# A Rasch analysis of the quality of marking of GCSE Science

William Pointer


## Summary

The reintroduction of an examination that combines biology, chemistry and physics has created the potential for a reduction in the reliability of marking. A Rasch analysis was conducted to explore whether the reliability of marking of science exams was being compromised by having markers mark outside their specialism.

This paper concludes that the differences between specialisms are minor at worst. Therefore there is no pressing need to split the marking of expert marked items by subject specialism. The most important factor in ensuring high marking reliability is likely to be the clarity of the mark scheme and the training of markers. Clear thought and planning should be used to ensure mark schemes are fit for purpose to allow marking to be completed reliably.


## Background

The reliability of marking in high stakes examinations, such as GCSE, is a very important and pertinent issue (see, for example, Jordan, 2014). One of the factors often cited as affecting the reliability of marking is the characteristics of the person marking the response. There is a limited amount of research literature regarding the traits that are necessary in an examiner to ensure a high level of marking reliability. Pinot de Moira (2003) found no link between personal characteristics and marking reliability but did find some evidence to support the notion that the work of more able candidates is harder to mark.

Royal-Dawson and Baird (2009) found that teaching experience was not necessary to reliably mark most of the examination questions on the UK's national curriculum Key Stage 3 English test. They proposed that questions with a high level of curriculum specificity should be marked by teachers but that other items could be marked by non-teachers. In the same subject area, Meadows and Billington (2010) attempted to split out the effects of subject knowledge, teaching experience and marking experience. They found that, overall, GCSE English examiners, trainee English teachers, English undergraduates and undergraduates from other disciplines marked GCSE English equally accurately, although there were some differences at item level.

Suto, Nadas and Bell (2011) found that a marker's highest level of education is a better predictor of marking accuracy than teaching or marking experience in IGCSE biology. This reinforced the findings of Suto and Nadas (2008) who reached the same conclusion for GCSE mathematics and physics. It led them to conclude that "education trumps experience".

Despite the existing research evidence to the contrary, there remains a concern that having examiners mark questions outside their subject specialism has a detrimental effect on the reliability of marking. Take for example GCSE Science A and GCSE Additional Science which are currently offered by AQA with two possible routes of certification; structural differences in the assessment gives rise to questions about marking reliability. Route 1 is assessed via three examined units, one in each of biology, chemistry and physics plus a controlled assessment.

Route 2 is assessed via two examined units that both contain a mix of biology, chemistry and physics questions plus a controlled assessment.

For Route 1, examiners are appointed to mark a particular unit where only one specialism is examined; therefore they can opt to mark their specialist subject area. For Route 2, examiners are appointed to mark a particular unit where all three specialisms are examined and they mark a quota of all items. It could be argued that, for Route 2, having biologists mark chemistry and physics questions, for example, would reduce the reliability of the marking. Although, that said, it is worth noting many teachers actually teach across all three subject areas and so the impact of marking outside of their specialism is likely to be much less pronounced than if they were marking something of which they had no relevant experience or knowledge.

This paper is interested in finding if there are any interactions between the examiner specialism and the subject matter in the item that is being marked. If any such interactions exist then it would suggest that the marking of this examination would be made more reliable by changing the way items are allocated to markers. Currently, all examined units for the GCSE science specifications are marked on-screen. On-screen marking allows each item to be marked separately. Items are split into one of three groups, which determine how that item is marked. Items can be marked automatically, marked by a general marker or by an expert marker. Changes might be made so that expert marked items, which by definition require subject knowledge, are allocated according to specialism rather than to all expert markers, regardless of specialism.

The paper focuses on the higher tier as there is a greater preponderance of expert marked items compared with the foundation tier. In particular it looks at the SCA2HP examination from summer 2013 as it had a large entry and therefore there were more examiners to study.

The reliability of the marking can be derived from operational data obtained about the seeds. Seeds are items that have been pre-marked by a senior examiner and are used as a method of quality control for the marking of the GCSE science units. By treating the seed mark as the ‘true’ mark we can compare the mark given by individual examiners to the seed mark to judge the reliability of their marking.

## Methodology

There were 68 expert markers for SCA2HP in summer 2013. For the analysis, the examiners were split into their respective specialisms, i.e. biology, chemistry or physics. The allocation of specialism to examiner was done on the basis of the units that the examiner had marked in the past. If the examiner had marked biology units in the past then he/she was designated a biologist. It was unclear what specialism two of the examiners had and so they were excluded from the analysis. This left 33 biologists, 17 chemists and 16 physicists.

For the live marking period, a total of 1,247 seeds were created and used to monitor the expert markers for SCA2HP. Since the seeds are presented at random, each seed was not marked by every examiner; on average each seed was marked by 42 different examiners. Seeds that were seen by less than half (33) of the examiners were excluded.

Each seed was tagged to the item number and the subject area that it was assessing. Some examiners marked the same seed more than once; in these cases the mark awarded by the examiner the first time the seed was presented was used. For each examiner and for each seed the absolute difference between the mark given by the examiner and the mark awarded to the seed was calculated.

A number of seeds were marked extremely accurately. The marks given for 723 seeds were in complete agreement with the ‘true’ mark. Seeds that do not have much variation will not throw

any light on differences between examiners and so seeds with a standard deviation less than 0.35 were excluded. This left a total of 125 seeds to analyse.

By focusing on the most discriminating seeds, in other words the seeds with the greatest variance, the model was designed to exaggerate any differences between the examiners as the majority of seeds were marked well by all examiners regardless of their specialism.

The partial credit model (PCM; Masters, 1982) was used to describe the ability of the examiner and the difficulty of the seed. This type of Rasch analysis has been used in the past to explore the performance of specific examinations (see Taylor, 2010) and has been discussed in detail in He & Wheadon (2008). One of the advantages of using the Rasch model is that it puts the two variables, examiner ability and difficulty to mark a seed, on the same scale. It is also very good at dealing with missing data, which is quite prevalent in this data set. The implementation of the PCM for this data is slightly different from standard examination analysis; instead of looking at the number of marks a student gained on different items, it looks at the difference between the 'true' mark and the mark awarded by the examiner. Consequently, an examiner with a high measure has a low performance and items with a high measure were easier to mark.

Winsteps (see http://winsteps.com) was used to fit the PCM.

## Results

### Differential group functioning (DGF)

The seeds were grouped by subject and the examiners by specialism and then the differential group functioning (DGF) size of the groups was calculated. DGF is a measure of the relative difficulty and can be used to show if a group of examiners is performing at their usual ability on a set of items. If the DGF is zero then the examiners are performing at their usual ability, if it is positive then they are marking better than their usual ability and if it is negative then they are marking worse. Linacre (2013) says that DGF should be greater than 0.5 logits to be noticeable.

Table 1 shows the DGF size in logits and the approximate standard error (SE) for each subject-specialism pairing. The DGF t is the approximate Student's t-statistic test, which is estimated as $\frac{\text{DGF size}}{\text{DGF SE}}$ with (the number of observations – 2) degrees of freedom, and Prob. is the probability of the t-value. The question being investigated is:

> Are examiners with a particular specialism better or worse at marking a given subject area when compared with their overall reliability of marking?

The table shows very little variability in the marking of the biologists; their performance is as expected for all three subjects. In fact the highest DGF for the biologists was for the chemistry questions, but this result is not significant. The chemists and physicists both show slightly more imbalance, with seemingly better performance in their area of expertise. However, none of the values are significant[1], even without applying any false discovery rate control (FDR) to allow for the fact that multiple hypothesis tests were being performed. Therefore it is not possible to conclude that any group of specialists are performing better or worse than expected in any particular subject.

