AlphaPlus

Statistical grade boundary setting approaches

Simulation analysis for the IB

September 2022

1. Contents

List of tables

2. Executive summary

It is recommended that this report be read in conjunction with the accompanying report from the first phase of this research project:

AlphaPlus (2021), 'Statistical grade boundary setting approaches: Literature review for the IB.' Manchester, U.K.: AlphaPlus Consultancy Ltd.

This report summarises the findings of analysis and modelling that comprise the final stage of a project aiming to review the International Baccalaureate's (IB's) Statistically Recommended Boundary (SRB) setting procedures, with a view to facilitating their refining and improvement. The ultimate goal of the project is that, ideally, SRBs would provide an accurate statistical estimate of where grade boundaries should be for a component, that aligns with other evidence and with the final grade boundaries (or at least, needs much more minor adjustments applying than current SRBs do).

In light of this, a brief summary of this analysis' aims are as follows:

1. 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.
2. 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. Five SRB-setting approaches were modelled:

a. Maintain prior standard
b. Common centres
c. Stable common centres
d. Circle-arc equating
e. Instant summary of achievement without grades (ISAWG)

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.

3. Context

3.1. IB's current standard setting procedures

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. IB's standard setting is done via a process called 'grade awarding'.

Historically the IB's grade awarding model can be described as "weak criterion referencing". In other words, a balance of the criterion (competence in the domain at hand) with how candidates have performed in prior years is used to set the standard. The following diagram (IBO, 2018) shows the three key sources of information that input into IB's grade awarding process.

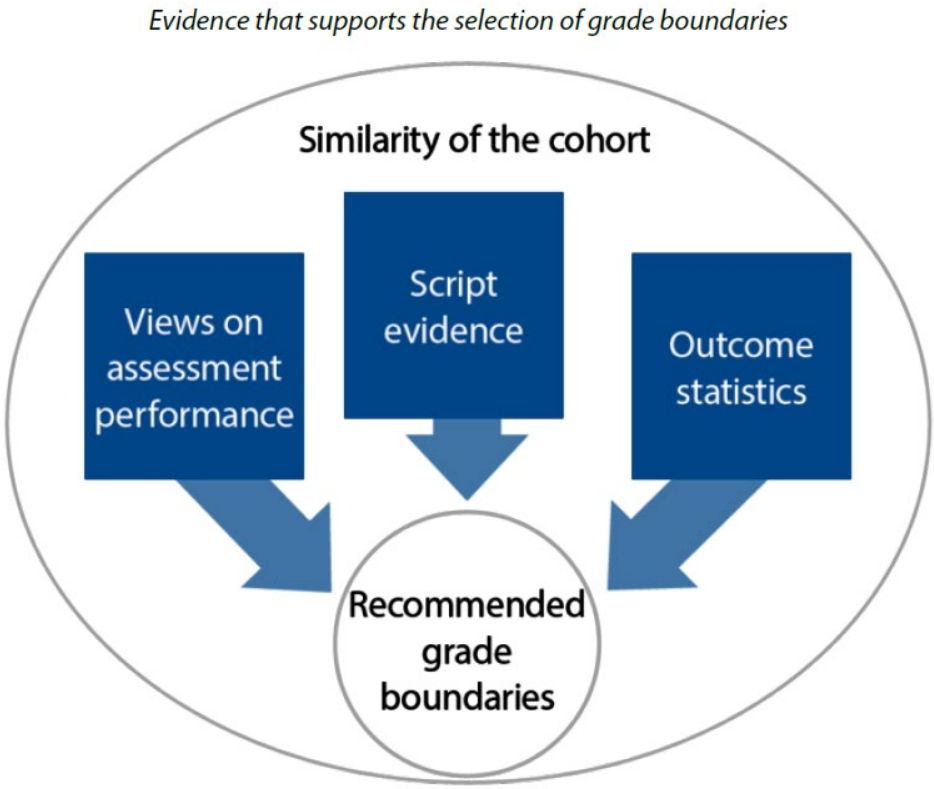

Figure 1: IB's current inputs into grade awarding activities

The first input into grade awarding is 'outcome statistics'. Whilst this is multifaceted, including histograms of mark distributions, mean marks, item level data and more, one of the most valuable statistics are the statistically recommended grade boundaries (SRBs) for key judgemental grades (3, 4, and 7). These use quantitative data from the prior and current session to ask the question "If the prior cohort had sat this year's exam, what grade boundaries would be needed to maintain the same overall grade distribution?". Notably, the SRBs are actually norm-referenced as opposed to criterion-referenced, which in the absence of other inputs would result in a maintenance of outcomes from year to year.

The criterion-referencing element of IB's grade awarding is introduced by the next input into the process, evidence from scrutiny of candidate scripts. This qualitative input is used by the awarding committee to decide whether the difficulty of the assessment and/or the ability of the cohort has changed relative to last year, and to determine final grade boundary positions.

Finally, the third input are views on assessment performance from key personnel, which are not as explicitly criterion- or norm-referenced. They tend to be used to contextualise the decision on final grade boundaries in light of the above two pieces of evidence; i.e. they may help the awarders decide whether the prior outcome can validly be maintained or should be adjusted.

