

Preliminary Advice

Combining AIEEE Scores and Board Assessments for Tertiary Selection Purposes in India

Prepared for Mr Vineet Joshi, Chairman, CBSE
by
Glenn Rowley



Australian Council for Educational Research

October 22, 2012

Contact

Dr Glenn Rowley
Principal Research Fellow
Australian Council for Educational Research
Private Bag 55
CAMBERWELL VIC 3124
Tel: +613 9277 5443 Fax: +613 9277 5500
Email: glenn.rowley@acer.edu.au

Table of Contents

The data supplied	3
The advice	3
Assumptions	4
The first datasets provided.....	5
The second datasets provided	7
Board AM.....	8
Board BR	9
Board CBS.....	10
Board GU.....	11
Board KL	12
Board MR	13
Board RJ	14
Board TN	15
Board UP	16
Differences by Board	17
Demonstration of the procedures	18
Conclusion.....	21

The Request

To provide advice on how a set of scores from national examinations (AIEEE) in Physics, Chemistry and Mathematics) can be combined with scores provided by 29 different Boards across India to generate a single score that can be used for tertiary selection in Engineering across all 29 Boards.

The data supplied

Initially, I was provided with three files:

EEECBS10.zip
EEECBS11.zip
EEECBS12.zip.

On October 18 I was provided with a further set of data files:

EEEAM12.xlsx
EEEBR12.xlsx
EEECBS12.xlsx
EEEGU12.xlsx
EEEKL12.xlsx
EEEMR12.xlsx
EEERJ12.xlsx
EETN12.xlsx
EEEUP12.xlsx.

This advice is provided on the basis of an appraisal of these files and some limited trialling of the proposed strategies.

The advice

On the basis of my examination of the data provided, I propose that the following procedures be considered.

1. For each Board, the AIEEE and Board data need to be scrutinised carefully, to ensure that the data are free from irregularities and evidence of manipulation. Once satisfied with the quality of the data, apply the procedures that follow within each set of Board scores.
2. Apply a transformation to the AIEEE Total scores (the sum of examination scores in Physics, Chemistry and Mathematics) to convert the distribution from the (typically) skewed raw score distribution to a normal distribution.
3. Adjust the mean and standard deviation of this normalised distribution to the same mean and standard deviation as that of the original AIEEE Total scores *from that Board*.
4. Apply a transformation to each set of Board scores to convert the distribution from the (typically) skewed raw score distribution to a normal distribution.
5. Adjust the mean and standard deviation of this normalised distribution to the same mean and standard deviation as that of the original AIEEE Total scores *from that Board*.

6. Form a composite score by combining the transformed AIEE scores and the transformed Board scores in 60:40 proportion (60% AIEEE Total and 40% Board scores).
7. The composite will have the same mean but not the same standard deviation as the original distribution of AIEE scores from that Board, and its distribution will be not quite normal. Normalise the distribution and adjust the mean and standard distribution to match the mean and standard deviation of the AIEEE Total scores *from that Board*.
8. Repeat the process in all 29 Boards.

At the conclusion of this process, each Board will have a set of scores that is normally distributed, and with means and standard deviations that match those of the scores obtained by its own candidates on the AIEEE Total. To the extent possible, scores from different Boards will now be comparable.

9. Aggregate the scores from all 29 Boards into a single data file. These scores are now suitable for use in tertiary selection.
10. If the scores are to be made publicly available, they need to be converted to a scale more suited for this purpose, and one that will not be confused with either AIEEE scores or Board scores. Three possibilities would be:
 - a. A normal distribution with an arbitrarily chosen mean and standard deviation (such as the mean 500 standard deviation 100 commonly to used to present the results of international surveys);
 - b. Percentiles (to one or two decimal places), ranging from 100 down to essentially zero; and
 - c. Percentiles (to one or two decimal places), ranging from 100 down to a number that reflects a relevant proportion; e.g. the percentage of the age cohort across all 29 Boards who present for the AIEEE examination (as used in the Australian Tertiary Entrance Rank).

Assumptions

The procedures described above are based on the following assumptions:

1. In populations of the size we are dealing with here, distributions of achievement will be normal, or approximately so.
2. AIEEE Total scores are directly comparable with one another, whether they come from the same Board or from different Boards.
3. Board scores from candidates within the same Board are directly comparable to one another.
4. Board scores from candidates from different Boards are not directly comparable to one another.
5. Differences in the distributions of AIEEE Total scores across Boards provide the best available indication of the differences in distributions of achievement across the Boards.

The first datasets provided

The first datasets provided to me were in the files EEECBS10.zip, EEECBS11.zip and EEECBS12.zip. For each of the years 20-10, 2011 and 2012, they contained data for the Board CBSE as follows:

AIEEE Physics

Board marks: Physics

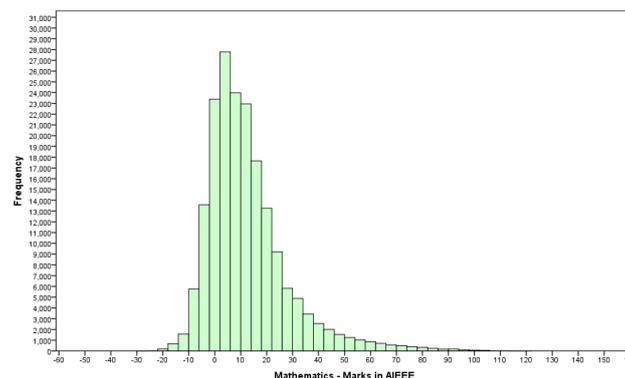
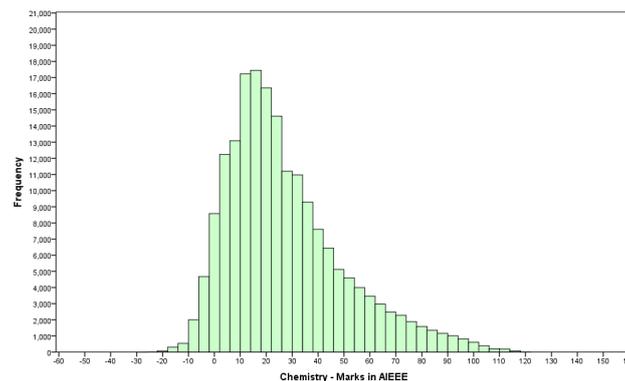
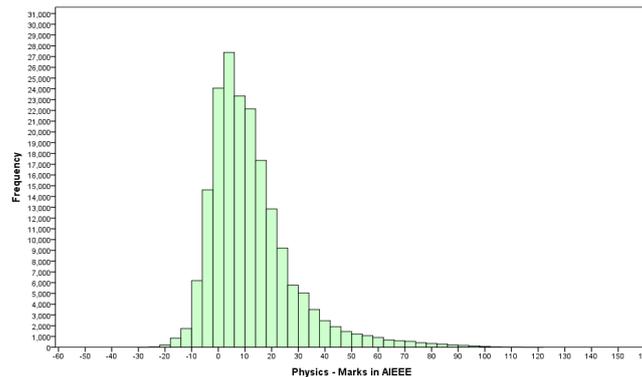
AIEEE Chemistry

Board marks: Chemistry

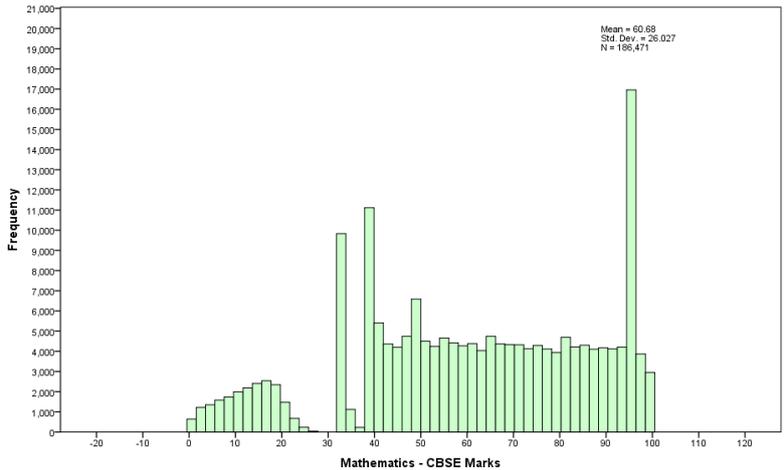
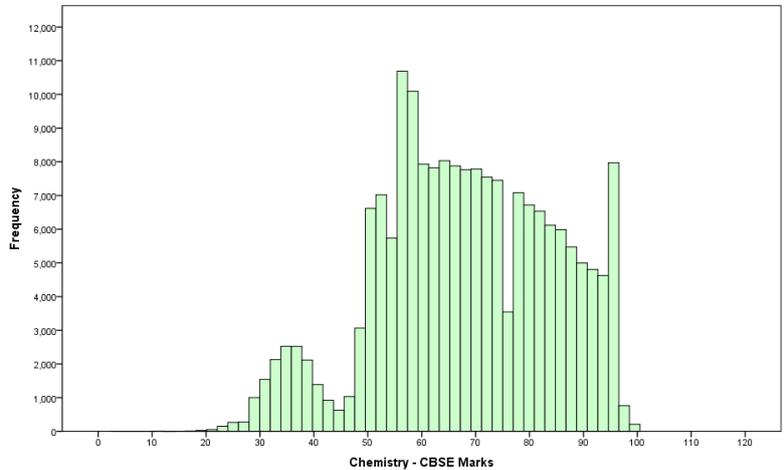
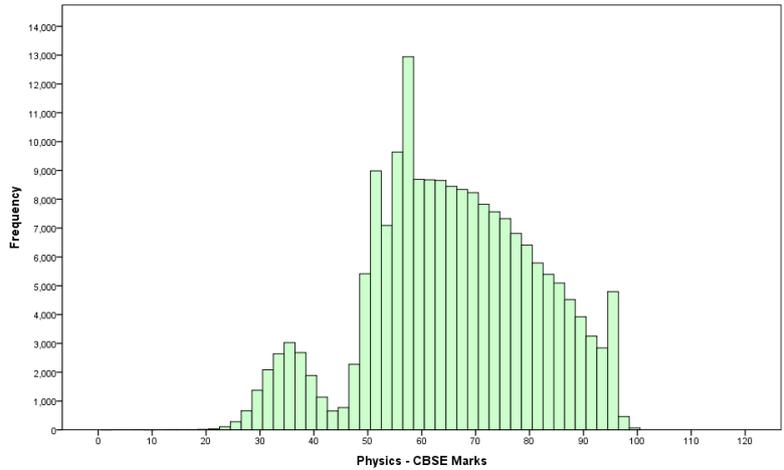
AIEEE Mathematics