[1] with a significance level α of 0.05

**Table 1 The differential group functioning (DGF) size in logits for each subject by specialism**

| Specialism | Subject | Number of observations | DGF size (in logits) | Standard error (in logits) | DGF t | Prob. |
|---|---|---|---|---|---|---|
| Biology | Biology | 1045 | 0.00 | 0.06 | 0.00 | 1.00 |
| Biology | Chemistry | 575 | 0.06 | 0.10 | -0.57 | 0.57 |
| Biology | Physics | 911 | 0.00 | 0.08 | 0.00 | 1.00 |
| Chemistry | Biology | 593 | 0.03 | 0.08 | -0.39 | 0.70 |
| Chemistry | Chemistry | 322 | 0.17 | 0.14 | -1.19 | 0.24 |
| Chemistry | Physics | 480 | -0.13 | 0.10 | 1.33 | 0.18 |
| Physics | Biology | 566 | 0.00 | 0.08 | 0.00 | 1.00 |
| Physics | Chemistry | 294 | -0.22 | 0.12 | 1.83 | 0.07 |
| Physics | Physics | 452 | 0.16 | 0.11 | -1.48 | 0.14 |

It is also possible to make pairwise comparisons, either by item subject or by examiner specialism and look at the DGF contrast, which is the difference between the DGF sizes. Table 2 shows the DGF contrast for item subjects. The question under investigation shifts to:

> Is there a group of specialists whose marking of one subject is better or worse than their marking of either of the other subjects?

The greatest DGF contrast was exhibited by the physicists, where the DGF contrast between the chemistry items and the physics items is -0.39. This could indicate that the physicists are better at marking the physics questions over the chemistry questions. The probability of obtaining the t-value for this example was 0.0191. Since nine hypothesis tests are being carried out, FDR control[2] is needed when looking for significance. This gives a critical value of 0.0056 and so the result is not significant. Therefore, there is no evidence to suggest differing performance by examiner specialism on the different subjects.

**Table 2 DGF contrast in logits for subjects**

| Subject | Subject | DGF contrast (in logits) | Joint SE (in logits) | t | d.f. | Prob. | Specialism |
|---|---|---|---|---|---|---|---|
| Biology | Chemistry | -0.06 | 0.12 | -0.49 | INF | 0.63 | Biology |
| Biology | Physics | 0.00 | 0.10 | 0.00 | INF | 1.00 | Biology |
| Biology | Chemistry | -0.14 | 0.16 | -0.84 | 670 | 0.40 | Chemistry |
| Biology | Physics | 0.16 | 0.13 | 1.28 | INF | 0.20 | Chemistry |
| Biology | Chemistry | 0.22 | 0.15 | 1.53 | 657 | 0.13 | Physics |
| Biology | Physics | -0.16 | -0.14 | 1.19 | 947 | 0.23 | Physics |
| Chemistry | Physics | 0.06 | 0.12 | 0.45 | INF | 0.65 | Biology |
| Chemistry | Physics | 0.30 | 0.17 | 1.74 | 701 | 0.08 | Chemistry |
| Chemistry | Physics | -0.39 | -0.16 | 2.35 | 692 | 0.02 | Physics |

The table of the DGF contrast for examiner specialisms is shown in Appendix 1.

[2] Benjamini-Hochberg procedure: $\alpha_k = k*0.05/9$

**Probabilities**

Logits are on an arbitrary scale and so it is difficult to clearly understand what the differences mean in the real world. To help contextualise the results, it is possible to convert logits into probabilities. This allows us to report the probability that the different sets of examiners will mark a particular item correctly. For simplicity, take the case of a dichotomous question:

$$P(\theta, d) = \frac{\exp(d\ -\theta)}{1 + \exp(d\ -\theta)} \qquad (1)$$

P(θ,d) is the probability that an examiner with ability θ can mark an item with difficulty d correctly. This is analogous to the standard case (see Rasch, 1960; Wright & Stone, 1979).

The PCM sets a scale such that the average difficulty of the items is zero. Therefore, an item that is of average difficulty has d = 0, which when substituted into equation (1) yields:

$$P(\theta, 0) = \frac{\exp(-\theta)}{1 + \exp(-\theta)} \qquad (2)$$

The average ability of all of the examiners is -2.64. Since a low score represents good performance, this means that the 'test' was easy for the examiners, this is good news as it means the examiners were, on the whole, able to mark accurately. If the average ability of the examiners was zero then, using equation (2) with θ=0, they would have a probability of only 0.5 of awarding the correct mark to a seed of average difficulty, this would mean the marking reliability was very low. With an average ability of -2.64, using equation (2), there is a 0.933 chance that a seed of average difficulty will be marked correctly.

Earlier it was noted that the greatest difference within specialisms was by the physicists when we compared their performance on the chemistry and physics questions. The difference was 0.39 logits; the logit scale is relative but does not give a sense of what this means in the real world. If we compute the probabilities that the physicists will mark a dichotomous chemistry question and a dichotomous physics question (both of average difficulty) correctly then we can make a meaningful comparison.

By computing the difference of P(θ,0) for the physicists on an average chemistry question (i.e. 0.22 logits worse than average) and an average physics question (i.e. 0.16 logits better than average),

$$P(2.64 - 0.22,0) - P(2.64 + 0.16,0) = P(2.42,0) - P(2.80,0) = 0.918 - 0.943 = -0.024 \qquad (3)$$

we find that the physicists are 2% less likely to award the 'true' mark to a one-mark chemistry item compared to a physics item.

Figure 1 shows the probabilities that an examiner will mark a dichotomous item of average difficulty correctly for the three subject areas by specialism. It shows the difference of 0.024 between chemistry and physics calculated in equation (3). It also shows that all other differences are very small. The error bars show the 95% confidence intervals (DGF ± 1.96 × SE); all of the confidence intervals overlap.

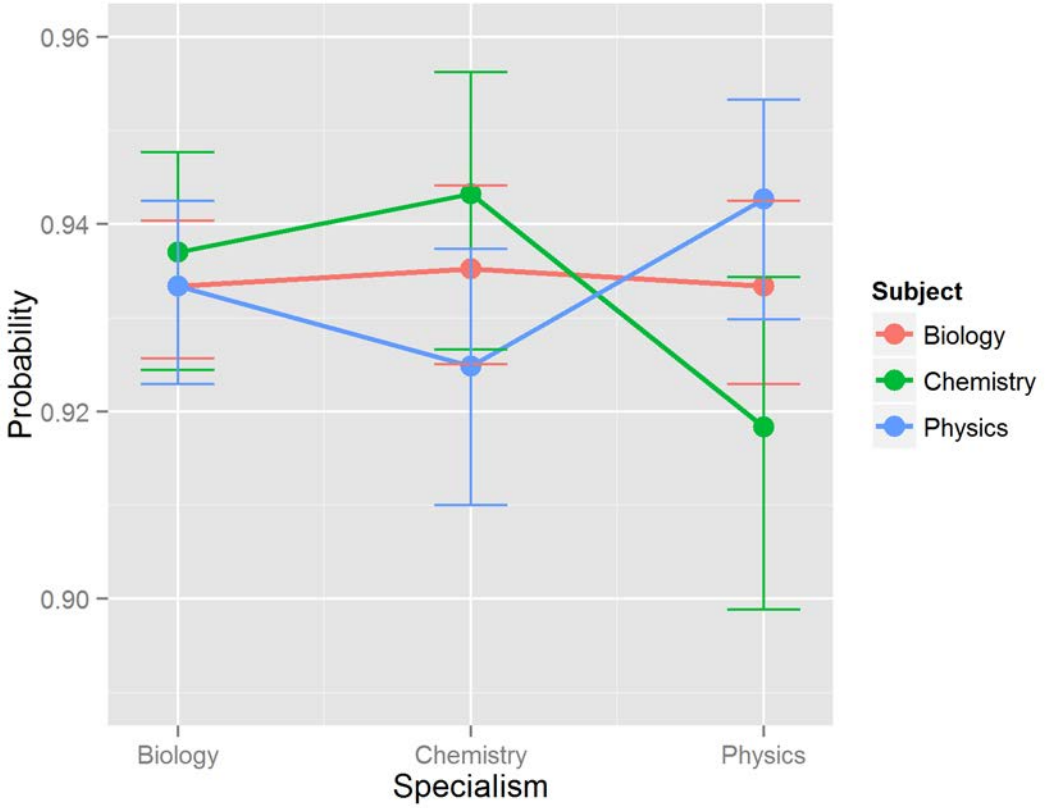


**Figure 1** **The probability that an examiner will correctly mark a dichotomous question of average difficulty based on their specialism.**

**Examiners**

The analysis can be done at examiner level, to identify whether particular examiners are marking a particular subject area badly relative to their ability to mark in general, as well as identifying the examiners who are the most unreliable. This could help in identifying which examiners to reappoint and which examiners may be best redeployed to a different unit, perhaps one in which there is only one specialism to mark.