3.1.1. Issues with current SRB setting procedures

The current norm-referenced SRB procedure is reliant on the ability of the cohort remaining constant relative to the prior session in order to be completely valid. Whilst it is likely that in many of the IB's awarding contexts this is the case, in many others it will not always be – and further to this, in some contexts there will be no prior session to benchmark against.

When the cohort's ability is not stable over time (or when there is no prior session to refer to) the current purely norm-referenced SRB setting method ceases to provide the best possible estimate of where grade boundaries should be placed. Below we list some common contexts and how they violate these assumptions, leading to SRBs potentially being inaccurate.

1. Large stable subjects (whilst large cohorts are likely to be stable, performance might still change)
2. Small subjects (small cohorts are inherently less stable in ability over time)
3. Growing subjects (the 'new' centres starting a subject are likely to cause a shift in cohort ability profile over time)
4. Changing curriculum or assessment models (the cohort may initially struggle with a new assessment reflecting a drop in effective ability)
5. New subjects (in these cases there is no prior data on which to base SRBs)

It is worth noting that these contexts are not always distinct – they frequently co-occur. For instance, a small subject can be growing, and a large stable subject can undergo a change to its assessment model.

3.2. Aim of this report

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. The ultimate goal of the project is that, ideally, SRBs would provide an accurate estimate of where grade boundaries should be that rarely needs adjusting (or at least, needs much more minor adjustments applying than current SRBs do).

The prior stages of this project comprised an initial literature review scoping out the 'universe' of possible SRB setting methodologies, resulting in the selection of several possible SRB-setting approaches that might be workable in the IB context. The second phase consisted of a scoping of different IB awarding contexts via focus groups involving a range of IB subject personnel, which was used to select a number of IB subjects that captured the full breadth of such contexts that the selected SRB-setting approaches could then be trialled upon. IB have subsequently shared these subjects' historic data for a number of sessions in order that simulations can be carried out to accomplish this.

In light of this, a brief summary of this analysis' aims are as follows:

1. 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.
2. To review how well each simulated approach's results lines up with indications of where the grade boundary 'should' be and draw conclusions which procedures might be most suitable or unsuitable for the IB's contexts.

4. Methodology

4.1. Subjects and SRB approaches

Earlier stages of the project focused on agreeing a) what SRB setting approaches to simulate and b) which subjects to conduct these simulations on. The conclusion of these decisions and therefore what was the focus of these simulations was as follows.

The subjects of interest are listed in the table below.

Table 1: Subjects simulated in this analysis

Grade Award context	Programme	Group	Subject / options /timezone	Sessions
Large, stable subjects	DP	5	Maths SL TZ2	May
Very small subjects	DP	1	Armenian A: Literature SL	May
Small and Stable Subjects	DP	1	Swedish A: Literature SL	May
Growing subjects: *gradual growth*	DP	1	English A: Language & Literature HL	May
Shrinking subjects	DP	1	English A: Literature HL TZ1	May
Growing subjects: *significant growth*	DP	3	Global Politics HL	May
Growing subjects: s*udden growth*	DP	3	ITGS HL	November
Changing curriculum & assessment models	DP	6	Film HL	M19
New subjects	DP	4	Sports Exercise & Health Science (SEHS) HL	M18
New cohort in existing subject	MYP	5	Mathematics	November
"Verification" model	DP	6	Theatre HL	May
Skewed distributions	DP	2	Chinese B SL	November

It is worth highlighting that the IB lacks a clear definition of what a 'growing' or 'shrinking' entry subject is. In lieu of this, we have adopted the informal definition of these terms in discussion with the IB – that being that 'stable' subjects are those growing at the same rate as the overall candidature, 'growing' as those whose cohorts are increasing in size at a faster rate than the suite as a whole, and 'shrinking' as subjects growing at a slower rate than general growth[1].

The methodological techniques of interest are as follows:

1. The current SRB setting method (Maintain prior standard)
2. Common centres (Maintain a subset of the cohort's outcomes)
    a. Stable common centres (Maintain a subset of the cohort's outcomes)
3. Circle-arc equating
4. Instant summary of achievement without grades (ISAWG)

Note that our modelling of SRBs is in all cases carried out at the component level. This is because this is the level that SRBs are set in practice within the IB's awarding process – whilst the subject level outcomes are ultimately what is most important and the final boundary positions are often

[1] Note that our modelled 'shrinking' subject does actually see an absolute drop in cohort size over time, so fits both possible definitions of shrinking.

dictated by this,[2] the SRBs are an interim stage in the boundary setting process that does not factor the subject outcome in. We do however present results at both a component and a subject level (the latter having aggregated component-level results for a given SRB-setting methodology to subject level) in order to provide this more ultimately impactful indication of the impact a method would have on subject level outcomes.

In general, all of the methods above were tested out in all of the subject scenarios – even in situations where we know it is inadvisable to carry them out (i.e., common centres with very few common centres) as this serves to demonstrate the situations in which we should avoid using each technique. The exceptions are where it was simply not possible to carry out the method – i.e., there are no common centres, or the "new subject" scenario offers no prior information to use to set the standard or to equate to. This also serves to generate information about the range of scenarios each method can accommodate. Such instances are discussed in detail where they occur in the results section.