Board marks: Mathematics

The AIEEE marks were for year 2011 were examined first, and were distributed as follows



The distributions are as would be expected in any situation where a set of common examinations have been administered to a large population of candidates. There is a positive skew indicating that the examination is able to separate candidates at the upper end better than at the lower end of the achievement scale – a desirable attribute given its current use. The distributions of Board scores were quite different:



These bear no resemblance to any known distribution, and defy explanation. They appear to contain at least two separate populations. One is small, nearly normal, and at the lower end of the reported scores. The other is larger and is almost completely separate from the first in Physics and Chemistry and completely separate in Mathematics. For each subject, the second, larger distribution has peaks that suggest human intervention to raise candidates over

some score, such as a passing score or perhaps the score that required to gain some reward, such as course selection or a scholarship. In Mathematics the distribution approximates a uniform distribution, suggesting that the bulk of the scores submitted were rankings or percentiles.

In the absence of any explanation for these irregularities, I cannot recommend that such data could contribute usefully to tertiary selection. AIEEE Total scores appear to be quite well-suited for tertiary selection, and to form any combination with scores distributed like these would detract seriously from their effectiveness.

Fortunately, these problems were absent from the 2102 data that were provided next. The remainder of the report will be based on 2012 data, which were free of these irregularities.

The second datasets provided

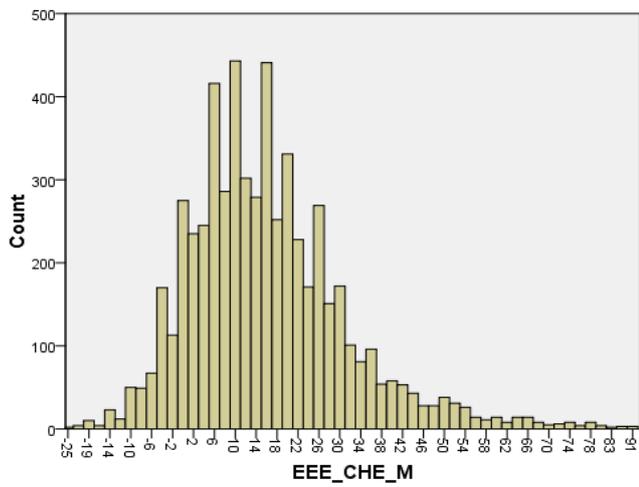
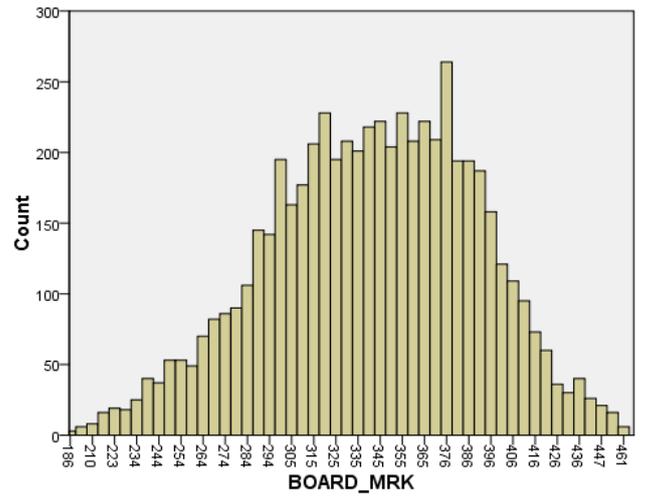
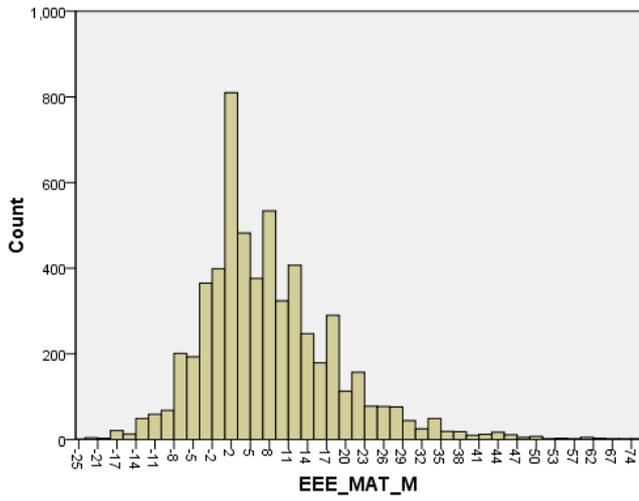
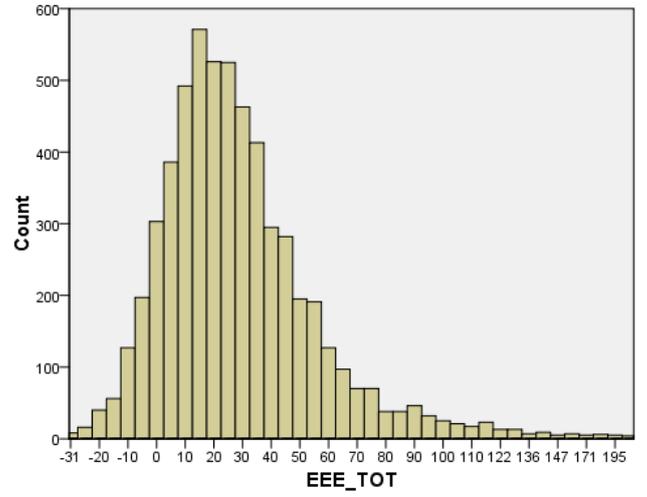
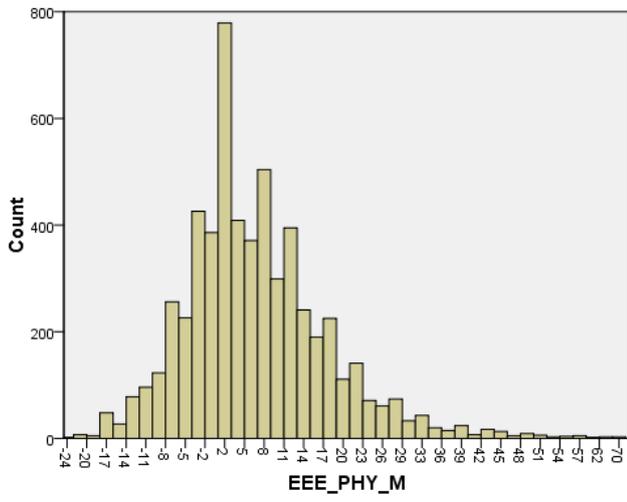
On October 18 I was provided with a further set of data files with the filenames

EEEAM12.xlsx
EEEBR12.xlsx
EEECBS12.xlsx
EEEGU12.xlsx
EEEKL12.xlsx
EEEMR12.xlsx
EEERJ12.xlsx
EEETN12.xlsx
EEEUP12.xlsx.