The PCM assigns a value to represent the ability of each examiner. As previously stated the average was -2.64 logits (the lower the value the better the examiner). The person measures have not been included in this paper but could be used in conjunction with other available data before making any decisions on re-hiring or not. However, care should be taken as the number of seeds marked by each examiner for each subject is generally quite small, often less than 30, and so the results are less dependable.

The DGF of the examiners by subject is shown in Appendix 2. Using the Student's t-statistic test, and applying false discovery rate (FDR) control with n=196 there are two examiners with statistically significant p values. However, these examiners had only marked two items in the subject that was deemed significant and so there is insufficient data to draw any conclusions. If no FDR control is applied, then three more examiners would be statistically significant. Two of the examiners had a large, positive DGF, which means that they were marking that subject particularly well; unsurprisingly it was their own subject specialisms that were marked particularly well. The other examiner had a DGF size of -0.91 on the physics questions, her specialism was biology. This means that this examiner's performance on the physics items was substantially worse than expected compared to her overall performance. It may well be that this constitutes sufficient evidence to move this examiner to another unit. Such a decision should

not be made on the basis of this evidence alone as the result was only significant without applying any control for the fact that multiple hypothesis tests were being performed. It is worth noting that this examiner's score was -2.79, which is slightly better than the average of -2.64. So there is no reason for her not to mark, just that a combined paper may not be the most appropriate setting for her.

**Items**

In previous research, it has been shown that there is no noticeable overall difference between the reliability of marking for markers with different backgrounds. However, differences have been found in the reliability of marking of particular questions (see, for example, Meadows and Billington, 2010). Therefore it seems prudent to compare the groups of specialists at an item level.

By grouping the different seeds into the item to which they referred, the PCM can also be used to show if any items showed any discrepancies between the different specialisms. Again, care has to be taken when interpreting results as the number of observations is sometimes small. The DGF for the items by examiner specialism is shown in Appendix 3. Using the Student's t-statistic test, and applying false discovery rate (FDR) control with n=93 there are no items that have statistically significant p values. If a less conservative approach is taken and FDR is not applied then three items have significant p values. Question 2bi, a biology question, was marked better than expected by the biologists and worse than expected by the chemists. Questions 3aii and 13aii were marked less well than expected by the physicists. This paper does not attempt to explain why such variation occurs, just to highlight any potential issues. Further investigation, involving a scrutiny of the mark scheme, may be needed to try and find a cause for this discrepancy.

Six of the seven hardest to mark seeds were from question 2bii. This question had the highest mark tariff on the paper and also included the assessment of quality of written communication; it is also the only question on the paper that used a level of response mark scheme. This question had the second highest seed failure rate[3] despite having a tolerance of one mark. The seed failure rate provides a good metric when looking for items that suffered from poor marking reliability. The DGF size for each of the three groups of specialists was approximately zero for this item, i.e. they marked it to the same level as their marking as a whole. So despite being hard to mark, it did not seem to cause any particular group of specialists any more problems than another.

## Discussion and Conclusion

The Rasch model has proved to be a useful tool in analysing seed data to look at marking reliability. The model and the subsequent analysis undertaken for this paper was on the (125) most discriminating seeds, this meant that the vast majority of the seeds, where there was a large degree of alignment of marking from examiners of all specialisms, were ignored. One effect of this was that 10 questions were not represented at all, leaving 31 (out of a total of 41 expert items) to be analysed.

Considering this analysis was based on the 125 seeds that were hardest to mark and that it does not show any clear difference between specialisms, it does not appear that there is any pressing need to restrict marking to a particular specialism. This could be because the markers have opted to mark this paper and therefore are confident in marking all three subjects. With the move to terminal exams in 2014, there will be a need for more examiners. It may not be

[3] The seed failure rate is the percentage of seeds that are marked outside of a pre-determined tolerance.

possible to extend the marking panel to include only examiners with the capability to mark all three subject areas to a good standard to ensure the reliability of marking is not compromised. Therefore it may be prudent to continue to monitor the reliability of marking for these units (SCA1, SCA2, AS1 and AS2) closely.

This paper has only looked at the output of the marking; factors such as examiners taking more time or care on areas which are unfamiliar have not been taken into account. These factors could result in an increased risk of marking not being completed in time.

The most important factor in ensuring marking is reliable is more likely to be the clarity of the mark scheme and the training given to markers at standardisation. This is where the focus of any improvements should be. Identifying items with a high seed failure rate may help to target particular types of question where mark scheme design could be further improved.

**Limitations**

The mark awarded to the seed was treated as the 'true' mark; if the 'wrong' mark was incorrectly awarded when the seed was created then this could give misleading results. It might have been advantageous to conduct the analysis using a different measure of the 'true' mark, for example the modal mark awarded to the seed during the marking period. However, Meadows and Billington (2010) did not find using a second measure of the correct mark made any difference to their conclusions.

## References

He, Q. & Wheadon, C. (2012). Using the dichotomous Rasch model to analyse polytomous items. *Journal of applied measurement, 14*(1), 44–56.

Jordan, C. (2014, January 4). We need to end the annual exam marking scandal. *The Telegraph*. Retrieved February 11, 2014, from http://www.telegraph.co.uk/education/educationopinion/10550133/We-need-to-end-the-annual-exam-marking-scandal.html

Linacre, J. (2013). *A user's guide to Winsteps.* winsteps.com. 3.80.0, 408

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Meadows, M. & Billington, L. (2005). *A Review of the Literature on Marking Reliability*. Report for the National Assessment Agency by AQA Centre for Education Research and Policy.

Meadows, M. & Billington, L. (2010). *The effect of marker background and training on the quality of marking in GCSE English.* RPA_10_MM_RP_028. Manchester, UK: AQA Centre for Education Research and Policy.

Pinot de Moira, A. (2003). *Examiner background and the effect on marking reliability*. RC218. Manchester, UK: AQA Centre for Education Research and Policy.

Royal-Dawson, L. & Baird, J. (2009). Is teaching experience necessary for reliable scoring of extended English questions? *Educational Measurement: Issues and Practice*, *28*(2), 2–8.

Suto, I. & Nadas, R. (2008). What determines GCSE marking accuracy? An exploration of expertise among maths and physics markers. *Research Papers in Education, 23*(4), 477–497.

Suto, I., Nadas, R., and Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education, 26*(1), 21–52.

Taylor, R. (2010). *A Rasch analysis of GCE Biology (BIOL4)*, January 2010. RPA_10_RT_TR_036. Manchester, UK: AQA Centre for Education Research and Policy.

Wright, B.D. and Stone, M.H. (1979). *Best Test Design. Rasch Measurement.* Chicago, IL: MESA Press.

# Appendix

## Appendix 1 DGF contrasts for examiner specialisms

Question: Is there a subject that is marked better or worse by a particular group of examiners?

| Specialism | Specialism | DGF contrast | Joint SE | t | d.f. | Prob. | Subject |
|---|---|---|---|---|---|---|---|
| B | C | -0.03 | 0.10 | -0.31 | INF | 0.76 | B |
| B | P | 0.00 | 0.10 | 0.00 | INF | 1.00 | B |
| B | C | -0.11 | 0.17 | -0.64 | 730 | 0.52 | C |
| B | P | 0.28 | 0.16 | 1.78 | 718 | 0.08 | C |
| B | C | 0.13 | 0.12 | 1.06 | INF | 0.29 | P |
| B | P | -0.16 | -0.13 | 1.22 | INF | 0.22 | P |
| C | P | 0.03 | 0.11 | 0.27 | INF | 0.78 | B |
| C | P | 0.39 | 0.19 | 2.10 | 613 | 0.04 | C |
| C | P | -0.29 | -0.15 | 1.99 | 923 | 0.05 | P |

There are no significant results and so there is no evidence that any of the subjects are marked better or worse by a particular group of examiners.