In a similar vein, all methods were run across three year's data, from 2017 to 2019. This repetition of the approach on multiple datasets generates a number of results for each, which helps increase the validity of our findings by not just having them be based on a single (potentially) anomalous session. The two exceptions are changing and new subjects, where we are specifically interested in the session where the change occurred.

Per the table, we conducted this analysis on full subjects – meaning several components within each subject/level/timezone combination (consider use of 'subject' moving forward shorthand for this combination of three variables). However, some subjects contain coursework components where the grade boundaries are not awarded in the usual way – they are retained from session to session as the task remains the same. As such we did not model these components, and when aggregating component-level marks to derive subject-level results, we simply applied the actual grade boundaries to coursework components, as this would be what occurred regardless of which SRB setting methodology was applied. The exception was Theatre which, as an entirely coursework subject, had to have each component follow our approach for non-coursework components so as not to be excluded from our analysis entirely. This is despite the fact that the subject usually follows a verification awarding model; SRBs are not typically generated for Theatre and a check that the established boundaries are still appropriate is conducted. As such we are deviating slightly from the 'normal' awarding process for this subject so as to include an entirely coursework subject in the analysis.

Note that like many awarding organisations, IB operates with a set of 'judgemental' boundaries that are set using SRBs (the 3/4, 6/7 and 2/3 boundaries)[3]. To mimic IB practice, we will only model these boundaries directly using our various SRB setting methods. The remaining boundaries are set using interpolation rules, which we will follow in our modelling to arrive at a complete set of boundaries for each method. Put simply, the rules are:

[2] But not always – it can be the case that external factors such as subject pairs analysis, or an exercise to align standards across a group of subjects (as examples) can have an impact on the final boundary positions selected. This is important to bear in mind when considering scenarios in our analysis where the final grade boundaries are quite different to both our modelled and the actual SRBs.

[3] Note however that in a small minority of cases SRBs might not be calculated and boundaries not interpolated, such as for extremely small subjects (i.e. Armenian A: Literature SL, which we model SRBs for below).

- Split the number of marks between the judgemental boundaries into the number of grades between them. I.e. the 3/4 to 6/7 difference is split into grades 4, 5 and 6, and below the 2/3 boundary the marks are split into 1 and 2.
- If there is a non-integer number of marks per grade after doing this, work out where these grade widths would lead each grade boundary to fall as a decimal. Then, round these decimal boundary marks to the nearest integer – unless it is 0.5, in which case round down. This results in the following:
    - For the 3/4 to 6/7 span:
        - If there are two 'remainder' marks, they are allocated into grades 4 and 6.
        - If there is one 'remainder' mark, it is allocated into grade 5.
    - For below the grade 2/3 boundary, a 'remainder' mark is allocated into grade 2.

4.2. Data

To enable all the above methods to be utilised, the following data was provided:

- For the Diploma programme:
    - The entire dataset for all subjects May and November 16-19 (including DP core), including:
        - Indications of centre membership
        - Indicators of resit status
        - Component-level mark data
        - Component-level grade data
        - Component-level maximum mark available information
        - Subject-level mark data
        - Subject-level grade data
        - The historic SRBs and the percentage of data these were based upon (N.B. Armenian A: Literature SL did not have SRBs so this will not be provided)
        - The historic zone of uncertainty
        - The historic actual grade boundaries
- For the MYP:
    - The Mathematics dataset for November 16-19 including:
        - Indications of centre membership
        - Component-level mark data
        - Component-level grade data
        - Component-level maximum mark available information
        - Subject-level mark data
        - Subject-level grade data
        - The historic SRBs
        - The historic actual grade boundaries
- General information:
    - Confirmation as to which boundaries are judgmental for each subject of interest.
    - Interpolation rules for non-judgemental boundaries.
    - Rules for SRB-setting in atypical cases; no candidates achieved a grade in the reference year data, etc.
    - Information on component weightings for subjects of interest in the years analysed.
    - Information on component mark aggregation to subject level, i.e., rounding rules.
    - Information on any changes made to assessments in the subjects of interest in the years modelled.

- Indications as to which components within each subject have static grade boundaries that do not change year-on-year.

Note that the data provided for the DP and MYP was post-Enquiry Upon Results (EUR) and at-risk, each of which can lead to some mark and/or grade changes, though typically very few within a given component. Similarly, the data provided is 'complete', unlike that used for the actual grade awards, during which marking may not have completely finished by the time of the award meeting.

This can lead to slight inconsistencies in our modelling vs the actual SRBs set, and is a key reason why we both present the actual SRBs alongside the 'maintain prior standard' scenario which effectively replicates the SRB-setting process, meaning any discrepancies from actual SRBs due to the above factors should be apparent.

4.2.1. Data cleaning

There were several unusual cases in the data which, for the purposes of this analysis, required data cleaning to be conducted in order to have the most valid dataset possible. Each such reason for cleaning is discussed below in turn.

4.2.1.1. School supported self-taught candidates

School supported self-taught (SSST) candidates are those who do not have a teacher of the particular language variant for a given DP Language A: Literature SL subject in their school, but still study the subject. Due to this, they cannot be internally assessed by the teacher as in normal circumstances, hence the distinction being important for IB operationally. Whilst it could be argued that such candidates are invalid to compare to 'typical' entry candidates due to their lack of a teacher, we consider that teacher effects are inherently already wrapped up in the data for other candidates, and 'no teacher' is essentially just another teacher effect.

