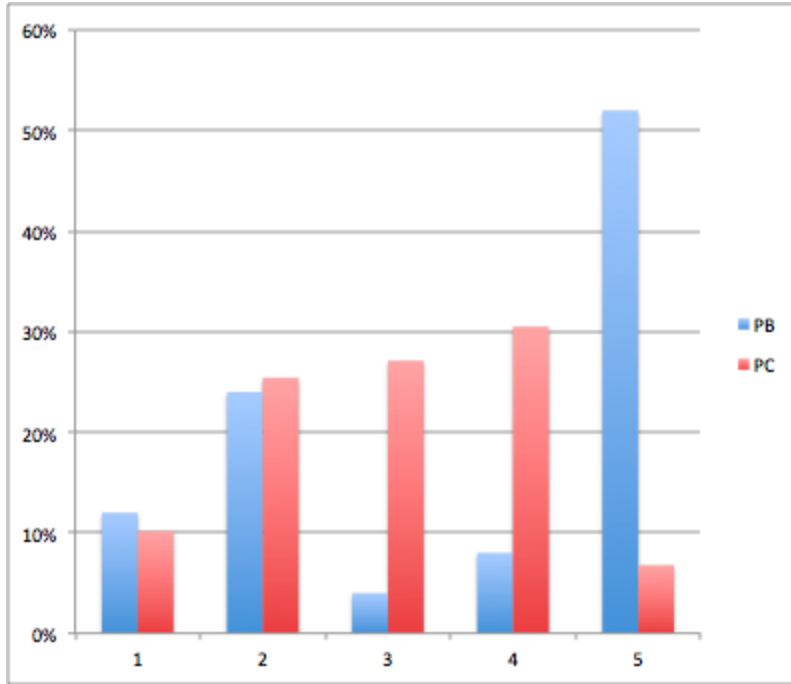
Even though 60% of the PB felt this was a good thing, the PC is almost evenly split with 38% positive and 35% negative. The program chairs believe this was very helpful in the PB meeting to have more input, but the picture from the reviewers seem to be less clear.

Asking the PC whether they were comfortable with having the PB member making the final decisions had a similar trend as above:

"I was comfortable leaving some accept/reject decisions to the program board"
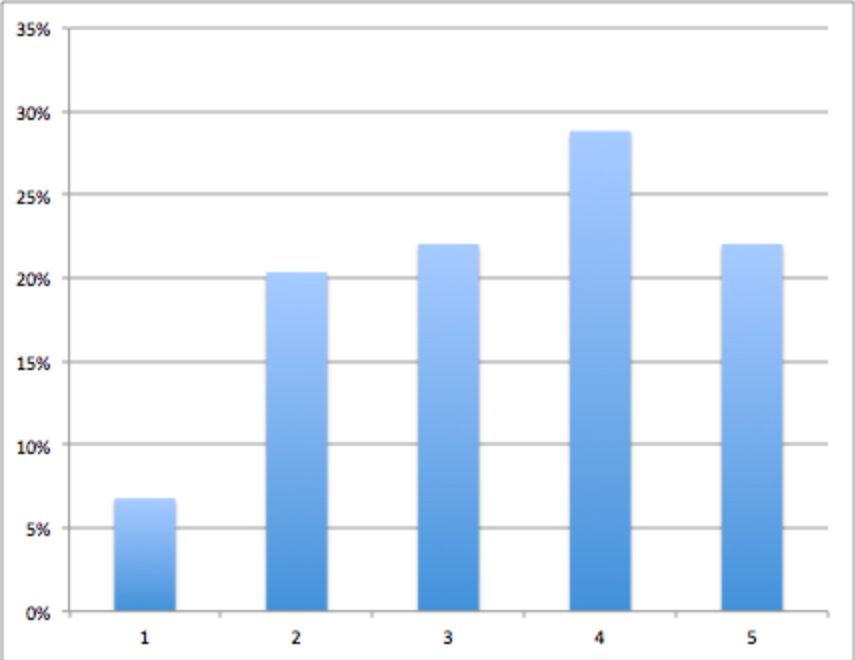
Only 51% was positive about leaving decisions to the PB and 27% negative. This doesn't compare favourably to the 2014 numbers where 70% was positive and 22% negative. This doesn't bode well for the model if there is this level of discontent from the PC. In the future, we need to continue to watch the sentiment of the PC toward the PB decision making.

The last question related to whether the final outcomes of the decisions at the PB meeting were good. This was asked to the PB, since they were in the meeting.