**Appendix 2 DGF by examiner and subject**

| Examiner | Count | Average | Expected | DGF score | DGF size | DGF SE | DGF t | Prob. | Subject |
|---|---|---|---|---|---|---|---|---|---|
| B301313 | 43 | 0.14 | 0.17 | -0.03 | 0.22 | 0.43 | -0.52 | 0.61 | B |
| B301313 | 15 | 0.20 | 0.12 | 0.08 | -0.57 | 0.59 | 0.97 | 0.35 | C |
| B301313 | 29 | 0.10 | 0.10 | 0.00 | -0.02 | 0.60 | 0.04 | 0.97 | P |
| B565986 | 34 | 0.18 | 0.24 | -0.07 | 0.38 | 0.43 | -0.87 | 0.39 | B |
| B565986 | 19 | 0.16 | 0.13 | 0.03 | -0.24 | 0.61 | 0.40 | 0.70 | C |
| B565986 | 30 | 0.20 | 0.15 | 0.05 | -0.35 | 0.43 | 0.82 | 0.42 | P |
| B567789 | 44 | 0.41 | 0.31 | 0.10 | -0.38 | 0.27 | 1.43 | 0.16 | B |
| B567789 | 18 | 0.06 | 0.19 | -0.13 | 1.23 | 0.82 | -1.50 | 0.15 | C |
| B567789 | 36 | 0.14 | 0.20 | -0.06 | 0.43 | 0.46 | -0.93 | 0.36 | P |
| B567798 | 43 | 0.51 | 0.45 | 0.07 | -0.19 | 0.25 | 0.75 | 0.45 | B |
| B567798 | 23 | 0.39 | 0.30 | 0.09 | -0.32 | 0.38 | 0.86 | 0.40 | C |
| B567798 | 32 | 0.19 | 0.34 | -0.15 | 0.72 | 0.42 | -1.74 | 0.09 | P |
| B567874 | 35 | 0.17 | 0.19 | -0.01 | 0.09 | 0.44 | -0.21 | 0.83 | B |
| B567874 | 20 | 0.10 | 0.11 | -0.01 | 0.14 | 0.74 | -0.19 | 0.85 | C |
| B567874 | 26 | 0.15 | 0.13 | 0.03 | -0.22 | 0.53 | 0.42 | 0.68 | P |
| B567911 | 38 | 0.26 | 0.29 | -0.03 | 0.12 | 0.35 | -0.34 | 0.74 | B |
| B567911 | 22 | 0.23 | 0.17 | 0.06 | -0.33 | 0.48 | 0.69 | 0.50 | C |
| B567911 | 31 | 0.19 | 0.20 | -0.01 | 0.06 | 0.44 | -0.13 | 0.90 | P |
| B567939 | 46 | 0.28 | 0.29 | -0.01 | 0.03 | 0.31 | -0.10 | 0.92 | B |
| B567939 | 23 | 0.04 | 0.18 | -0.14 | 1.39 | 0.75 | -1.85 | 0.08 | C |
| B567939 | 40 | 0.28 | 0.19 | 0.09 | -0.45 | 0.33 | 1.38 | 0.17 | P |
| B568225 | 32 | 0.63 | 0.53 | 0.10 | -0.23 | 0.26 | 0.87 | 0.39 | B |
| B568225 | 22 | 0.27 | 0.36 | -0.09 | 0.36 | 0.45 | -0.80 | 0.43 | C |
| B568225 | 30 | 0.33 | 0.37 | -0.04 | 0.14 | 0.36 | -0.40 | 0.69 | P |
| B568589 | 36 | 0.44 | 0.39 | 0.05 | -0.17 | 0.29 | 0.59 | 0.56 | B |
| B568589 | 16 | 0.31 | 0.24 | 0.07 | -0.32 | 0.50 | 0.64 | 0.53 | C |
| B568589 | 26 | 0.15 | 0.27 | -0.12 | 0.67 | 0.50 | -1.33 | 0.19 | P |
| B573473 | 30 | 0.50 | 0.59 | -0.09 | 0.22 | 0.30 | -0.72 | 0.48 | B |
| B573473 | 23 | 0.43 | 0.39 | 0.05 | -0.15 | 0.36 | 0.41 | 0.69 | C |
| B573473 | 33 | 0.45 | 0.41 | 0.04 | -0.13 | 0.30 | 0.44 | 0.66 | P |
| B573478 | 40 | 0.17 | 0.29 | -0.11 | 0.59 | 0.39 | -1.49 | 0.14 | B |
| B573478 | 18 | 0.11 | 0.20 | -0.09 | 0.64 | 0.70 | -0.91 | 0.38 | C |
| B573478 | 32 | 0.34 | 0.16 | 0.18 | -0.91 | 0.33 | 2.76 | 0.01 | P |
| B573748 | 46 | 0.28 | 0.32 | -0.04 | 0.16 | 0.31 | -0.52 | 0.61 | B |
| B573748 | 20 | 0.35 | 0.21 | 0.14 | -0.63 | 0.41 | 1.55 | 0.14 | C |
| B573748 | 38 | 0.18 | 0.21 | -0.03 | 0.17 | 0.40 | -0.43 | 0.67 | P |
| B573774 | 20 | 0.05 | 0.11 | -0.06 | 0.86 | 0.91 | -0.95 | 0.36 | B |
| B573774 | 5 | 0.00 | 0.09 | -0.09 | 0.21 | < 1.68 | -0.12 | 0.91 | C |
| B573774 | 18 | 0.17 | 0.07 | 0.09 | -0.92 | 0.60 | 1.54 | 0.14 | P |
| B575685 | 34 | 0.32 | 0.29 | 0.04 | -0.16 | 0.34 | 0.46 | 0.65 | B |
| B575685 | 18 | 0.06 | 0.17 | -0.12 | 1.14 | 0.84 | -1.37 | 0.19 | C |
| B575685 | 30 | 0.20 | 0.18 | 0.02 | -0.15 | 0.44 | 0.34 | 0.73 | P |
| B579075 | 37 | 0.24 | 0.27 | -0.03 | 0.12 | 0.37 | -0.33 | 0.74 | B |
| B579075 | 19 | 0.21 | 0.20 | 0.01 | -0.06 | 0.54 | 0.11 | 0.91 | C |
| B579075 | 28 | 0.21 | 0.19 | 0.02 | -0.14 | 0.44 | 0.31 | 0.76 | P |
| B579507 | 34 | 0.56 | 0.41 | 0.15 | -0.43 | 0.26 | 1.66 | 0.11 | B |
| B579507 | 15 | 0.07 | 0.22 | -0.15 | 1.23 | 0.83 | -1.48 | 0.16 | C |
| B579507 | 28 | 0.14 | 0.25 | -0.11 | 0.64 | 0.50 | -1.28 | 0.21 | P |
| B580360 | 2 | 0.00 | 0.15 | -0.15 | 0.00 | < 1.95 | 0.00 | 0.00 | B |
| B580360 | 5 | 0.20 | 0.22 | -0.02 | 0.09 | 1.08 | -0.08 | 0.94 | C |
| B580360 | 19 | 0.26 | 0.24 | 0.02 | -0.09 | 0.49 | 0.18 | 0.86 | P |
| B580368 | 34 | 0.35 | 0.34 | 0.01 | -0.05 | 0.33 | 0.16 | 0.87 | B |
| B580368 | 16 | 0.25 | 0.25 | 0.00 | 0.00 | 0.55 | 0.00 | 1.00 | C |
| B580368 | 32 | 0.22 | 0.24 | -0.02 | 0.10 | 0.41 | -0.24 | 0.81 | P |
| B580532 | 2 | 0.00 | 0.12 | -0.12 | 0.00 | < 2.19 | 0.00 | 0.00 | B |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| B580532 | 6 | 0.33 | 0.15 | 0.18 | -0.97 | 0.79 | 1.23 | 0.29 | C |
| B580532 | 8 | 0.00 | 0.11 | -0.11 | 0.80 | < 1.43 | -0.56 | 0.60 | P |
| B580845 | 38 | 0.39 | 0.34 | 0.05 | -0.19 | 0.30 | 0.64 | 0.53 | B |
| B580845 | 13 | 0.15 | 0.21 | -0.06 | 0.38 | 0.74 | -0.51 | 0.62 | C |
| B580845 | 36 | 0.19 | 0.23 | -0.04 | 0.20 | 0.41 | -0.50 | 0.62 | P |
| B581128 | 33 | 0.67 | 0.54 | 0.13 | -0.30 | 0.25 | 1.21 | 0.24 | B |
| B581128 | 22 | 0.27 | 0.37 | -0.10 | 0.40 | 0.45 | -0.89 | 0.39 | C |
| B581128 | 26 | 0.31 | 0.39 | -0.08 | 0.31 | 0.39 | -0.79 | 0.44 | P |
| B581326 | 10 | 0.10 | 0.08 | 0.02 | -0.30 | 1.07 | 0.28 | 0.79 | B |