For Literature subjects in M18 and prior sessions, SSST candidates' marks for the internally assessed component needed adjusting due to a change in how examiners submitted marks in M19. Both taught and SSST had a moderated max mark of 30 in 2016-2018, but SSST changed to 60 in 2019 for operational reasons (whilst taught remained at 30).

To resolve this, rows in the data where GROUP_NO==1 & ASSESSMENT_TYPE==”INTERNAL” & SCALING_FACTOR==0.5 had their moderated mark halved (rounding 0.5s up) and scaling factors re-set to 1.0. This puts M19 data for self-taught candidates on the same scale as non-SSST candidates.

4.2.1.2. Timezoned components

Some IB subjects are timezoned; despite sometimes sharing papers or questions, they have two separate awards, one for various regions of the globe. The two timezones are labelled timezone 1 and timezone 2. However, such subjects share coursework components. These components (and indeed all components in any non-timezoned subjects) are labelled timezone 0, i.e., they are timezone agnostic.

For the ISAWG metric's derivation, each individual component is input to the model separately. Because timezoned components are awarded completely separately in practice, they are treated

as different components, so this was mirrored for the ISAWG and timezone was used as a factor in identifying unique components[4].

For the grade modelling of each subject of interest, timezone selections (for timezoned) subjects are listed in Table 1. Note that because coursework components were not modelled and were solely used to aggregate candidates' results to subject level, these components all being labelled as timezone 0 did not pose any issue for the analysis; the other timezone candidates' coursework results would simply not be aggregated to subject level due to their results for the modelled components not being included in the analysis, and these individuals thus naturally 'fall out' of the analysis.

4.2.1.3. Candidates with no grades

Some candidates in the data received "N" or "P" grades, which equate to in turn "No grade awarded" and "Grade pending". Both were infrequent but Ps were much rarer, with under five instances in the entire dataset.

N grades typically occur when a candidate misses one (or several) assessment(s) for a given subject; some components will have a grade, others Ns. For the component level main analysis, we excluded component results scored as Ns, as they did not receive a grade. However, all other components that did receive a grade were included in the analysis. Therefore, for some candidates, we did not have a "complete data set" for analysis if they received an N grade in some components and not in others. At subject-level, however, candidates with any N grades were excluded from the results reported in this paper entirely, because it is not possible to aggregate to subject level without a full set of valid component grades. Note that any "missing" data for components resulted in a similar occurrence to Ns in one or more components – candidates' data was not aggregated to subject level in these situations.

For the very small number of candidates with P results, these component results were completely removed from the analysis, as their minimal incidence means this has extremely low impact on our findings.

4.2.1.4. Extended essay components

The extended essay (EE) is one of the DP Core components which all diploma candidates will take. However, it can be on almost any subject the IB offers – Biology, Literature, Arts, etc. In the data this is recorded in the component fields, which, by default, results in an EE component appearing under each subject at least one candidate picks as their EE topic. However, regardless of subject selection and presentation in the data, the EE component is marked to a common set of marking criteria and graded as a single component. To ensure this treatment in the data, the subject selection in the data was overwritten with an "EE component" dummy value to ensure that regardless of subject, all the component's data would be analysed together (for the purposes of ISAWG modelling).

4.2.1.5. Validation of mark scaling

[4] Technically NO-LANG_CODE was used to determine this, but timezone is a component of this variable.

As part of our data validation, we checked that moderated marks multiplied by the scaling factor equated to the scaled mark in all cases in the data. For a handful of rows (some in visual arts, and one in psychology) this was not the case, however.

The visual arts cases were because for a brief period of time, a minimum scaled mark of 1 was applied for certain assessment components and so all scaled marks of 0.47 got rounded up. These instances of mismatched scaled scores were left to stand for the analysis.

The psychology case was for a candidate who switched from Psychology SL to HL after results were released. The internal assessment is the same for both HL and SL Psychology, so the mark was carried over, along with the SL scaling factor, hence the mark being higher than it looks like it should be. Given this is a highly anomalous case, it was removed from the data for the analysis.

4.3. SRB-setting approaches

This section details the intricacies of each simulated approach.

4.3.1. Maintain prior standard

This is the most straightforward methodology to implement. The boundaries for the current year are set to as closely as possible approximate the grade distribution for the prior year.

4.3.2. Common centres

This method is slightly more complex than the method above. Instead of carrying forward the outcome for the entire cohort, both the reference and current year's cohorts are first subsetted 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).

There is an additional step however – weighting the common centres' prior grade distributions by current year entries, to ensure that centres have the influence on the overall prediction their current entry size (rather than an older out of date entry size) dictates they should. Pinot de Moira (2019) documents this approach and notes its use by the WJEC and CCEA awarding bodies in the UK.

It is also advised in Pinot de Moira's (2019) review of common centres approaches, that where multiple specifications exist for the same cognate subject area (i.e., HL and SL; potentially subjects like Language A: Language and Literature vs Language A: Literature) which entrants have a tendency to move between, that one prediction is derived for the entire group of specifications. In other words, the overall proportion of candidates receiving each grade across the group of specifications would be controlled, with changes to each individual specification's grade distribution permitted if stronger centres are known to be moving to/from it over time.