"Final decisions of acceptance/rejection were balanced and justified."
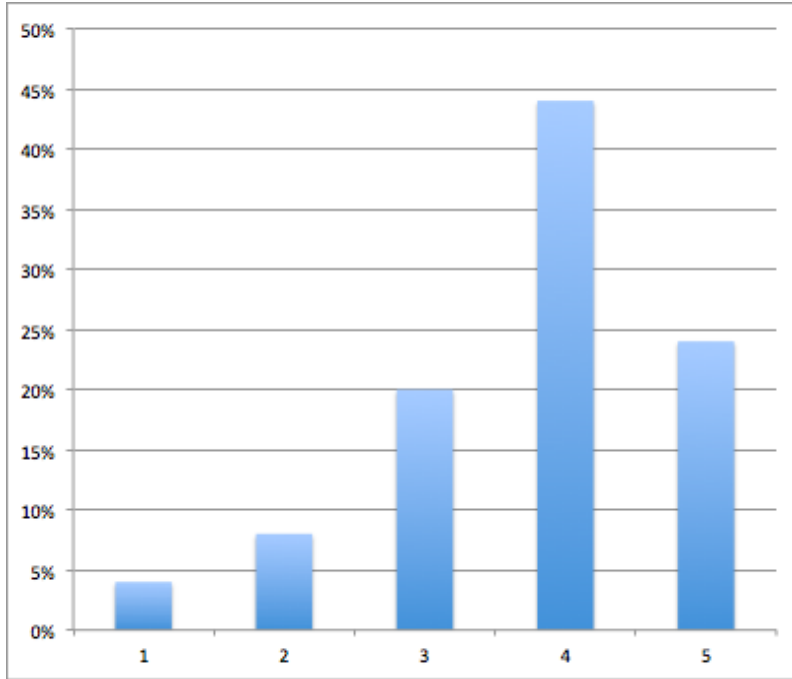

The PB was happy in that 68% was positive and only 12% negative. In 2014 85% was positive, but again due to the Likert scales being different, the 10% being negative indicates the scores are closer than it would seem.

4.4 Double Blind

For some time now, there has been a discussion whether ICSE should go double blind in its reviewing process. It was decided to ask this question to all authors of submitted papers to see what the community feel is.

"ICSE should use double blind reviewing"
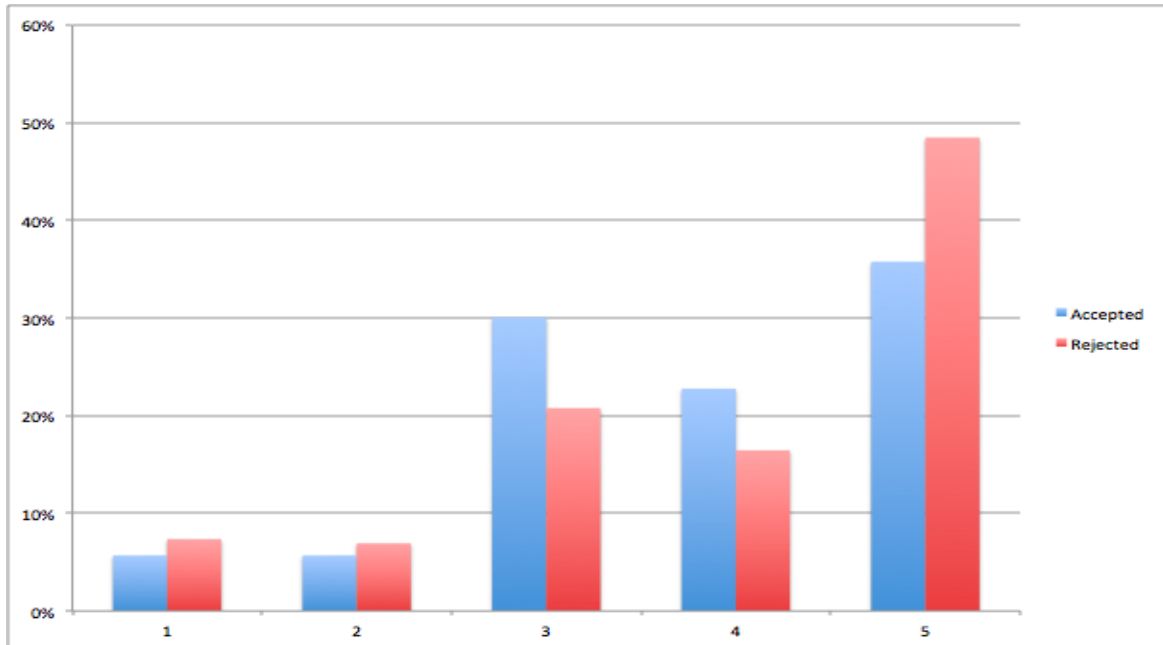
This is again broken out into accept and reject, and one can see the authors of rejected papers are more in favour of double blind. Overall, the trend is overwhelmingly towards double blind: 63% of the total authors polled were in favour, only 15% against and 24% neutral.

5 Chair Comments

Here we would like to comment on a few things we observed along the way.