| B581326 | 11 | 0.09 | 0.05 | 0.04 | -0.52 | 1.10 | 0.47 | 0.65 | C |
| B581326 | 9 | 0.00 | 0.07 | -0.07 | 0.52 | < 1.49 | -0.35 | 0.74 | P |
| B582475 | 30 | 0.27 | 0.32 | -0.06 | 0.25 | 0.39 | -0.63 | 0.53 | B |
| B582475 | 23 | 0.22 | 0.23 | -0.01 | 0.04 | 0.49 | -0.09 | 0.93 | C |
| B582475 | 30 | 0.30 | 0.24 | 0.06 | -0.29 | 0.37 | 0.77 | 0.45 | P |
| B584385 | 27 | 0.30 | 0.30 | -0.01 | 0.03 | 0.40 | -0.09 | 0.93 | B |
| B584385 | 19 | 0.21 | 0.21 | 0.00 | 0.00 | 0.54 | 0.00 | 1.00 | C |
| B584385 | 32 | 0.22 | 0.21 | 0.01 | -0.03 | 0.41 | 0.08 | 0.93 | P |
| B584842 | 3 | 0.00 | 0.03 | -0.03 | 0.00 | < 3.65 | 0.00 | 1.00 | B |
| B584842 | 6 | 0.00 | 0.04 | -0.04 | 0.00 | < 2.15 | 0.00 | 1.00 | C |
| B584889 | 40 | 0.43 | 0.42 | 0.01 | -0.03 | 0.28 | 0.10 | 0.92 | B |
| B584889 | 20 | 0.15 | 0.28 | -0.13 | 0.72 | 0.58 | -1.24 | 0.23 | C |
| B584889 | 28 | 0.36 | 0.28 | 0.07 | -0.29 | 0.36 | 0.83 | 0.42 | P |
| B592283 | 41 | 0.34 | 0.35 | 0.00 | 0.00 | 0.30 | 0.00 | 1.00 | B |
| B592283 | 20 | 0.35 | 0.22 | 0.13 | -0.57 | 0.41 | 1.37 | 0.19 | C |
| B592283 | 34 | 0.15 | 0.22 | -0.07 | 0.46 | 0.46 | -1.00 | 0.32 | P |
| B592361 | 32 | 0.16 | 0.20 | -0.05 | 0.30 | 0.47 | -0.64 | 0.53 | B |
| B592361 | 20 | 0.15 | 0.13 | 0.02 | -0.18 | 0.61 | 0.29 | 0.77 | C |
| B592361 | 25 | 0.20 | 0.16 | 0.04 | -0.26 | 0.48 | 0.53 | 0.60 | P |
| B592584 | 35 | 0.43 | 0.34 | 0.09 | -0.31 | 0.30 | 1.05 | 0.30 | B |
| B592584 | 18 | 0.17 | 0.22 | -0.05 | 0.31 | 0.61 | -0.50 | 0.62 | C |
| B592584 | 29 | 0.14 | 0.22 | -0.08 | 0.52 | 0.51 | -1.02 | 0.32 | P |
| B706392 | 36 | 0.31 | 0.28 | 0.03 | -0.12 | 0.34 | 0.34 | 0.73 | B |
| B706392 | 22 | 0.23 | 0.17 | 0.05 | -0.32 | 0.48 | 0.66 | 0.52 | C |
| B706392 | 33 | 0.12 | 0.19 | -0.07 | 0.50 | 0.51 | -0.98 | 0.34 | P |
| B709623 | 15 | 0.33 | 0.49 | -0.16 | 0.50 | 0.49 | -1.01 | 0.33 | B |
| B709623 | 15 | 0.27 | 0.33 | -0.06 | 0.26 | 0.55 | -0.48 | 0.64 | C |
| B709623 | 18 | 0.56 | 0.38 | 0.18 | -0.53 | 0.36 | 1.46 | 0.16 | P |
| B713789 | 38 | 0.61 | 0.62 | -0.02 | 0.04 | 0.24 | -0.15 | 0.88 | B |
| B713789 | 19 | 0.37 | 0.41 | -0.04 | 0.15 | 0.43 | -0.34 | 0.74 | C |
| B713789 | 29 | 0.45 | 0.40 | 0.05 | -0.15 | 0.32 | 0.46 | 0.65 | P |
| B713806 | 37 | 0.22 | 0.31 | -0.09 | 0.44 | 0.38 | -1.15 | 0.26 | B |
| B713806 | 24 | 0.25 | 0.20 | 0.05 | -0.24 | 0.45 | 0.54 | 0.59 | C |
| B713806 | 36 | 0.28 | 0.22 | 0.06 | -0.31 | 0.35 | 0.89 | 0.38 | P |
| C301281 | 30 | 0.40 | 0.43 | -0.03 | 0.11 | 0.33 | -0.32 | 0.75 | B |
| C301281 | 16 | 0.13 | 0.23 | -0.11 | 0.69 | 0.70 | -0.99 | 0.34 | C |
| C301281 | 35 | 0.34 | 0.27 | 0.07 | -0.31 | 0.32 | 0.95 | 0.35 | P |
| C301379 | 33 | 0.33 | 0.31 | 0.03 | -0.11 | 0.34 | 0.32 | 0.75 | B |
| C301379 | 22 | 0.23 | 0.24 | -0.01 | 0.05 | 0.49 | -0.11 | 0.91 | C |
| C301379 | 30 | 0.20 | 0.22 | -0.02 | 0.14 | 0.44 | -0.31 | 0.76 | P |
| C301812 | 27 | 0.22 | 0.18 | 0.04 | -0.23 | 0.44 | 0.51 | 0.61 | B |
| C301812 | 19 | 0.11 | 0.13 | -0.03 | 0.24 | 0.73 | -0.33 | 0.74 | C |
| C301812 | 27 | 0.11 | 0.13 | -0.02 | 0.20 | 0.60 | -0.33 | 0.74 | P |
| C567551 | 41 | 0.41 | 0.47 | -0.06 | 0.17 | 0.28 | -0.62 | 0.54 | B |
| C567551 | 19 | 0.16 | 0.29 | -0.13 | 0.69 | 0.58 | -1.19 | 0.25 | C |
| C567551 | 28 | 0.50 | 0.33 | 0.17 | -0.54 | 0.30 | 1.77 | 0.09 | P |
| C567966 | 42 | 0.29 | 0.36 | -0.08 | 0.30 | 0.32 | -0.94 | 0.35 | B |
| C567966 | 22 | 0.23 | 0.23 | 0.00 | 0.00 | 0.49 | 0.00 | 1.00 | C |
| C567966 | 35 | 0.34 | 0.25 | 0.09 | -0.38 | 0.32 | 1.18 | 0.25 | P |
| C568249 | 30 | 0.17 | 0.14 | 0.03 | -0.23 | 0.48 | 0.47 | 0.64 | B |
| C568249 | 21 | 0.10 | 0.08 | 0.01 | -0.17 | 0.73 | 0.23 | 0.82 | C |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C568249 | 35 | 0.06 | 0.09 | -0.03 | 0.50 | 0.69 | -0.73 | 0.47 | P |
| C572593 | 32 | 0.16 | 0.18 | -0.02 | 0.14 | 0.48 | -0.29 | 0.77 | B |
| C572593 | 20 | 0.20 | 0.10 | 0.10 | -0.84 | 0.51 | 1.65 | 0.12 | C |
| C572593 | 32 | 0.06 | 0.11 | -0.05 | 0.59 | 0.68 | -0.87 | 0.39 | P |
| C573099 | 41 | 0.56 | 0.48 | 0.08 | -0.21 | 0.24 | 0.89 | 0.38 | B |
| C573099 | 17 | 0.29 | 0.29 | 0.01 | -0.03 | 0.50 | 0.05 | 0.96 | C |
| C573099 | 24 | 0.21 | 0.36 | -0.15 | 0.66 | 0.46 | -1.42 | 0.17 | P |
| C573415 | 33 | 0.21 | 0.30 | -0.09 | 0.45 | 0.41 | -1.10 | 0.28 | B |
| C573415 | 24 | 0.21 | 0.18 | 0.02 | -0.14 | 0.48 | 0.30 | 0.77 | C |
| C573415 | 30 | 0.27 | 0.19 | 0.08 | -0.42 | 0.38 | 1.11 | 0.28 | P |
| C576461 | 42 | 0.43 | 0.39 | 0.04 | -0.12 | 0.27 | 0.42 | 0.67 | B |
| C576461 | 18 | 0.22 | 0.25 | -0.03 | 0.14 | 0.54 | -0.25 | 0.81 | C |
| C576461 | 28 | 0.25 | 0.29 | -0.04 | 0.18 | 0.41 | -0.44 | 0.67 | P |
| C579871 | 15 | 0.47 | 0.44 | 0.02 | -0.07 | 0.44 | 0.16 | 0.88 | B |
| C579871 | 7 | 0.14 | 0.20 | -0.05 | 0.37 | 1.05 | -0.35 | 0.74 | C |