Whilst we did not implement any such grouping here, this should nonetheless be borne in mind as a key variant of the approach.

4.3.2.1. 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".

Generally, the definition of stable centres is determined by 'how similar the entry size of the centre is from reference to current year'. Typically, this is defined as 'within X% of the size of the centre's entry in the reference year', though the precise value X takes can vary (Eason, 2006). However,

Benton and Sutch (2014) expanded on this by suggesting that historically large centres, those with stable results over time, and those with stable candidate characteristics could also be used to help define what a truly stable centre was.

However, these additional constraints seem likely to be problematic in the IB context and would likely reduce the pool of common centres to an unreasonably small number (as well as rendering the identification of common centres much more complex), so we followed the 'in use' stable common centres practice in WJEC (Pinot de Moira, 2019) and simply defined stable common centres as those with +/-15% of the cohort size in the reference year.

4.3.3. Circle-arc equating

Circle-arc equating comes from a completely different school of approaches to all the others in this analysis plan – score equating rather than prediction-based (see section 2.2 of AlphaPlus (2022) for an explanation of these two different schools of standards maintenance approaches). In brief, it 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 (Livingstone & Kim, 2009).

It is therefore somewhat similar to mean equating in that it applies a shift to the marks on one form according to how well its midpoint lines up with that observed on the other form, but the circle arc also ensures accuracy of the equating throughout the mark range better than mean equating. Ultimately it results in a similar output to equipercentile equating but is feasible with much smaller sample sizes.

The main decision for this approach is how to define the three points. The midpoint is typically the mean mark achieved on each form, and the maximum mark is typically the maximum mark achieved on. Whilst in some cases the midpoint can be chained linear equated, in the IB context without common items between sessions this will not be possible. As such this method does not attempt to control for variation in overall cohort ability between sessions as common centres or ISAWG does.

However, the minimum mark varies – it is sometimes zero and sometimes the minimum meaningful mark, though Livingstone and Kim (2009) observe that the lowest meaningful mark concept is generally useful in MCQ papers. Given that IB assessments are not heavily MCQ based in their entirety, we used zero.

4.3.4. ISAWG

The instant summary of achievement without grades method, or ISAWG, is by far the most complex and involved approach to be modelled here. ISAWG is 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.

In perhaps easier to digest terms, the ISAWG essentially says "based on the marks you have for all components marked so far, the best estimate of your 'general academic ability' is X". Focusing into a grade award for a particular component, this generation of a 'general academic ability' based on data across all assessments sat is done for every candidate – both for a reference year, and the current year. Having done this, we then know the distribution of 'general academic ability' in each

year, and crucially, how different it is this year relative to the reference year. Based on this ability difference, the method works out how many more or fewer of each grade to award this year.

There are a number of decisions to be made in an investigating ISAWG approach:

1. Which subjects/components are included in the ISAWG metric, and should any be excluded?
2. Which subjects should ISAWG be used to award; all or just a subset?
3. For a modelling exercise such as this, whether to follow an 'in session' or 'end of session' approach, and if 'in session' what delay to implement for results being available?
4. What equating approach should be utilised to establish the year-to-year ISAWG relationship?
5. For each subject, which means of adjusting the prior outcome should be used (prediction matrix, logistic regression or a score equating approach?)
6. For each subject, which candidates should be included when adjusting the prior outcome; all candidates, just those from common centres, non-resitters, or another subset of candidates?

We address each of the above decisions in turn below.

4.3.4.1. Which data should be included in ISAWG's derivation

The derivation of an ISAWG score amounts to using raw mark data from all assessments all candidates have sat to infer their overall 'general ability'. Note however that we will use final unscaled marks in this work[5]. The more data the better, as we are likely to have a more informed estimate of the impact of score on a given component on the ISAWG with as much data as possible included.

There are several reasons why a particular subject or component might be removed, however. If a subject is overly influential on the data it could prove unhelpful, particularly for some subjects. For instance, because Mathematics is such a central subject most candidates do, it might be more beneficial to Arts subjects to exclude it from the ISAWG – perhaps even to derive a separate Arts ISAWG.

Per discussion with IB, we conducted an ISAWG analysis for the Diploma programme using all available subjects and components – including the DP core (all EE subjects were aggregated however, as they share a common mark scheme and standard). This is largely for reasons of keeping a practical scope; investigating the optimal subjects to include for each component would be a substantial undertaking.

Notably Benton (2017) did exclude some subjects from his initial work on ISAWG, but our understanding is that this was done because he was more interested in assessing the efficacy of the approach in principle, rather than directly applying it to a real session where all subjects need SRBs setting. Additionally, we are specifically interested in the IB context where there are more small entry subjects than the English GCSE context used by Benton and colleagues to apply a 500

[5] In rare cases these needed adjusting due to an incomparability between self-taught and other candidates. Cases where GROUP_NO==1 & ASSESSMENT_TYPE=="INTERNAL" & SCALING_FACTOR==0.5 had their final unscaled mark halved (rounding 0.5s up) and scaling factors re-set to 1.0.

candidate per component threshold would remove a substantial proportion of the available data and likely negatively impact the ISAWG overall.

