

Statistical grade boundary setting approaches

Executive Summary

October 2022

Executive summary

Context

The International Baccalaureate (IB) is a major international non-profit foundation which offers a suite of educational programmes to students aged between 3 and 19. These educational programmes are alternatives to “in country” programmes, with their own curricula and assessments. As a result, one of the myriad roles for the IB in their programmes’ running is in setting and maintaining the standard of these assessments, in order to ensure fairness and comparability from year to year.

About this study

This study focused on one element of the way in which the IB converts the marks for its assessments into grades and maintains standards from one year to the next. Specifically, it concentrated on the various possible approaches to creating statistically recommended boundaries (SRBs) for the Diploma Programme (DP), Career-Related Programme (CP) and Middle Years Programme (MYP), their strengths and weaknesses and the situations when one approach should be preferred over another. For the purpose of this study, SRBs were defined as the estimates for grade boundaries (the minimum mark at which each grade is awarded) based on a statistical comparison of the relative strength of one year’s cohort compared to the previous year (IB, 2018).

Towards this aim, the study focused on:

- Conducting a review of current and seminal literature regarding different approaches to calculating SRBs, including a review of their strengths and weaknesses and when they are most likely to be appropriate. Furthermore, the review also included other statistical approaches to boundary setting when SRBs are not appropriate, such as when there is no previous cohort taking an assessment.
- Documenting a range of different real-life grade boundary setting contexts relevant to the IB, based around features such as cohort size, rate of growth etc. for individual subjects.
- Analysing the accuracy and appropriateness of these different approaches to calculating SRBs in a range of different contexts identified as relevant to the IB.

Literature Review

The literature review formed the first stage of the project aiming to review and improve the IB’s SRB setting procedures. The ultimate goal of the project was that, ideally, SRBs would provide an accurate estimate of where grade boundaries should be that rarely needs adjusting (or at least, needs more minor adjustments applying than current SRBs do).

In light of this, this literature review aimed to accomplish the following:

1. Map out the ‘universe’ of statistical standard setting procedures, including:
 - a. Any requirements for them being able to be utilised
 - b. Any advantages and disadvantages relative to other approaches
2. Make initial judgments as to which procedures might be most suitable or unsuitable for the IB’s contexts

A wealth of literature was reviewed to gather information on the statistical standard setting methodologies in use. Broadly they fall into one of two categories, score equating and prediction-

based approaches. We summarise the techniques covered by each category, including but not limited to:

- Score equating
 - Basic equating techniques (mean, linear, equipercentile, etc)
 - Smoothing techniques
 - Nonequivalent groups designs
 - Item response theory
- Prediction-based
 - Ways of deriving a prediction
 - External indicators of cohort differences
- Ways of combining multiple approaches

Based on this, we draw some initial conclusions about the standard setting approaches likely to be viable (or not) for the IB's contexts. Prior attainment-based approaches seem infeasible due to a lack of such information, whilst nonequivalent groups designs are also likely impractical since anchor items would compromise the security of IB's assessments.

This leaves, to generalise, three broad approaches which seem promising for the IB:

- a. Basic equating techniques
- b. Concurrent attainment approaches
- c. Approaches seeking to maintain the prior outcome (i.e., via common centres)

Basic equating techniques, as a whole, are suitable in situations where the two cohorts are comparable in ability. However, only around half of IB's contexts meet this assumption. It is worth noting that basic equating approaches *can* be applied in almost any circumstance (they need only a small sample size), which might mean that in some cases they are the only viable option. The question is whether it is advisable to do so (i.e., if cohorts are likely to be dissimilar), or whether it would be preferable to rely on judgemental approaches alone.

Concurrent equating approaches like the 'Instant summary of achievement without grades' or 'ISAWG' method developed by Benton (2017) are powerful, and are suitable for the IB's programmes due to their featuring a broad range of subjects. Further, it offers (by some margin) the most convincing equating approach for some of the most awkward contexts, including very small subjects, those with complete cohort change, and completely new subjects. However, ISAWG approaches are extremely complex, with a huge wealth of available options and modifications (even when compared to the other approaches in this paper). It seems likely that ISAWG would be a method that *can* offer solutions for IB's most challenging contexts, but would require a substantial amount of effort to adequately trial and implement it – effort which might be disproportionate to the benefits it offers. The approach also has other drawbacks, being tricky to implement and a black box in terms of ease of explanation to laypersons.

Approaches seeking to maintain the prior outcome for a subset of the cohort (such as the 'common centres' approach) are a well-established means of attempting to account for cohort changes that is viable as long as there is a large enough cohort, and sufficient centres taking the subject from one year to the next. Whilst not as strong of a method as prior attainment for maintaining outcomes, it is still superior to many other approaches as it aims to account for any change in cohort ability over time. It is also appropriate in just about all of IB's contexts, with the exception of very small cohorts and completely new subjects (though there is the possibility of using common centres to link to a similar existing subject, dubious as this may be).