| C579881 | 35 | 0.14 | 0.27 | -0.12 | 0.74 | 0.45 | -1.64 | 0.11 | B |
| C579881 | 20 | 0.30 | 0.19 | 0.11 | -0.55 | 0.44 | 1.26 | 0.22 | C |
| C579881 | 22 | 0.27 | 0.18 | 0.10 | -0.51 | 0.44 | 1.17 | 0.25 | P |
| C581328 | 39 | 0.72 | 0.61 | 0.11 | -0.22 | 0.22 | 1.03 | 0.31 | B |
| C581328 | 18 | 0.00 | 0.39 | -0.39 | 2.00 | < .61 | -3.30 | 0.00 | C |
| C581328 | 29 | 0.52 | 0.42 | 0.09 | -0.27 | 0.30 | 0.89 | 0.38 | P |
| C583200 | 43 | 0.21 | 0.19 | 0.02 | -0.14 | 0.37 | 0.37 | 0.71 | B |
| C583200 | 27 | 0.11 | 0.11 | 0.00 | 0.00 | 0.61 | 0.00 | 1.00 | C |
| C583200 | 33 | 0.09 | 0.12 | -0.03 | 0.30 | 0.59 | -0.51 | 0.61 | P |
| C709029 | 34 | 0.44 | 0.43 | 0.01 | -0.02 | 0.30 | 0.07 | 0.95 | B |
| C709029 | 19 | 0.32 | 0.32 | 0.00 | 0.00 | 0.46 | 0.00 | 1.00 | C |
| C709029 | 32 | 0.31 | 0.32 | -0.01 | 0.03 | 0.35 | -0.09 | 0.93 | P |
| C709682 | 40 | 0.57 | 0.55 | 0.02 | -0.05 | 0.24 | 0.21 | 0.83 | B |
| C709682 | 13 | 0.15 | 0.38 | -0.23 | 1.07 | 0.67 | -1.59 | 0.14 | C |
| C709682 | 27 | 0.44 | 0.37 | 0.07 | -0.24 | 0.33 | 0.71 | 0.48 | P |
| C709686 | 36 | 0.14 | 0.27 | -0.14 | 0.80 | 0.44 | -1.81 | 0.08 | B |
| C709686 | 20 | 0.25 | 0.17 | 0.08 | -0.47 | 0.48 | 0.98 | 0.34 | C |
| C709686 | 33 | 0.27 | 0.18 | 0.10 | -0.52 | 0.36 | 1.47 | 0.15 | P |
| P567764 | 41 | 0.05 | 0.04 | 0.01 | -0.17 | 0.72 | 0.24 | 0.81 | B |
| P567764 | 19 | 0.05 | 0.02 | 0.03 | -0.76 | 1.15 | 0.66 | 0.52 | C |
| P567764 | 32 | 0.00 | 0.03 | -0.03 | 0.75 | < 1.43 | -0.53 | 0.60 | P |
| P567783 | 48 | 0.25 | 0.26 | -0.01 | 0.07 | 0.32 | -0.22 | 0.83 | B |
| P567783 | 18 | 0.33 | 0.17 | 0.16 | -0.82 | 0.43 | 1.88 | 0.08 | C |
| P567783 | 26 | 0.08 | 0.16 | -0.09 | 0.82 | 0.66 | -1.24 | 0.23 | P |
| P567850 | 39 | 0.51 | 0.42 | 0.10 | -0.28 | 0.26 | 1.08 | 0.29 | B |
| P567850 | 21 | 0.10 | 0.22 | -0.13 | 0.93 | 0.65 | -1.43 | 0.17 | C |
| P567850 | 37 | 0.22 | 0.25 | -0.03 | 0.16 | 0.38 | -0.41 | 0.68 | P |
| P568175 | 30 | 0.77 | 0.59 | 0.18 | -0.36 | 0.24 | 1.50 | 0.14 | B |
| P568175 | 23 | 0.30 | 0.38 | -0.08 | 0.29 | 0.42 | -0.69 | 0.50 | C |
| P568175 | 31 | 0.32 | 0.44 | -0.12 | 0.40 | 0.35 | -1.14 | 0.26 | P |
| P572607 | 32 | 0.22 | 0.19 | 0.02 | -0.14 | 0.41 | 0.35 | 0.73 | B |
| P572607 | 21 | 0.14 | 0.13 | 0.02 | -0.13 | 0.61 | 0.22 | 0.83 | C |
| P572607 | 31 | 0.10 | 0.13 | -0.04 | 0.36 | 0.59 | -0.61 | 0.54 | P |
| P573066 | 49 | 0.31 | 0.41 | -0.10 | 0.37 | 0.29 | -1.30 | 0.20 | B |
| P573066 | 15 | 0.40 | 0.26 | 0.14 | -0.56 | 0.45 | 1.25 | 0.23 | C |
| P573066 | 26 | 0.38 | 0.28 | 0.11 | -0.41 | 0.35 | 1.15 | 0.26 | P |
| P573972 | 39 | 0.38 | 0.32 | 0.06 | -0.22 | 0.29 | 0.76 | 0.45 | B |
| P573972 | 20 | 0.25 | 0.18 | 0.07 | -0.37 | 0.48 | 0.76 | 0.46 | C |
| P573972 | 22 | 0.00 | 0.17 | -0.17 | 1.85 | < .83 | -2.21 | 0.04 | P |
| P580577 | 35 | 0.20 | 0.30 | -0.10 | 0.48 | 0.40 | -1.21 | 0.24 | B |
| P580577 | 20 | 0.35 | 0.19 | 0.16 | -0.78 | 0.40 | 1.96 | 0.07 | C |
| P580577 | 25 | 0.20 | 0.20 | 0.00 | 0.00 | 0.48 | 0.00 | 1.00 | P |
| P584438 | 31 | 0.48 | 0.42 | 0.07 | -0.20 | 0.30 | 0.67 | 0.51 | B |
| P584438 | 20 | 0.25 | 0.26 | -0.01 | 0.07 | 0.49 | -0.14 | 0.89 | C |
| P584438 | 30 | 0.20 | 0.26 | -0.06 | 0.33 | 0.44 | -0.74 | 0.46 | P |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| P584678 | 20 | 0.60 | 0.52 | 0.08 | -0.19 | 0.33 | 0.56 | 0.58 | B |
| P584678 | 10 | 0.60 | 0.37 | 0.23 | -0.66 | 0.46 | 1.45 | 0.19 | C |
| P584678 | 23 | 0.26 | 0.43 | -0.17 | 0.64 | 0.43 | -1.47 | 0.16 | P |
| P584681 | 36 | 0.39 | 0.39 | 0.00 | 0.00 | 0.30 | 0.00 | 1.00 | B |
| P584681 | 14 | 0.57 | 0.35 | 0.23 | -0.68 | 0.39 | 1.73 | 0.11 | C |
| P584681 | 31 | 0.23 | 0.33 | -0.10 | 0.46 | 0.40 | -1.15 | 0.26 | P |
| P708856 | 35 | 0.29 | 0.28 | 0.00 | 0.00 | 0.36 | 0.00 | 1.00 | B |
| P708856 | 20 | 0.20 | 0.18 | 0.02 | -0.12 | 0.54 | 0.22 | 0.83 | C |
| P708856 | 31 | 0.16 | 0.18 | -0.02 | 0.12 | 0.48 | -0.26 | 0.80 | P |
| P708858 | 13 | 0.38 | 0.27 | 0.12 | -0.47 | 0.50 | 0.94 | 0.37 | B |
| P708858 | 12 | 0.25 | 0.24 | 0.01 | -0.03 | 0.64 | 0.04 | 0.97 | C |
| P708858 | 9 | 0.00 | 0.18 | -0.18 | 1.33 | < 1.24 | -1.07 | 0.32 | P |
| P708860 | 39 | 0.21 | 0.26 | -0.06 | 0.30 | 0.38 | -0.78 | 0.44 | B |
| P708860 | 19 | 0.32 | 0.19 | 0.12 | -0.61 | 0.44 | 1.39 | 0.18 | C |
| P708860 | 28 | 0.18 | 0.19 | -0.01 | 0.04 | 0.48 | -0.09 | 0.93 | P |
| P708882 | 40 | 0.43 | 0.59 | -0.16 | 0.43 | 0.27 | -1.57 | 0.12 | B |
| P708882 | 22 | 0.41 | 0.38 | 0.03 | -0.10 | 0.38 | 0.26 | 0.79 | C |
| P708882 | 28 | 0.61 | 0.40 | 0.20 | -0.55 | 0.27 | 2.02 | 0.05 | P |
| P710043 | 39 | 0.49 | 0.51 | -0.02 | 0.05 | 0.27 | -0.19 | 0.85 | B |
| P710043 | 20 | 0.25 | 0.30 | -0.05 | 0.22 | 0.49 | -0.45 | 0.66 | C |
| P710043 | 42 | 0.38 | 0.34 | 0.04 | -0.14 | 0.28 | 0.48 | 0.64 | P |