4.3.4.2. Which subjects should ISAWG be used to award

Whilst the ISAWG will need to include all components in its modelling, we will only be considering how it produces SRBs for the subjects of interest listed above. This keeps things manageable; otherwise the ISAWG modelling would be orders of magnitude more complex than all other methods outlined above combined. In practice it may emerge that we only recommend ISAWG be used to award particular subjects, however.

4.3.4.3. Whether to model 'in session' or 'end session' ISAWG

As a concurrent attainment approach, in a real session ISAWG would be re-computed regularly throughout as more data became available. However, this also means that the effectiveness of the approach is based on the point at which an assessment falls within the session and thus how much other data is available with which to derive the ISAWG.

Per discussion with the IB, we solely modelled an 'end of session' ISAWG approach in order to provide a 'best case' indication of how ISAWG is able to be used to set SRBs, when it has all data from the session available. This is particularly pertinent as if ISAWG is seriously considered the order of assessments within the session might be tweaked to maximise the strengths of the approach – so any modelling of 'in session' ISAWG done now may no longer be valid after order changes.

4.3.4.4. Equating the reference and current years' ISAWG

Equating can be used in the ISAWG approach in two places. The first is in the calibration of the ISAWG in the reference session to the ISAWG separately derived for the current session. Here the aim is to put both ISAWGs onto the same scale. Note that if we are content to award (across the board in the whole suite) the same proportion of each grade every year, this step is not necessarily needed; we can simply use the raw ISAWG scores. But if there is a shift in overall cohort ability from year to year (i.e. from a substantial influx of weak or strong candidates), this approach becomes risky; we might see an over-award of grade 7s and a lowering of the standard after an influx of weak candidates, for instance.

As such we followed Benton (2017)'s example and utilised linear equating on a subset of candidates to attempt to account for potential shifts in ability between the sessions. Benton used all centres with 50 or more candidates in both sessions (and 30 per cent or less change between sessions) as common centres within which overall ability could be assumed to remain reasonably stable over time. We propose to start with this rule of thumb, though we are conscious that small centres being common might mean we have to adjust this limit.

The common centres subset of each year's ISAWG data was input into a linear equating model and used to transform the current year's ISAWG scores onto the scale of the reference year's. Note that whilst not all candidates' data was included in this equating exercise, the linear model can be used to transform every candidate's ISAWG score even when they are in a small centre.

4.3.4.5. What means of adjusting the prior outcome should be used

Once the ISAWG metrics for both sessions have been calibrated onto the same scale, it remains to actually use it to generate SRBs. There are a number of potential approaches, and either score equating or prediction-based ones can work. Benton (2017) utilises a linear equating method,

treating ISAWG as if it were an anchor test (i.e. carrying out a non-equivalent groups design such as Tucker or Levine), whilst Johns and Evans (2019) utilise a prediction matrix approach.

We used a decile-based prediction matrix approach, for a few reasons. Firstly, the ISAWG approach is obtuse already, and adding in the complexity of a statistical equating approach with an anchor assessment makes it even more challenging to explain. By contrast, the prediction matrix approach is much more straightforward to explain to relative laypersons.

Secondly, there are a huge variety of different non-equivalent groups equating designs available – Benton (2017) discusses the trade-off between chained and frequency equating, but within each there are a huge range of possible approaches to choose. With prediction matrices there is broadly one agreed approach, and if this approach appears to work, a substantial avenue of investigation can be somewhat closed off.

4.3.4.6. Which learners should predictions be generated for

As outlined above, we utilised a prediction-based approach. However, predictions may be more or less valid for particular candidates. If a candidate is a resitter taking a single component or a candidate not sitting the full DP, then their ISAWG score might be unstable and we would arguably not want to include them in predictions. In practice, in the context of awarding bodies based in England where the matrix approach is commonly used, excluding some learners from predictions is commonplace for similar reasons; candidates sitting the assessment at an unusual age are often removed[6].

Ultimately we should consider excluding candidates from the prediction matrix for a particular subject if we consider it likely that they will cause the component to have a different ISAWG to component score relationship relative to the reference year. Resitters are a possible example, particularly if increasing or decreasing in number; candidates who only sit a small number of subjects may also need to be excluded for similar reasons of irregularity. However, because removing non-diploma candidates would result in around 50% attrition of the data, it was agreed with IB that only resitters should be removed – though we note that investigating an approach using only diploma candidates might be fruitful in future.

One other possible scenario where it might be prudent to exclude candidates from the matrix is if there is considerable cohort growth, with the new candidates being comparatively strong or weak at the component relative to their performance on the rest of the DP. This does risk complicating the ISAWG approach significantly, but is certainly worth further consideration and investigation.

4.3.4.7. Summary

To summarise the above we present the following to briefly outline the decisions on the precise approach for ISAWG modelling:

1. All scaled score data for all subjects and components within the DP is included when deriving ISAWG metrics in each year.
2. We solely modelled ISAWG as an SRB setting approach for the subjects of interest outlined above.

[6] Note that even did we adopt an equating approach rather than a prediction-based one, we would still need to consider whether any candidates might throw off the ISAWG to component score relationship and make similar exclusions.