5.1 Program Board Model

Different reviewing models try to address different aspects of this intractable optimization problem. However, one thing is clear, the number of submitted papers are going up steadily and having all reviewers in the room at the face-to-face meeting is becoming increasingly impractical. We strongly believe the PB model is the way to address this problem since the extensive online discussions that this model requires to work is what makes the feedback to the authors more valuable. Even the authors of rejected papers liked the feedback they got in the discussion summaries. This year, we also had 2 PB members that intimately knew each paper that came up for discussion, although the extra review from a PB member did seem to cause some discomfort from the PC. Lastly, the PB model allows more people that did not review a paper to contribute to the discussion. This happened in a large number of cases this year; and mostly this was positive for acceptance as can be seen from the 65% acceptance rate at the meeting (much higher than the last 2 years). We share some lessons learned:

- We utilized the three PC reviews, the rebuttal, and an extensive online discussion to reduce the number of papers that got a full PB review. However, this process lead to the need for two major online discussion cycles, the second being right before the PB meeting. PC members felt the duration of their involvement was too long.
- The PC members commented that they perceived a hierarchy where the PB member overturned their decisions. While the ultimate decisions do not show many papers being overturned by the PB, the PC chairs did need to actively manage the dynamics when the PB member was more negative resulting in fewer overturned papers than might have happened. The PC members sometimes chose to disengage from the discussion when the PB member persisted in having a dissenting voice. PB members should be repeatedly reminded through the process to respect the opinions of the PC team for each paper.
- Based upon the above two bullet points, we recommend that papers who have clear consensus (accept or reject) from the three PC members do not get a PB review (alleviating the perception (or reality) of the hard-earned decision of the PC being overturned). A PB review is conducted on the undecided papers prior to rebuttal phase. Jointly, the PB and PC discuss the additional PB review and rebuttal.
- In some cases, the PB reviewer had a sub-reviewer which made the PC member question the true value of the PB review beyond the PC reviews. The PB reviewer should not be allowed to have a sub-reviewer.
- The PB should be given more time to do their reviews than occurred in our schedule.
- One person commented that a PB member should have to have been on the PC within the last two years so they can experience the dynamics between PB and PC members.

- Several commented that PC members should be allowed to call and/or skype into the PB meeting. This practice would be beneficial though very difficult to implement with conflicts of interest and the schedule of the PB meeting being so unpredictable.
- The only way the PB model will succeed is if the PC members feel they are first-class citizens and valued. The PC chairs and the PB should keep this in mind in all ways possible.

5.2 Author Involvement

We decided each paper should get at least 3 reviews, since in 2014 it was observed from the survey data that authors of papers that only got 2 reviews were much less happy with the outcome than those with 3 reviews and also a rejection.

We also introduced an author response option that was met with overwhelming approval from the authors. At the same time, the authors were asked to provide feedback on their reviewers. Our process showed that asking for reviewer feedback before decisions are made is beneficial. That also shows that early rejections should be avoided in most cases, especially for PC/PB papers.

We strongly recommend keeping the author response option. We also strongly recommend keeping the review rating feature in future. We used author feedback on reviewers in our decision process for identifying the reviewers to be awarded for being the top reviewers.

However, the authors were not given the opportunity to do a rebuttal of the PB review. Authors commented that the fourth review was often the most critical and that they wished they had an opportunity to respond. Additionally, many authors commented that they felt the rebuttal was not considered in the final outcome. One PC member suggested that the ultimate review (and/ or review summary) should specifically comment on the impact of the rebuttal on the review process.

5.3 Matching Reviewers to Papers

The current process for doing paper bidding is not bad, but unfortunately it is based on faulty information: (1) authors often pick the wrong categories for their papers; and (2) PC and PB member primarily choose based upon paper titles alone, and paper titles can be misleading. Reviewers simply don't always read all the 530 abstracts, nobody can blame them. Then you also get reviewers that simply bid low on everything and the perennial problem of not bidding high (or low) on enough papers.

What is required is more automation: matching authors by analyzing the text of at least title and abstracts to match it against self-extracted categories from say the last 5 years of publications and doing the same for each submission. This will not be hard and would do miles better than the current system. This was also suggested in the 2015 report, it might be time to act on it.

5.4 Double Bind

Clearly the community wants this and it should be introduced at the next opportunity, i.e. ICSE 2018.

5.5 Title Changes

An annoying snag we hit this year was not setting a firm deadline for when title changes must be made. Some reviews requested title changes and authors were keen to make them, but sent new titles until the camera-ready due date, by that time the old titles have proliferated through the various processes (copyright forms, session assignments, best paper selections, etc.).

Future chairs, set a title change deadline for no later than 2 weeks after the accept decisions have been sent out.

5.6 Communication

This year, we made the official slide deck that was shown to the PB and PC available online on a Google+ page and also held a Hangout sessions where we discussed its contents. See https://plus.google.com/communities/110196757571521827561. We believe this level of transparency into the process is good for the community and should be done in the future as well.

5.7 Reviewing Platform

CyberChairPro and the tireless support of Richard van de Stadt was priceless, as always. Without Richard, this tool will be hard to use. But, he was on tap to solve every conceivable issue we had and to run special custom jobs. Anybody that wants to use a different system should think long and hard, since it would be hard to believe this level of service would be available at any price. Thanks Richard!