**Appendix 3 DGF by item and specialism**

| Item | Count | Average | Expected | DGF score | DGF size | DGF SE | DGF t | Prob. | Specialism |
|---|---|---|---|---|---|---|---|---|---|
| 2bi | 67 | 0.07 | 0.17 | -0.10 | 0.92 | 0.41 | -2.23 | 0.03 | B |
| 2bi | 40 | 0.35 | 0.22 | 0.13 | -0.60 | 0.29 | 2.08 | 0.04 | C |
| 2bi | 33 | 0.24 | 0.21 | 0.04 | -0.20 | 0.39 | 0.51 | 0.61 | P |
| 2bii | 514 | 0.41 | 0.41 | 0.01 | -0.02 | 0.08 | 0.32 | 0.75 | B |
| 2bii | 300 | 0.44 | 0.44 | 0.00 | 0.00 | 0.10 | 0.00 | 1.00 | C |
| 2bii | 275 | 0.41 | 0.43 | -0.02 | 0.05 | 0.11 | -0.45 | 0.65 | P |
| 3ai | 20 | 0.30 | 0.26 | 0.04 | -0.16 | 0.46 | 0.34 | 0.74 | B |
| 3ai | 11 | 0.18 | 0.29 | -0.11 | 0.55 | 0.73 | -0.76 | 0.47 | C |
| 3ai | 10 | 0.30 | 0.26 | 0.04 | -0.18 | 0.64 | 0.27 | 0.79 | P |
| 3aii | 46 | 0.09 | 0.19 | -0.10 | 0.84 | 0.48 | -1.77 | 0.08 | B |
| 3aii | 25 | 0.16 | 0.16 | 0.00 | 0.03 | 0.54 | -0.06 | 0.95 | C |
| 3aii | 25 | 0.40 | 0.22 | 0.18 | -0.79 | 0.34 | 2.34 | 0.03 | P |
| 3d | 42 | 0.19 | 0.23 | -0.04 | 0.21 | 0.38 | -0.55 | 0.59 | B |
| 3d | 22 | 0.14 | 0.23 | -0.09 | 0.60 | 0.59 | -1.02 | 0.32 | C |
| 3d | 25 | 0.40 | 0.26 | 0.14 | -0.56 | 0.35 | 1.58 | 0.13 | P |
| 4a | 43 | 0.12 | 0.15 | -0.03 | 0.25 | 0.46 | -0.54 | 0.59 | B |
| 4a | 23 | 0.22 | 0.15 | 0.07 | -0.46 | 0.47 | 0.98 | 0.34 | C |
| 4a | 21 | 0.14 | 0.16 | -0.02 | 0.15 | 0.61 | -0.24 | 0.81 | P |
| 4bii | 21 | 0.14 | 0.15 | 0.00 | 0.00 | 0.61 | 0.00 | 1.00 | B |
| 4bii | 13 | 0.15 | 0.14 | 0.02 | -0.13 | 0.75 | 0.17 | 0.87 | C |
| 4bii | 7 | 0.14 | 0.17 | -0.03 | 0.19 | 1.06 | -0.18 | 0.86 | P |
| 5a | 165 | 0.18 | 0.18 | 0.00 | 0.02 | 0.20 | -0.12 | 0.90 | B |
| 5a | 80 | 0.25 | 0.20 | 0.05 | -0.28 | 0.24 | 1.14 | 0.26 | C |
| 5a | 86 | 0.17 | 0.22 | -0.04 | 0.26 | 0.28 | -0.95 | 0.34 | P |
| 5b | 42 | 0.12 | 0.15 | -0.03 | 0.28 | 0.46 | -0.61 | 0.55 | B |
| 5b | 19 | 0.16 | 0.16 | 0.00 | 0.00 | 0.62 | 0.00 | 1.00 | C |
| 5b | 19 | 0.26 | 0.19 | 0.07 | -0.38 | 0.48 | 0.79 | 0.44 | P |
| 5cii | 149 | 0.32 | 0.30 | 0.02 | -0.08 | 0.16 | 0.51 | 0.61 | B |
| 5cii | 81 | 0.30 | 0.30 | -0.01 | 0.03 | 0.23 | -0.12 | 0.91 | C |
| 5cii | 73 | 0.30 | 0.34 | -0.04 | 0.15 | 0.24 | -0.64 | 0.53 | P |
| 5d | 19 | 0.05 | 0.14 | -0.09 | 1.00 | 0.87 | -1.15 | 0.27 | B |
| 5d | 9 | 0.33 | 0.14 | 0.20 | -1.05 | 0.66 | 1.58 | 0.16 | C |
| 5d | 13 | 0.15 | 0.16 | -0.01 | 0.07 | 0.75 | -0.09 | 0.93 | P |
| 6b | 166 | 0.22 | 0.23 | -0.01 | 0.06 | 0.18 | -0.35 | 0.73 | B |
| 6b | 85 | 0.29 | 0.21 | 0.08 | -0.41 | 0.22 | 1.87 | 0.06 | C |
| 6b | 91 | 0.19 | 0.25 | -0.06 | 0.33 | 0.26 | -1.27 | 0.21 | P |
| 6c | 57 | 0.26 | 0.22 | 0.04 | -0.22 | 0.28 | 0.78 | 0.44 | B |
| 6c | 30 | 0.20 | 0.22 | -0.02 | 0.13 | 0.44 | -0.30 | 0.77 | C |
| 6c | 38 | 0.18 | 0.23 | -0.05 | 0.29 | 0.40 | -0.71 | 0.48 | P |
| 7a | 55 | 0.40 | 0.37 | 0.03 | -0.11 | 0.25 | 0.44 | 0.66 | B |
| 7a | 39 | 0.28 | 0.34 | -0.06 | 0.25 | 0.34 | -0.73 | 0.47 | C |
| 7a | 36 | 0.36 | 0.35 | 0.01 | -0.05 | 0.32 | 0.15 | 0.88 | P |
| 7b | 60 | 0.28 | 0.25 | 0.04 | -0.16 | 0.27 | 0.60 | 0.55 | B |
| 7b | 30 | 0.27 | 0.25 | 0.02 | -0.08 | 0.39 | 0.21 | 0.84 | C |
| 7b | 38 | 0.21 | 0.28 | -0.07 | 0.36 | 0.38 | -0.94 | 0.36 | P |
| 7c | 42 | 0.17 | 0.16 | 0.01 | -0.05 | 0.40 | 0.13 | 0.90 | B |