3. Given the preliminary nature of this work and to produce a ‘best case’ scenario, we modelled ‘end of session’ ISAWG.
4. Linear equating on a subset of centres (those entering 50+ candidates in both years with under 30 per cent change in size) was used to link the reference year’s ISAWG to the current year’s. (This is more lenient than stable common centre approaches but also uses a considerably larger dataset.)
5. A prediction matrix approach using ISAWG deciles was used to generate predictions for each subject based on its entrants’ ISAWG scores.
6. Non-resit candidates were included in the data used to generate predictions.

Note that our focus on the DP means we cannot check the efficacy of the ISAWG approach for the MYP subject of interest.

4.4. SRB-setting implementation

4.4.1. Sparse Mark Distributions

An issue with any standard setting methodology that relies on cumulative mark distributions to set grade boundaries, in the case of this report the Maintain Prior Standard, Common Centres and Stable Common Centres methods, is that if there are gaps in the distribution where no candidate has scored a particular mark then it can be ambiguous as to where to set the grade boundary. For example, suppose that the top 35% of persons achieved a grade 6 or higher on a reference examination, and in the current year the top 34% of persons scored say a mark of 30 or higher, with no persons scoring 29 or 28, and 36% of persons scoring a mark 27 or higher, the boundary for grade 6 could reasonably set at multiple points between a mark of 27 and 29. The solution to this issue used in our modelling was that if the reference boundary fell within the sparse distribution then to stick with the reference year mark, otherwise the limit of the range (minimum or maximum mark) of sparse marks closes to the reference boundary mark should be used. In the final analysis none of the methods were affected by sparse mark distributions, but it is useful to bear in mind this approach should any of the standard setting techniques outlined in this report be implemented.

4.4.2. Missing Judgemental Boundaries

Because judgemental boundaries of 2/3, 3/4 and 6/7 are used to set the remainder of the current year boundaries, if no person scored grades 3, 4 or 7 in the reference year then the Maintain Prior Standards, Common Centre and ISAWG methods would not work to set the current year grade boundaries in this scenario.

5. Results

In this section, we will cover each subject of interest in turn, presenting subject-level outcomes of our modelling (i.e., component-level results scaled and aggregated to subject-level). Due to the size of the tables, each session's results are presented in a distinct table, with the actual boundaries in the real session shown for reference in the first row. Component-level results can be found in the Appendix for reference to avoid lengthening the paper significantly. We note in brackets in each sub-section heading which sessions (and where relevant, timezones) were analysed for reference.

5.1. DP Mathematics SL (TZ2, May sessions)

This subject is the example of a 'large, stable entry subject' for which the current SRB setting process works quite well. It is, as such, to some degree a control condition – as IB is relatively confident the current boundary setting methodology works fairly well in such instances, we would not be expecting significant deviations away from the current boundaries to emerge in any of our alternative approaches.

Table 2: DP Mathematics SL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	19	38	51	62	73	83	100.0	99.5	91.6	75.3	54.0	29.2	9.6
Maintain prior outcome	0	19	40	52	63	73	84	100.0	99.5	89.7	73.3	51.4	29.2	7.9
Common centres	0	19	39	51	62	73	84	100.0	99.5	90.6	75.3	54.0	29.2	7.9
Stable common centres	0	18	38	51	62	73	84	100.0	99.5	91.6	75.3	54.0	29.2	7.9
Circle-arc equating	0	19	39	51	62	73	83	100.0	99.5	90.6	75.3	54.0	29.2	9.6
ISAWG	0	19	39	51	62	73	83	100.0	99.4	90.6	75.2	54.0	29.2	9.6

Table 3: DP Mathematics SL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	17	34	48	59	71	83	100.0	99.2	92.0	73.8	52.0	28.0	9.1
Maintain prior outcome	0	17	35	47	59	71	82	100.0	99.2	91.0	75.2	52.0	28.0	10.2
Common centres	0	16	34	47	59	71	82	100.0	99.3	92.0	75.2	52.0	28.0	10.2
Stable common centres	0	16	34	46	58	70	82	100.0	99.3	92.0	76.9	53.7	29.6	10.2
Circle-arc equating	0	17	36	49	60	70	81	100.0	99.2	90.1	71.8	49.5	29.6	11.5
ISAWG	0	16	34	47	58	70	82	100.0	99.3	92.0	75.2	53.7	29.6	10.2

Table 4: DP Mathematics SL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	16	32	45	57	69	81	100.0	99.0	90.7	73.8	51.1	27.1	8.3
Maintain prior outcome	0	15	31	45	56	69	80	100.0	99.1	91.5	73.8	52.9	27.1	9.3
Common centres	0	15	31	45	56	69	80	100.0	99.1	91.5	73.8	52.9	27.1	9.3
Stable common centres	0	14	29	43	55	67	79	100.0	99.3	93.1	77.0	54.7	30.6	10.5
Circle-arc equating	0	15	32	45	57	69	81	100.0	99.1	90.7	73.8	51.1	27.1	8.3
ISAWG	0	15	31	45	56	68	80	100.0	99.1	91.5	73.8	52.9	28.7	9.3

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 5: Summary of candidates included in common centre models (DP Mathematics SL TZ2)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	23,168	21,543	1,069	8,054	330
2018	24,980	24,114	1,182	8,886	346
2019	26,031	24,980	1,267	9,962	386

The status of this subject as a 'large, stable entry subject' appears confirmed by the three tables above. The five trialled SRB-setting approaches provide results that are comparable to the current methodology, save for one or two marks. Notably in all cases grade 7 is identical to or more lenient than the actual boundary set, for the lower two judgemental grades there is slightly more variation.