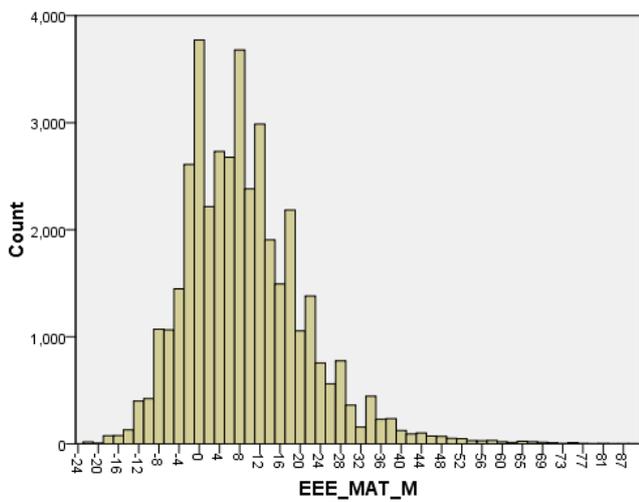
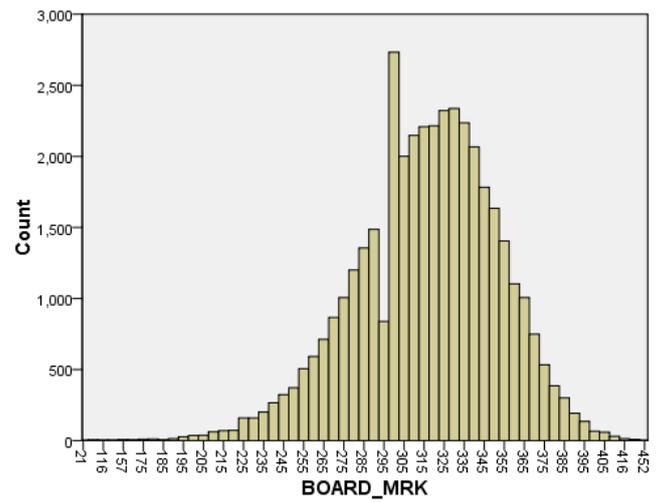
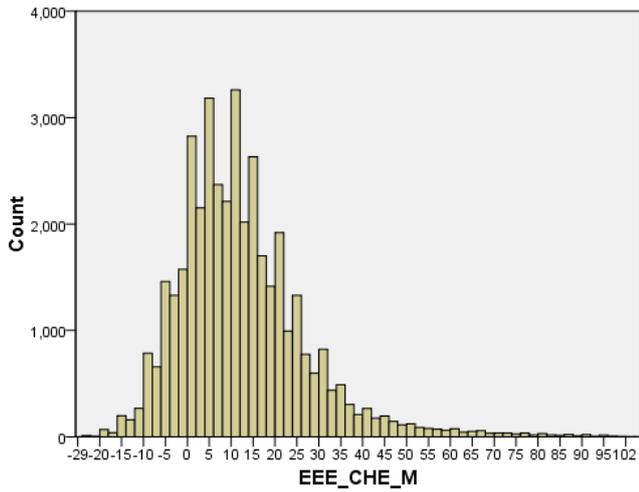
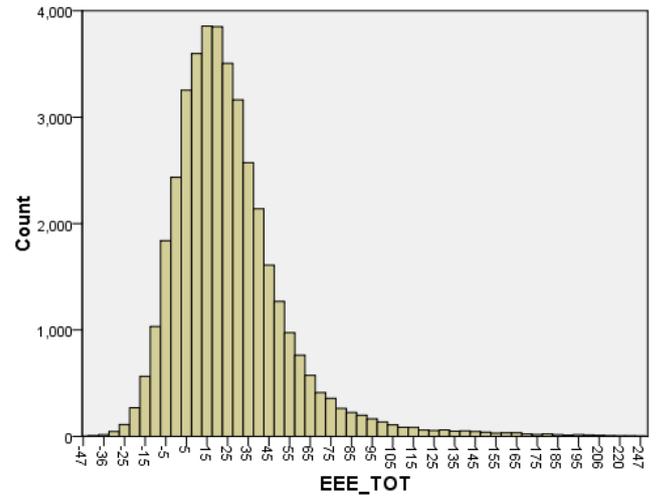
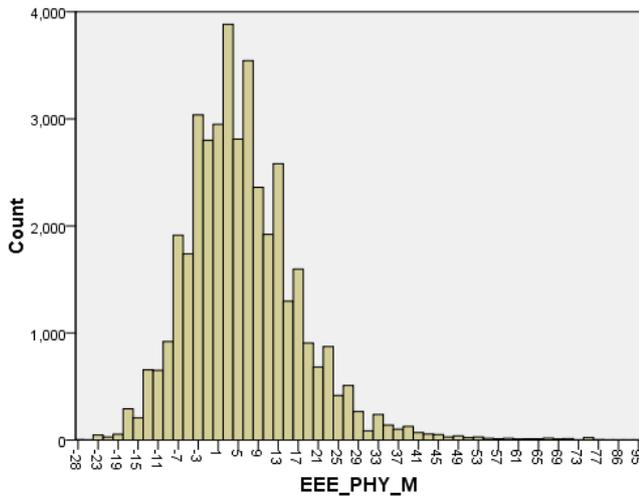
These appear to be 2012 data from nine Boards, for which I will use the acronyms AM, BR, CBS, GU, KL, MR, RJ, TN and UP. Each file includes variables labelled EEE_PHY_M, EEE_CHE_M, EEE_MAT_M, EEE_TOT and BOARD_MRK, which I interpret as AIEE marks in Physics, Chemistry and Mathematics, Total AIEEE scores and Board scores. It is not clear how the Board scores were arrived at, but if they were arrived at by a process that is uniform within each Board, they are potentially usable for selection.

For each of the nine Boards in turn, the distributions of the four AIEEE scores (Physics, Chemistry and Mathematics) and the Board scores are portrayed in the next nine pages of charts.

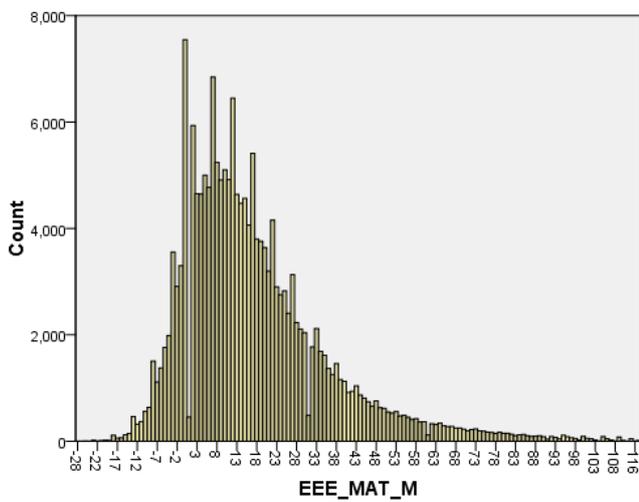
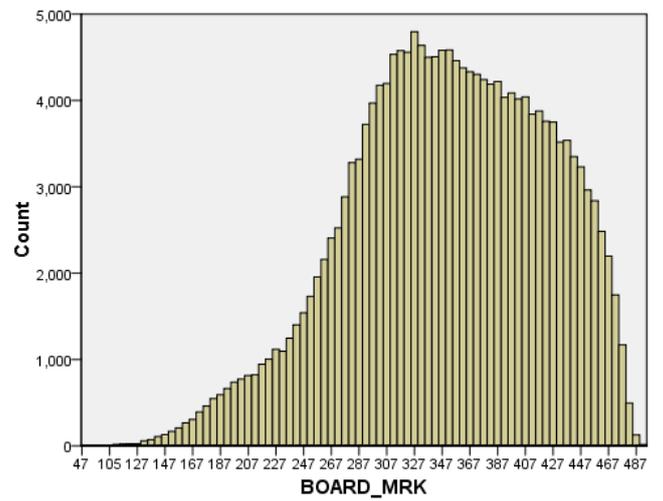
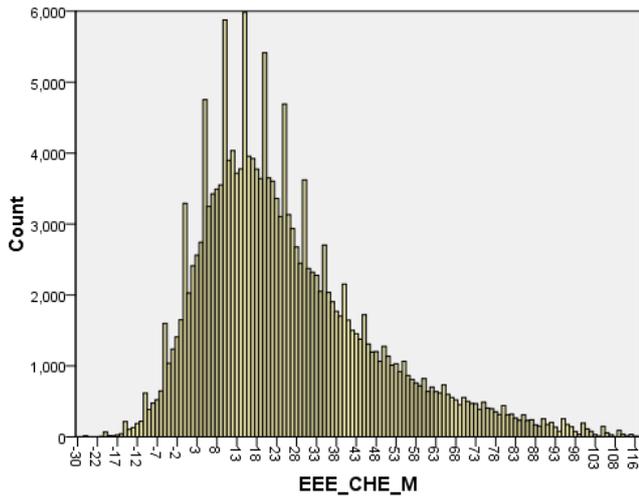
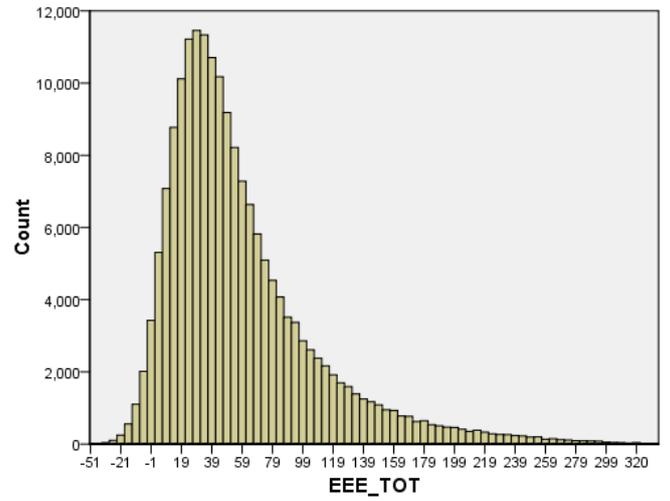
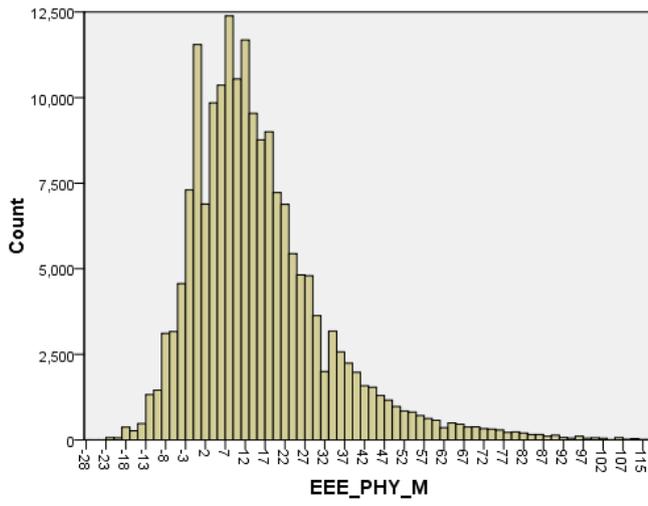
Board AM