Later stages of the project can draw upon this review to determine the approaches which are worthwhile carrying out further modelling on to evaluate their appropriateness for the IB's varied awarding contexts.

SRB Simulation analysis

This report summarises the findings of analysis and modelling that comprise the final stage of a project aiming to review the IB's SRB setting procedures, with a view to facilitating their refining and improvement. In light of this, a brief summary of this analysis' aims are as follows:

- To simulate where grade boundaries would fall under selected potentially feasible SRB-setting approaches, across a range of subjects that capture the full range of important contexts for the IB.
- To review how closely each simulated approach's results line up with both one another, and the actual SRBs and grade boundaries set in practice, and draw conclusions about which procedures might be most suitable or unsuitable for the IB's contexts.

Twelve subjects were selected for modelling, covering a broad range of awarding contexts IB typically faces:

Programme	Subject	Grade Award context
DP	Mathematics	Large, stable subjects
DP	Armenian A: Literature	Very small subjects
DP	Swedish A: Literature	Small and Stable Subjects
DP	English A: Language & Literature	Growing subjects: <i>gradual growth</i>
DP	English A: Literature	Shrinking subjects
DP	Global Politics	Growing subjects: <i>significant growth</i>
DP	Information Technology in a Global Society (ITGS)	Growing subjects: <i>sudden growth</i>
DP	Film	Changing curriculum & assessment models
DP	Sports Exercise & Health Science (SEHS)	New subjects: <i>HL introduced</i>
MYP	Mathematics	New cohort in existing subject
DP	Theatre	"Verification" model
DP	Chinese B	Skewed distributions

Five SRB-setting approaches were modelled:

1. **Maintain prior standard:** The boundaries for the current year are set to as closely as possible approximate the grade distribution for the prior year.
2. **Common centres:** Instead of carrying forward the outcome for the entire cohort, both the reference and current year's cohorts are first subsetting to a defined group (here a group of centres present in both years). The outcome for just this group is then what is matched from the reference to the current year (Pinot de Moira, 2019).
3. **Stable common centres:** This is a variant on the above method, where the pool of centres defined as common is further narrowed according to additional criteria. As such the only

further requirement for this method is the defined criteria by which centres are defined as “stable”.

4. **Circle-arc equating:** Uses a graph of the marks on the current assessment against those on a reference assessment, and plots a circle arc through three points: the maximum mark, the midpoint mark achieved, and the minimum mark (Livingston & Kim, 2009).
5. **Instant summary of achievement without grades (ISAWG):** A prediction-based approach that uses concurrent attainment as its external indicator of cohort differences (Benton, 2017). In short, it uses an amalgam of all components’ marks to derive an overall indicator of candidate ability across the entire suite. This is done for the reference and the current year, then the two are equated to establish a year-to-year relationship. The resulting equated ISAWG metric can then be used instead of prior attainment to predict outcomes.

Conclusions

Our broad conclusions about which methods are most suited to which contexts can be summarised as follows:

- In very small subjects of 30-50 candidates or fewer circle-arc equating is the only viable method.
- In growing/shrinking subjects with fairly sizeable growth (of around 25-33 per cent difference per year or higher) then ISAWG seems the best approach, though common centres approaches are viable if there is sufficient data to support this subsetting of the candidature (i.e., it is easier for common centres to be viable given IB’s cohort sizes).
- In new or changing assessments, the key factor is defining what the benchmark to reference the subject to is – the specific SRB-setting methodology is not so important.
- In other contexts not yet mentioned, there are minimal differences between methods and it is difficult to identify which is ‘best’ due to the lack of an objective ‘truth’. Often all methods deviate from the actual boundaries set by a similar distance and direction. If applying another method to account for growing/shrinking subjects however, it would certainly make sense to utilise that method in other contexts too to mitigate the possibility of unexpected cohort ability change having an adverse effect on the standard.
- Broadly, the choice between ISAWG and any other approach is one of pragmatism vs methodological rigour given the similarity of their SRBs – ISAWG is the only method that explicitly allows for “in common centre” ability changes within a subject, but is markedly more complex to implement. One possibility would be to implement common centres as a ‘quick win’, (possibly with stable common centres when this results in sufficient centres and candidates being included in the model), and further investigate ISAWG.

References

Benton, T. (2017, November). Pooling the totality of our data resources to maintain standards in the face of changing cohorts. Paper presented at the 18th annual AEA-Europe conference, Prague, Czech Republic.

IBO (2018). *Assessment principles and practices—Quality assessments in a digital age*. Cardiff, U.K.: International Baccalaureate Organisation.

Livingston, S. A., & Kim, S. (2009). The Circle-Arc Method for Equating in Small Samples. *Journal of Educational Measurement*, 46(3), 330-343.

Pinot de Moira, A. (2019). *Common Centres: In the context of maintenance of standards for the GCSE*. Unpublished report for the WJEC and CCEA.



Unit 109 Albert Mill
10 Hulme Hall Road
Castlefield
Manchester
M15 4LY

www.alphaplus.co.uk