| 7c | 22 | 0.14 | 0.17 | -0.04 | 0.26 | 0.60 | -0.44 | 0.67 | C |
| 7c | 23 | 0.17 | 0.16 | 0.02 | -0.13 | 0.54 | 0.24 | 0.81 | P |
| 8a | 141 | 0.31 | 0.31 | 0.00 | 0.00 | 0.17 | 0.00 | 1.00 | B |
| 8a | 83 | 0.30 | 0.30 | 0.00 | 0.00 | 0.23 | 0.00 | 1.00 | C |
| 8a | 79 | 0.32 | 0.31 | 0.00 | 0.00 | 0.23 | 0.00 | 1.00 | P |
| 8bii | 39 | 0.33 | 0.35 | -0.02 | 0.08 | 0.32 | -0.25 | 0.81 | B |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 8bii | 23 | 0.26 | 0.32 | -0.05 | 0.25 | 0.46 | -0.54 | 0.60 | C |
| 8bii | 27 | 0.44 | 0.37 | 0.07 | -0.24 | 0.33 | 0.73 | 0.47 | P |
| 9a | 44 | 0.23 | 0.20 | 0.02 | -0.12 | 0.34 | 0.35 | 0.72 | B |
| 9a | 16 | 0.06 | 0.19 | -0.13 | 1.14 | 0.85 | -1.35 | 0.20 | C |
| 9a | 25 | 0.28 | 0.24 | 0.04 | -0.18 | 0.42 | 0.44 | 0.67 | P |
| 9b | 42 | 0.21 | 0.23 | -0.01 | 0.07 | 0.37 | -0.20 | 0.84 | B |
| 9b | 18 | 0.22 | 0.27 | -0.05 | 0.23 | 0.55 | -0.42 | 0.68 | C |
| 9b | 15 | 0.33 | 0.25 | 0.09 | -0.39 | 0.50 | 0.77 | 0.45 | P |
| 9c | 41 | 0.39 | 0.32 | 0.07 | -0.26 | 0.29 | 0.89 | 0.38 | B |
| 9c | 22 | 0.23 | 0.34 | -0.11 | 0.52 | 0.48 | -1.07 | 0.30 | C |
| 9c | 15 | 0.33 | 0.36 | -0.02 | 0.09 | 0.51 | -0.18 | 0.86 | P |
| 10aii | 136 | 0.25 | 0.26 | -0.01 | 0.05 | 0.19 | -0.28 | 0.78 | B |
| 10aii | 78 | 0.22 | 0.25 | -0.04 | 0.18 | 0.27 | -0.69 | 0.49 | C |
| 10aii | 69 | 0.32 | 0.26 | 0.06 | -0.26 | 0.24 | 1.07 | 0.29 | P |
| 10b | 160 | 0.20 | 0.22 | -0.02 | 0.12 | 0.19 | -0.60 | 0.55 | B |
| 10b | 92 | 0.21 | 0.21 | 0.00 | 0.00 | 0.25 | 0.00 | 1.00 | C |
| 10b | 82 | 0.28 | 0.24 | 0.04 | -0.18 | 0.23 | 0.79 | 0.43 | P |
| 11aii | 21 | 0.24 | 0.14 | 0.10 | -0.66 | 0.46 | 1.44 | 0.17 | B |
| 11aii | 12 | 0.08 | 0.14 | -0.06 | 0.57 | 0.98 | -0.58 | 0.57 | C |
| 11aii | 10 | 0.00 | 0.15 | -0.15 | 1.24< | 1.28 | -0.97 | 0.36 | P |
| 11b | 22 | 0.18 | 0.14 | 0.04 | -0.27 | 0.53 | 0.51 | 0.61 | B |
| 11b | 13 | 0.08 | 0.14 | -0.06 | 0.66 | 0.96 | -0.68 | 0.51 | C |
| 11b | 12 | 0.17 | 0.17 | 0.00 | 0.02 | 0.76 | -0.03 | 0.98 | P |
| 12bii | 64 | 0.34 | 0.30 | 0.04 | -0.16 | 0.24 | 0.67 | 0.51 | B |
| 12bii | 33 | 0.18 | 0.27 | -0.09 | 0.50 | 0.43 | -1.15 | 0.26 | C |
| 12bii | 33 | 0.27 | 0.27 | 0.01 | -0.03 | 0.37 | 0.07 | 0.94 | P |
| 13aii | 40 | 0.13 | 0.26 | -0.14 | 0.84 | 0.43 | -1.95 | 0.06 | B |
| 13aii | 25 | 0.36 | 0.29 | 0.07 | -0.26 | 0.38 | 0.69 | 0.50 | C |
| 13aii | 14 | 0.57 | 0.30 | 0.27 | -0.87 | 0.40 | 2.20 | 0.05 | P |
| 13bi | 21 | 0.10 | 0.15 | -0.06 | 0.52 | 0.71 | -0.74 | 0.47 | B |
| 13bi | 9 | 0.11 | 0.14 | -0.03 | 0.26 | 1.03 | -0.25 | 0.81 | C |
| 13bi | 11 | 0.27 | 0.14 | 0.13 | -0.80 | 0.60 | 1.34 | 0.21 | P |
| 13bii | 22 | 0.18 | 0.09 | 0.09 | -0.75 | 0.49 | 1.55 | 0.14 | B |
| 13bii | 13 | 0.00 | 0.08 | -0.08 | 1.00< | 1.41 | -0.71 | 0.49 | C |
| 13bii | 9 | 0.00 | 0.09 | -0.09 | .75< | 1.44 | -0.52 | 0.62 | P |
| 14a | 99 | 0.26 | 0.31 | -0.05 | 0.21 | 0.22 | -0.98 | 0.33 | B |
| 14a | 56 | 0.41 | 0.31 | 0.10 | -0.37 | 0.24 | 1.54 | 0.13 | C |
| 14a | 44 | 0.30 | 0.32 | -0.02 | 0.09 | 0.31 | -0.30 | 0.77 | P |
| 14bii | 131 | 0.31 | 0.24 | 0.07 | -0.33 | 0.17 | 1.91 | 0.06 | B |
| 14bii | 73 | 0.16 | 0.23 | -0.07 | 0.40 | 0.30 | -1.31 | 0.20 | C |
| 14bii | 54 | 0.15 | 0.24 | -0.09 | 0.56 | 0.36 | -1.56 | 0.12 | P |