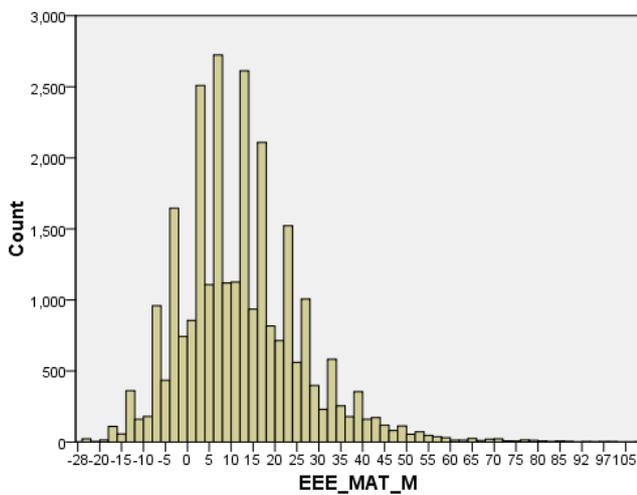
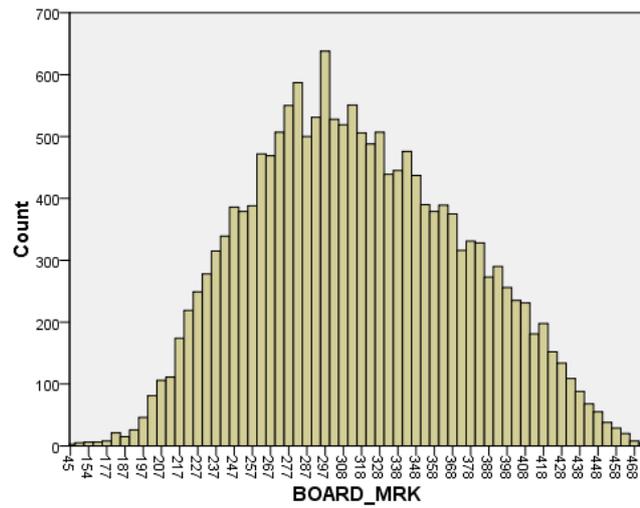
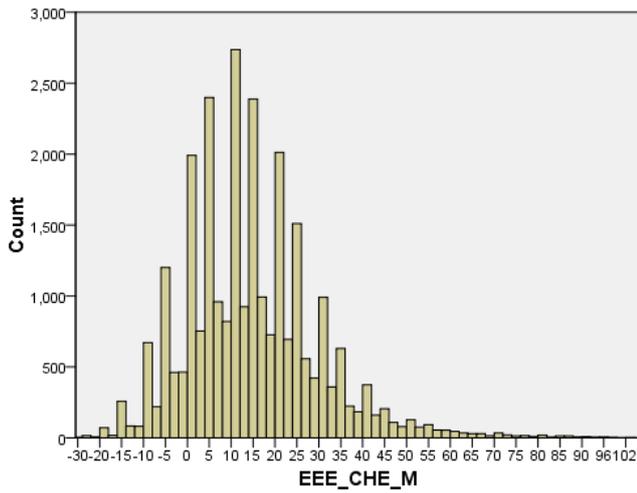
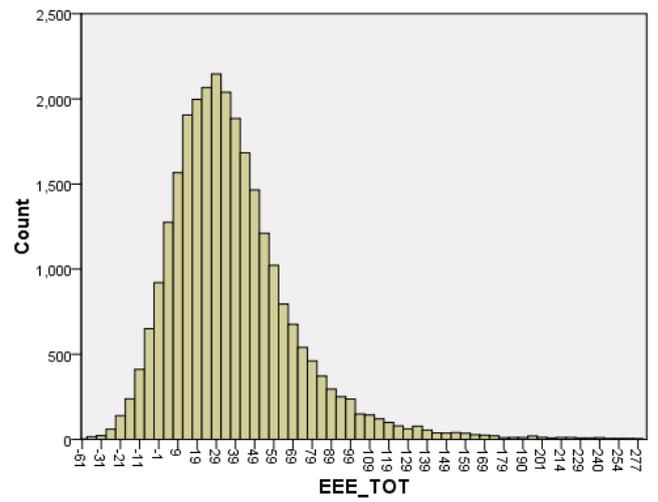
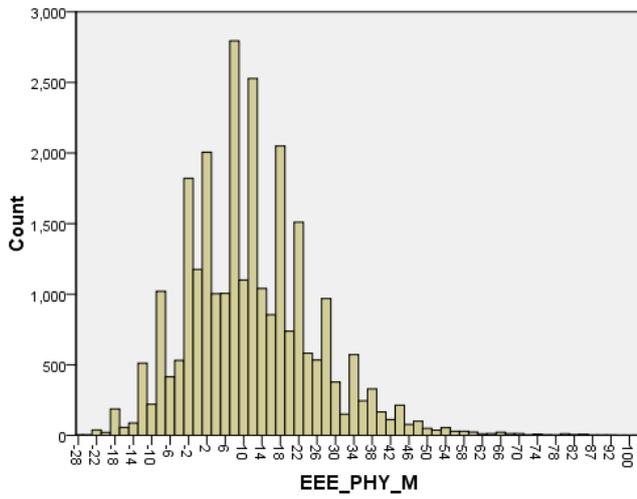
Board BR



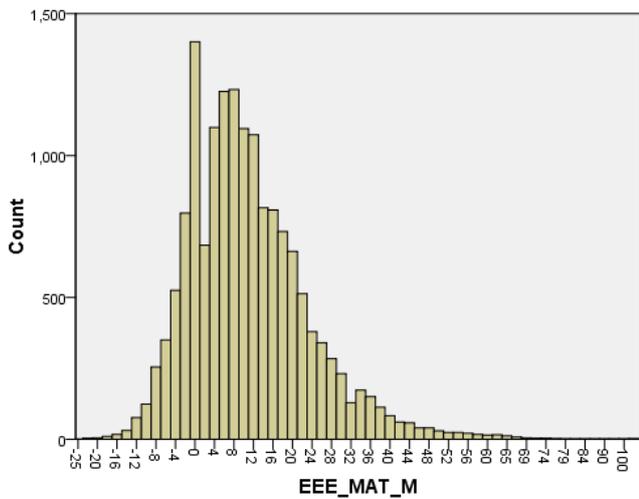
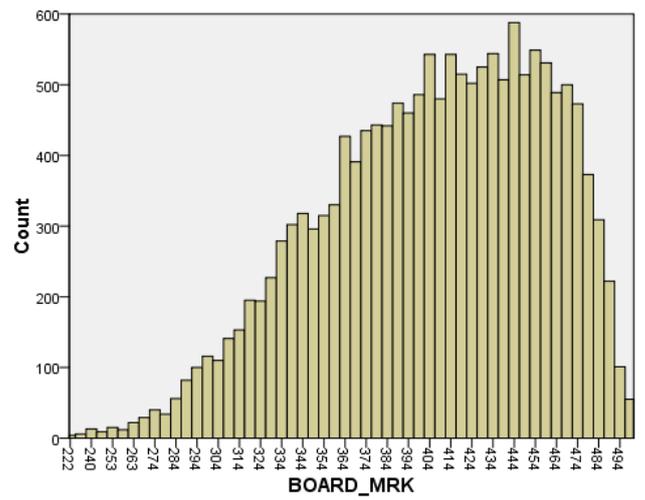
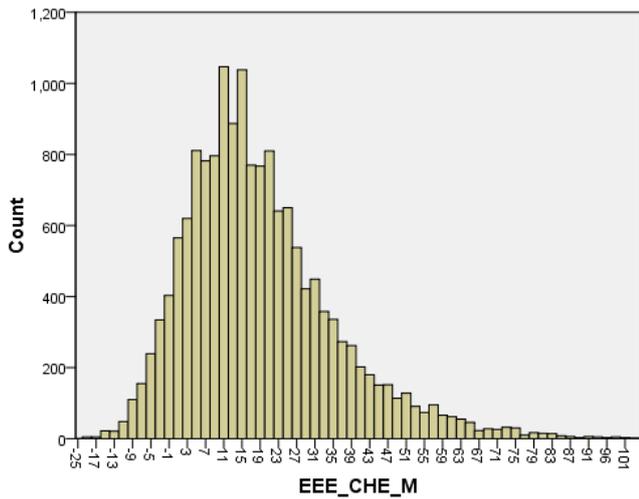
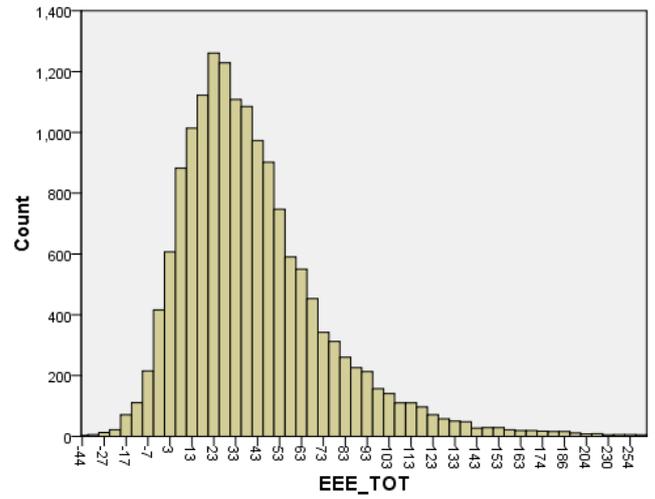
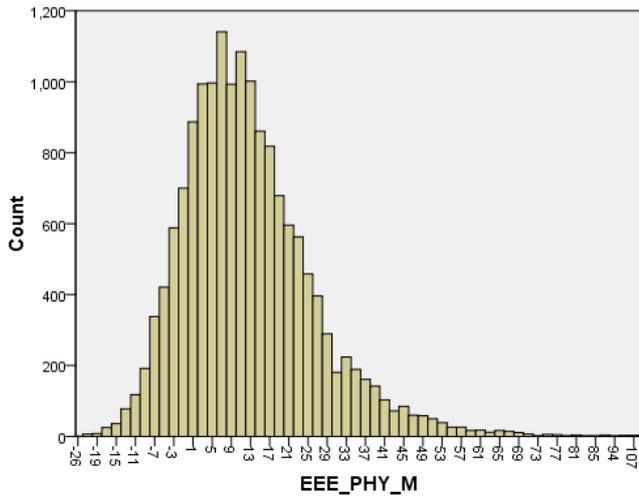
Board CBS



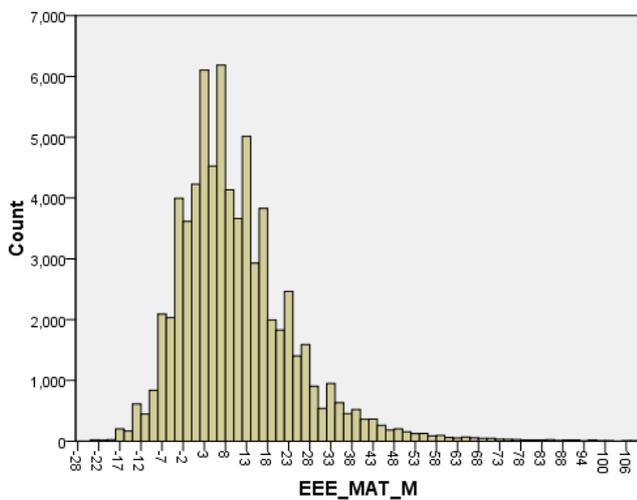
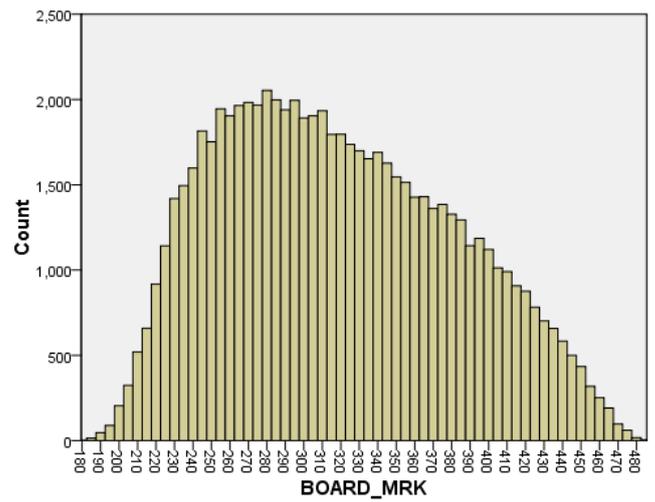
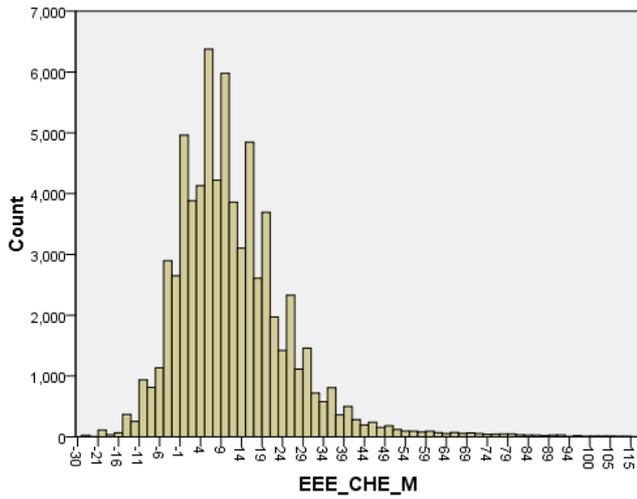
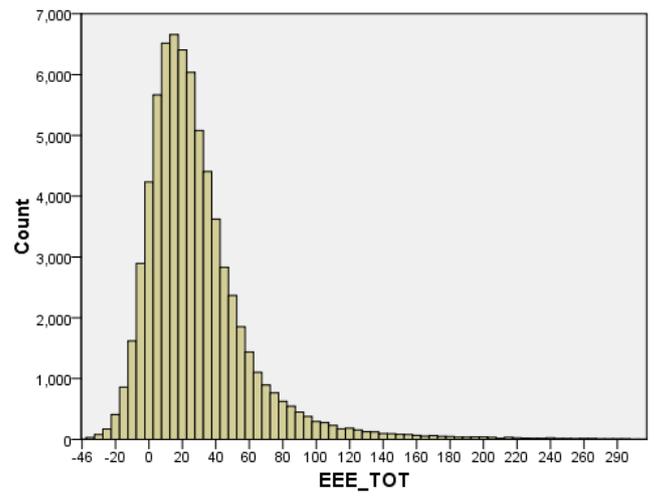
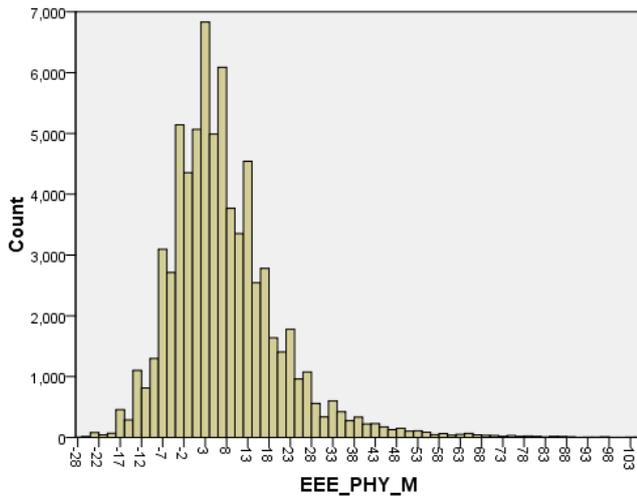
Board GU



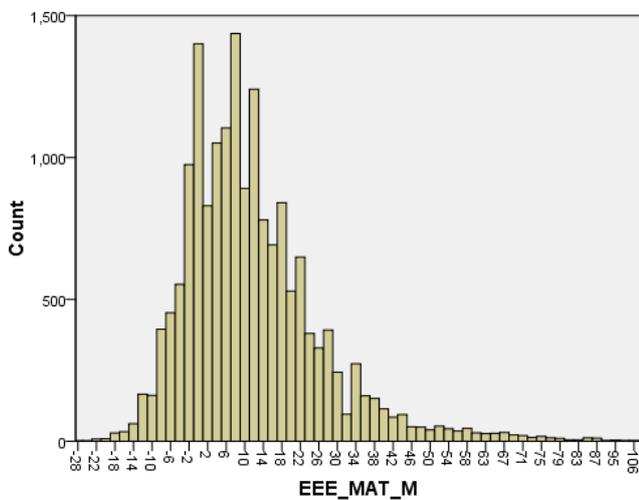
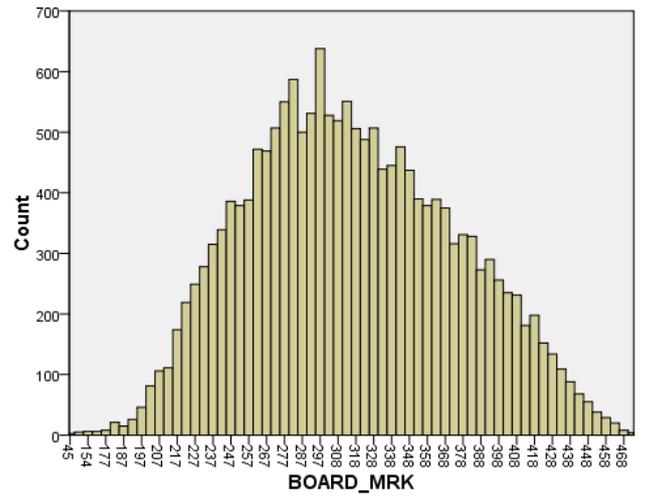
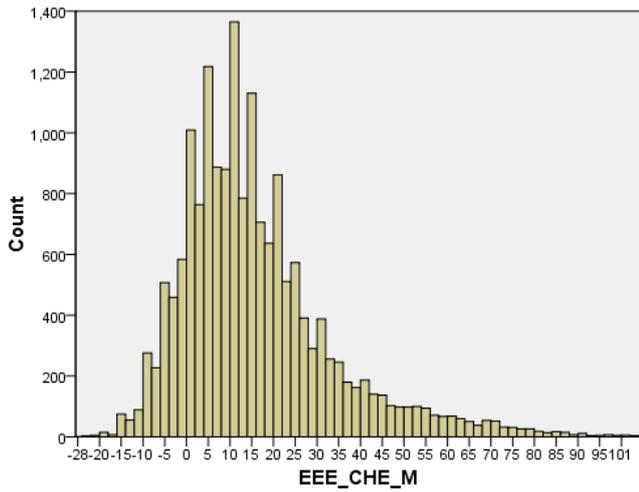
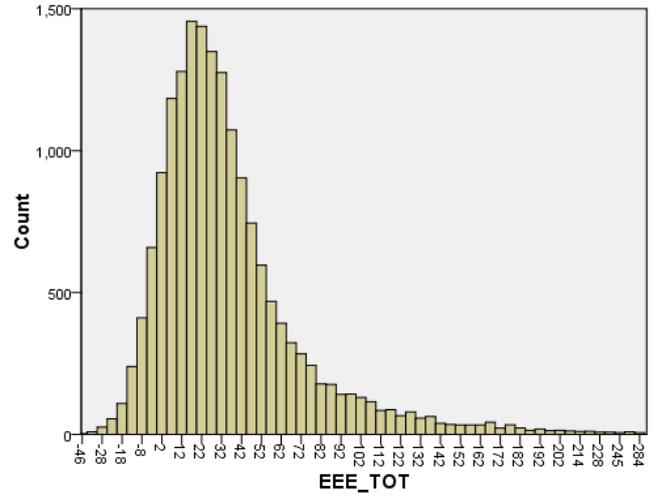
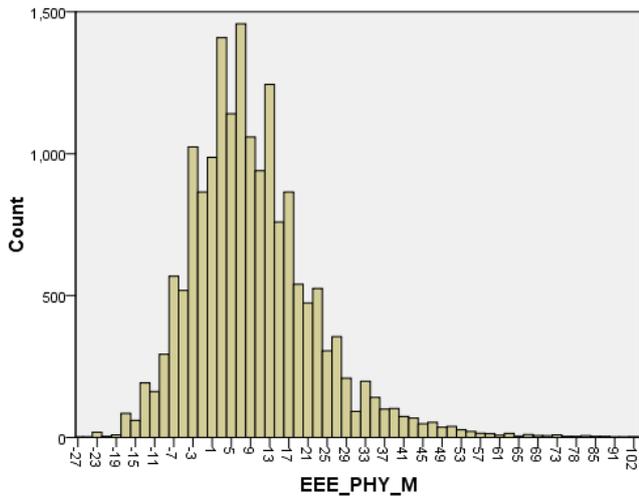
Board KL



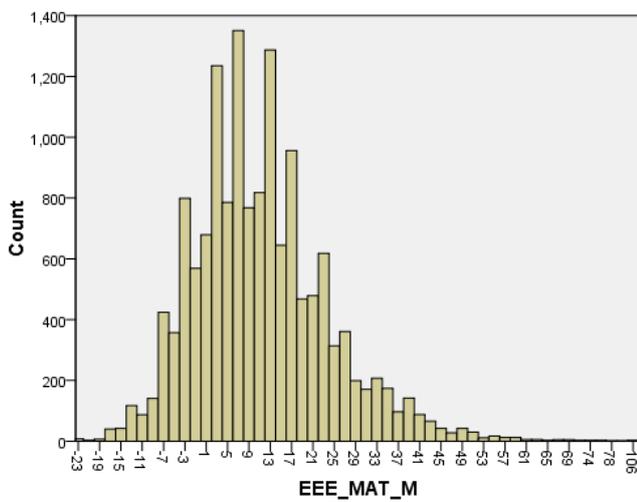
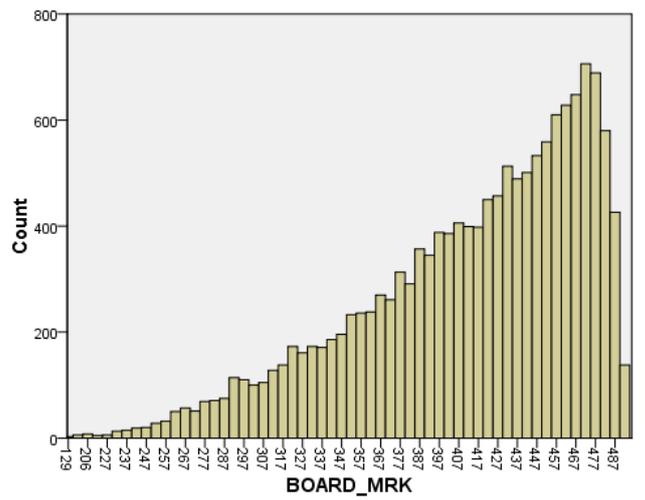
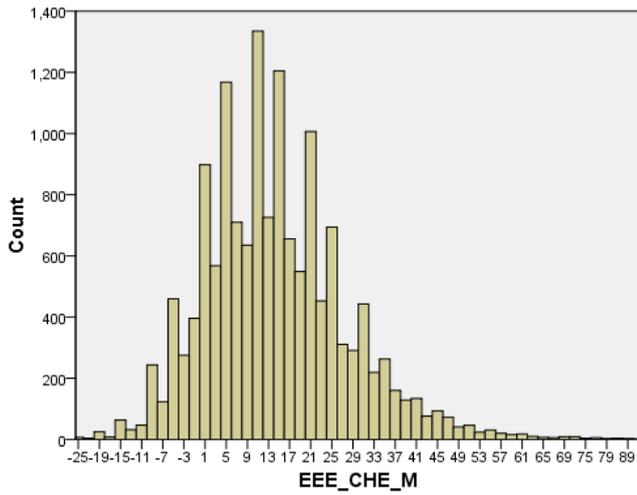
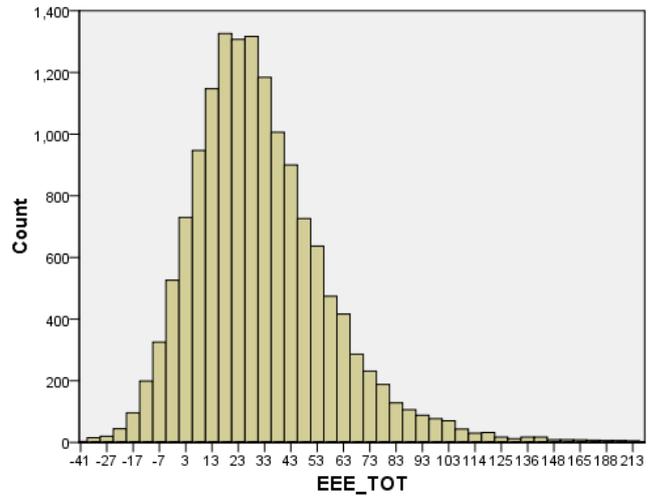
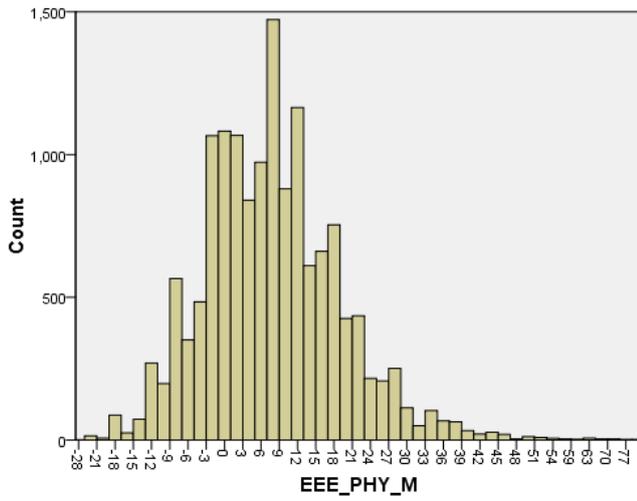
Board MR



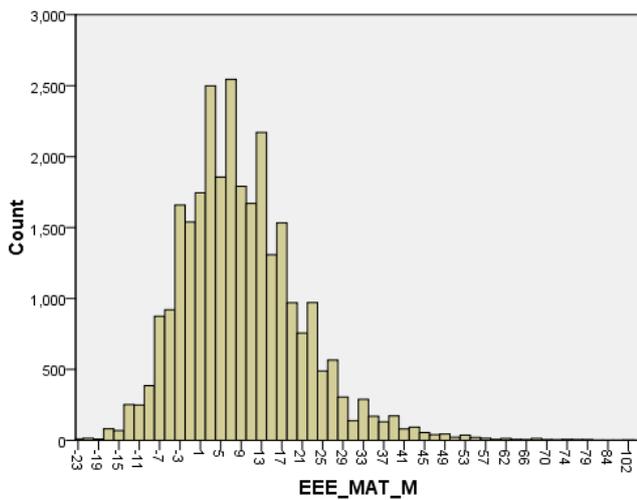
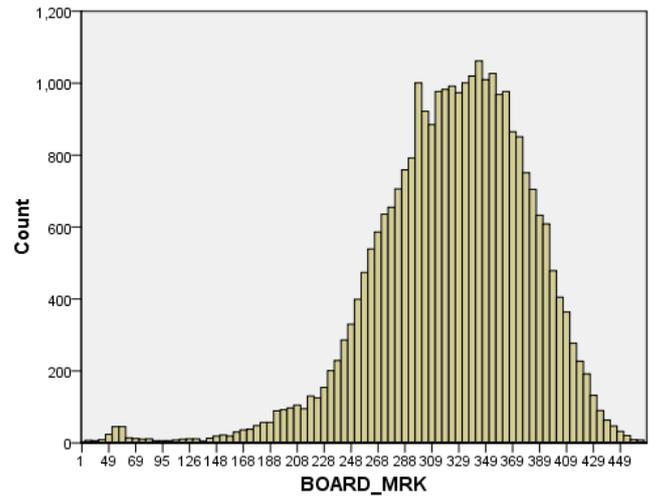
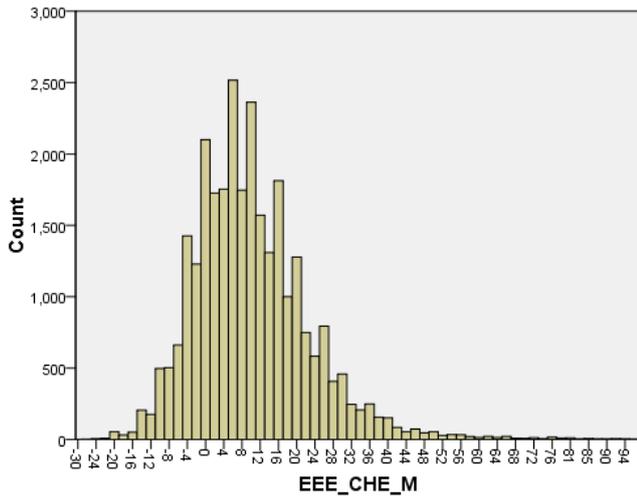
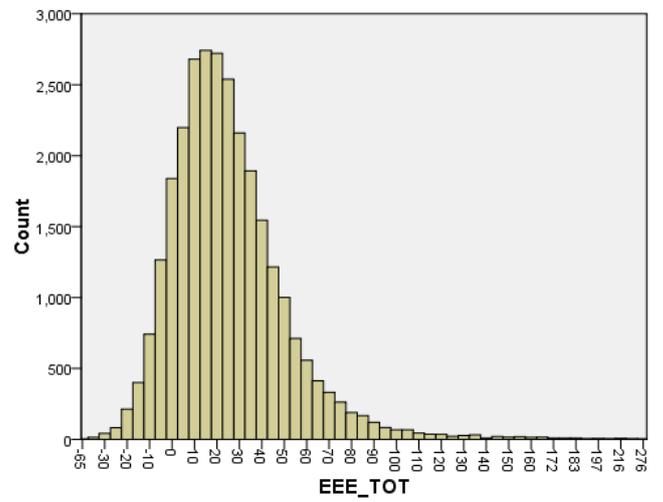
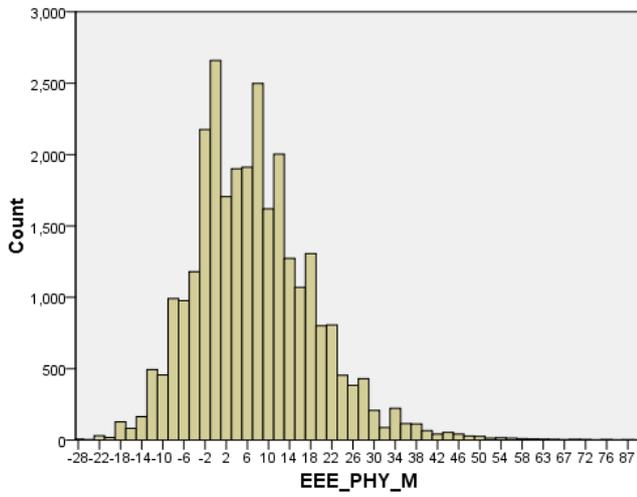
Board RJ



Board TN



Board UP



From these charts, it appears that:

1. In each of the nine Boards, the AIEEE scores are distributed in the manner expected. The single-subject scores are more “chunky,” because they have less possible score points. The AIEEE Total scores in each Board are smoother and have similar distributions – close to normal, but with some positive skew.

The Board scores are more varied:

2. For Board AM the score distribution is close to normal, with slight “bunching” at three points. These scores could be used, but the precision of selection would suffer slightly because of the inability of the scores to discriminate between candidates at these levels. Given that the levels are below the point at which the keenest selection decisions are made, this may not be a major issue.
3. For Board AM the score distribution is close to normal, with slight “bunching” at one point, suggesting human intervention after the marks had been awarded. These scores could be used for selection. As for Board AM, the problematic scores are below average and well below the point at which the keenest selection decisions are made, so this need not be a major issue.
4. For Board CBS, the score distribution is both positively skewed and asymmetrical. Neither of these presents any problem, and can be taken care of by normalising the scores. There will be some loss of precision because the examination discriminated less well at the crucial upper end of the score range.
5. For Board GU, the scores are reasonably close to normally distributed and very suitable.
6. For Board KL, the scores are significantly positively skewed, but this can be corrected by normalisation. The scores do not discriminate as well at the top end of the scale as they do in the middle and lower score ranges. The scores are usable, but would be more useful for selection if there was greater discrimination in the upper score ranges.
7. For Board MR, the scores are slightly positively skewed, but close enough to normally distributed and very suitable.
8. For Board RJ, the scores are close to normally distributed and very suitable.
9. For Board TN the score distribution is highly skewed in the negative direction, indicating that in the very highest score ranges, very little discrimination was made. This can be corrected by normalisation, but the consequence will be that candidates who are not exceptionally high scorers will be moved significantly downwards. The scores are insufficiently precise for us to know whether or not they deserve better than this. This is a disappointing set of scores, and it could be argued that the addition of the Board scores to the AIEEE Total scores would detract from the quality of selection.
10. For Board UP, the scores are close to normally distributed and very suitable.

In summary, for seven of the nine Boards, the Board scores could contribute usefully to selection. For the remaining two (KL and TN), the Board scores are too heavily concentrated in the high score range, which detracts significantly from their usefulness.

Differences by Board

There are substantial differences in average achievement levels by Board, as evidenced in the Means and standard deviations of the AIEEE scores presented in the table that follows.

Table 1. AIEEE Total scores: Comparisons by Board

Board	AIEEE Total Scores			Percent of Candidates in the top:			
	N	Mean	SD	1%	2%	5%	10%
AM	5,764	28.56	28.22	0.05%	0.24%	1.02%	2.91%
BR	40,103	26.01	28.32	0.06%	0.32%	1.21%	2.70%
CBS	182,399	58.24	50.14	1.94%	3.78%	9.07%	17.33%
GU	27,501	37.50	33.19	0.27%	0.53%	1.91%	5.22%
KL	15,819	41.15	33.90	0.16%	0.52%	2.28%	6.83%
MR	70,602	28.83	33.02	0.39%	0.75%	1.85%	3.96%
RJ	17,188	35.84	37.21	0.34%	1.07%	3.23%	7.07%
TN	14,739	32.01	26.81	0.04%	0.14%	0.65%	2.49%
UP	28,628	25.13	26.41	0.07%	0.19%	0.74%	1.82%
All	402,743	43.09	43.36	1.00%	2.00%	5.00%	10.00%

From Table 1 it is apparent that the differences in achievement by Board are too great to ignore. Board scores from different Boards cannot be treated as equivalent.

Scores from Board CBS have a dramatically higher mean, combined with an equally dramatically higher standard deviation than all other boards. The combined effect of these two elements is that there are many more candidates from Board CBS at the highest level of achievement than in any other Board. This is clearly seen by examining the percent of candidates in each Board who score in the top 1%, 2%, 5% and 10% of scores, shown in the last four columns of Table 1. Given that these are the ranges within which many selection decisions are likely to be made, it is important that any selection scores that we generate must preserve these differences.

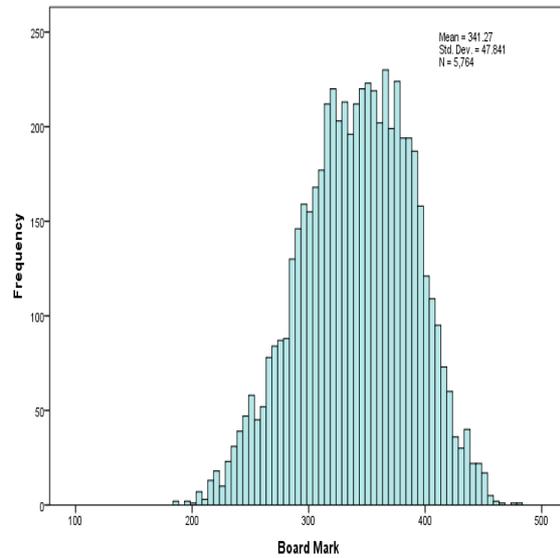
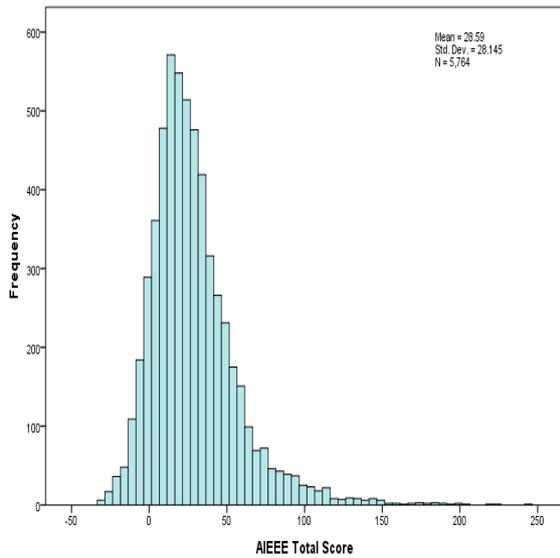
The proposal in this paper is that this be done by transforming each Board's scores (however they have been arrived at) so that they are normally distributed, and to adjust the mean and standard of each set of Board scores to match the mean and standard deviation of that Board's candidates on their AIEEE total scores. In this way we can have confidence that both the AIEEE Total scores and the adjusted Board scores, before they are combined, adequately represent differences within and between the Boards.

Demonstration of the procedures

To demonstrate how the scores would be combined and to show how the score distributions would appear at each stage, I will use the first of the Board data sets provided to me (Board AM) for purposes of demonstration.

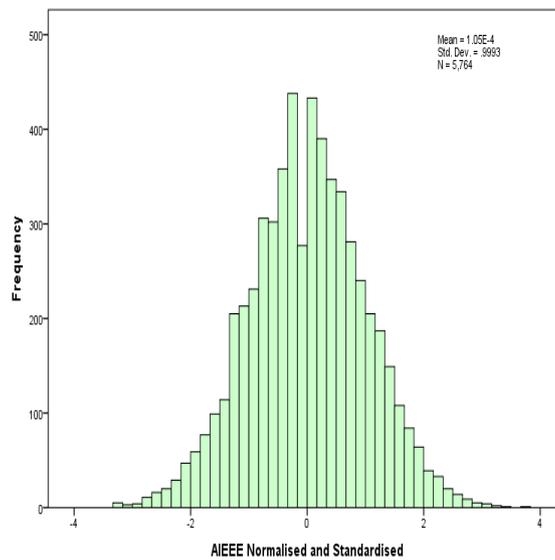
The steps involved are as outlined in the first section.

1. For each Board, the AIEEE and Board data need to be scrutinised carefully, to ensure that the data are free from irregularities and evidence of manipulation. Once satisfied with the quality of the data, apply the procedures that follow within each set of each Board scores.

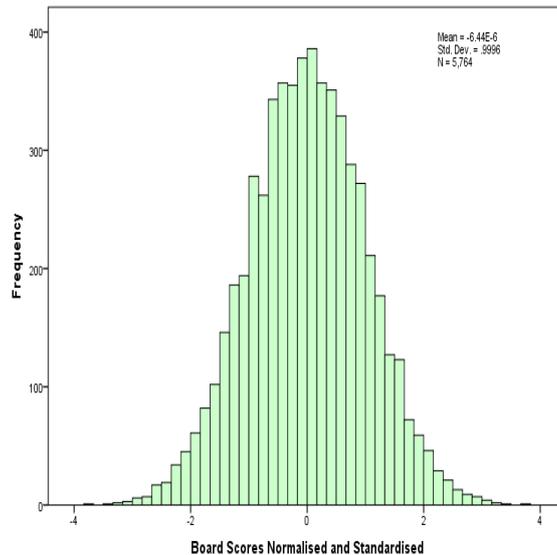


Comment: Both sets of scores are appropriate and can be converted to normal with little change.

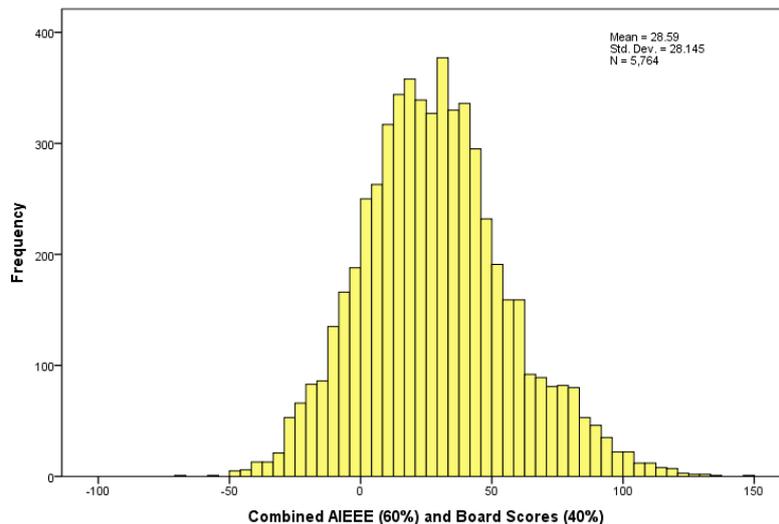
2. Apply a transformation to the AIEEE Total scores (the sum of examination scores in Physics, Chemistry and Mathematics) to convert the distribution from the (typically) skewed raw score distribution to a normal distribution.



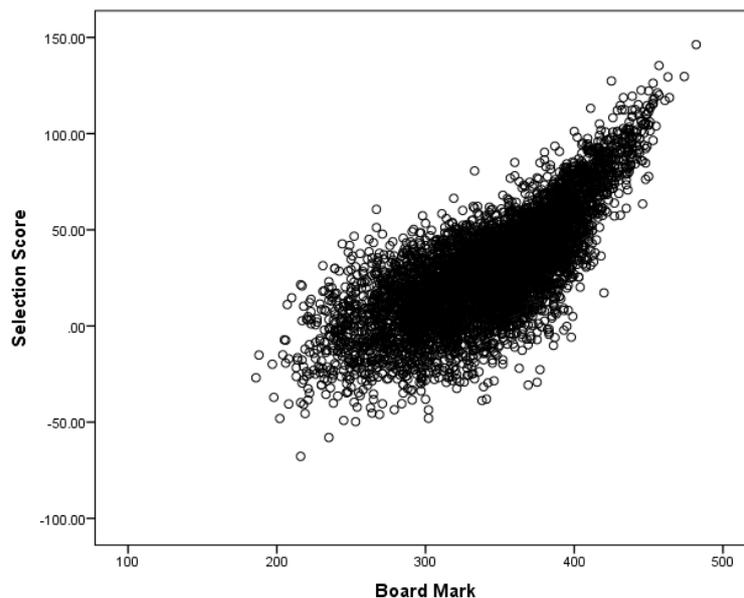
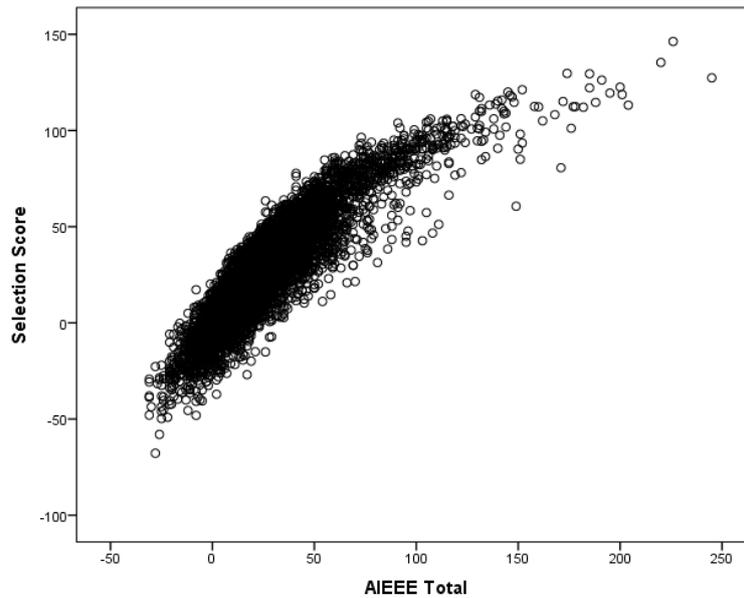
3. Adjust the mean and standard deviation of this normalised distribution to the same mean and standard deviation as that of the original AIEEE Total scores *from that Board*.
4. Apply a transformation to the Board scores to convert the distribution from the (typically) skewed raw score distribution to a normal distribution.



5. Adjust the mean and standard deviation of this normalised distribution to the same mean and standard deviation as that of the original AIEEE Total scores *from that Board*.
6. Form a composite score by combining the transformed AIEE scores and the transformed Board scores in 60:40 proportion (60% AIEEE Total and 40% Board scores).
7. The composite will have the same mean but not the same standard deviation as the original distribution of AIEE scores from that Board, and its distribution will be near to but not quite normal. Normalise the distribution and adjust the mean and standard distribution to match the mean and standard deviation of the AIEEE Total scores *from that Board*:



The final scores used for selection will correlate with both AIEEE scores and Board scores. The correlation with AIEEE scores is likely to be higher than with Board scores because of the higher weight accorded to the AIEEE scores. Scatter plots are shown below.



As anticipated, the selection scores correlate with both AIEEE Total scores and Board score, but more closely with the former.

Conclusion

The processes described are simple and manageable enough to be applied routinely. But the quality of the selection decisions made depends heavily on the quality of the data that are used to facilitate these decisions.

To some extent the process may be self-correcting. If there is a tendency in a Board to award greater numbers of high scores than is warranted, normalisation will adjust most of those high scores downwards, with the unfortunate result that the highest achievers in the Board will be disadvantaged. As this occurs (and particularly if it can be understood and anticipated), the

temptation to inflate the scores of higher achievers quickly soon disappear. This has been the experience when such procedures have been used here.

Data quality is an important issue. There appears to be no problem with the quality of AIEEE data, provided the total scores are used. The quality of Board data appears mixed. There are no major problems in the 2012 data that I have received from seven of the nine Boards. For the remaining two (Boards KL and TN), there is a tendency to push too many candidates into the highest score ranges, with w resulting lack of discrimination where it matters most.

In my view, the stated policy can be implemented if all the data are of the quality of Boards AM, BR, CBS, GU, MR, RJ and UP. It can be implemented with some loss of precision where the data are of lesser quality (as in Boards KL and TN).

Implementation of this policy needs to be accompanied by a concerted effort to provide professional development at the Board level. The professional development should include:

1. Setting examinations to discriminate across the whole achievement range.
2. Developing scoring schemes to achieve discrimination across the whole achievement range.
3. Training examiners to ensure uniform standards of marking throughout the Board.
4. Developing an understanding of the importance of data quality in the selection process and the detriment to data quality when proper procedures are not followed.