Stable common centres appears the most lenient of the methods, suggesting that the candidates in said subset of centres found the subject marginally harder than the other methods would imply. With the actual boundaries dropping each year (just not as much as by any of our modelled approaches) this suggests that the assessment did ramp in difficulty over time, just by slightly more than was accounted for in practice. Notably even maintain prior outcome suggests lower boundaries than were implemented in practice, lending credence to this.

5.2. DP Armenian A: Literature SL (May sessions)

This subject is the example of a 'very small subject', where the current approach may not work well due to significant changes in cohort ability from one session to the next. Due to its small size, it is also likely that some methods might not function particularly well. Because it relies on data from other subjects, ISAWG is therefore a method we would hope produces promising results in this scenario.

Table 6: DP Armenian A: Literature SL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	22	35	49	62	73	85					100.0	96.3	37.0
Maintain prior outcome	0	22	44	57	67	76	85					100.0	85.2	37.0
Common centres														
Stable common centres														
Circle-arc equating	0	22	43	56	67	76	86					100.0	85.2	37.0
ISAWG	0	24	47	60	70	77	86				100.0	96.3	74.1	37.0

Table 7: DP Armenian A: Literature SL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	22	35	49	62	73	85					100.0	76.2	33.3
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	19	37	51	63	74	85					100.0	76.2	33.3
ISAWG														

Table 8: DP Armenian A: Literature SL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	22	35	49	63	74	87				100.0	93.1	79.3	41.4
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	17	33	47	61	72	85				100.0	93.1	82.8	48.3
ISAWG														

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 9: Summary of candidates included in common centre models (DP Armenian A: Literature SL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	27	26	6	17	<5
2018	21	17	5	<10	<5
2019	29	26	5	<10	<5

In fact, only circle-arc equating is sufficiently robust to small sample sizes to provide results across all three years. There are simply not enough common centres to run those methods, and for maintain prior outcome, there are no candidates at grade 3 or below, meaning the full set of judgemental boundaries cannot be derived[7]. The 'maintain prior outcome' and ISAWG method are the only other methods that function, and provide a result in 2017 only.

Circle arc returns somewhat higher boundaries than the actual ones in 2017, marginally higher ones in 2018, and somewhat lower boundaries in 2019. This is likely down to the mean mark

[7] Whilst IB use the "reference year's boundary" in such cases in live awarding, we consider that it is useful to emphasize that a strict 'maintain prior outcome' methodology does not function in such cases. It is also worth noting that small Literature subjects such as this one also have the approach of having a "standard Literature boundaries" reference set. These are based on the most commonly used grade boundaries across all Literature subjects.

shifting quite drastically in each sitting. Notably maintain prior outcome produces similar results as circle arc, quite different to the actual boundaries. ISAWG however, in the one case it does function, produces even higher boundaries than the other two methods, suggesting that the 2017 cohort is lower ability than the 2016 one in terms of their ISAWG scores.

5.3. DP Swedish A: Literature SL (May sessions)

This subject is the example of a 'small and stable subject'. In other words, it is somewhat analogous to DP Maths SL (TZ2) in terms of the entry being similar each year, but is much lower entry size overall. This is likely to have an impact on some of the approaches we model, as they rely to varying degrees on volumes of scale to ensure accuracy and validity.

Table 10: DP Swedish A: Literature SL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	33	44	57	70	82		100.0	99.6	98.0	83.1	52.8	15.7
Maintain prior outcome	0	18	34	44	57	70	82		100.0	99.2	98.0	83.4	53.0	15.8
Common centres	0	17	33	44	57	71	83		100.0	99.6	98.0	83.4	46.6	13.4
Stable common centres	0	18	32	44	57	70	82			100.0	98.0	83.4	53.0	15.8
Circle-arc equating	0	17	33	45	58	69	81		100.0	99.6	96.8	80.2	54.7	17.4
ISAWG	0	18	35	44	57	70	82		100.0	99.2	98.0	83.1	52.8	15.7

Table 11: DP Swedish A: Literature SL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	16	30	43	56	69	81			100.0	96.9	83.8	49.0	15.4
Maintain prior outcome	0	17	31	42	56	68	81			100.0	97.7	83.8	52.5	15.4
Common centres	0	16	30	42	55	67	79			100.0	97.7	84.9	54.8	18.9
Stable common centres	0	16	30	44	58	72	85			100.0	96.1	79.2	37.5	8.9
Circle-arc equating	0	16	32	43	56	69	81			100.0	96.9	83.8	49.0	15.4
ISAWG	0	14	28	42	56	68	81			100.0	97.7	83.8	52.5	15.4

Table 12: DP Swedish A: Literature SL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	17	31	43	56	69	81		100.0	99.5	96.5	81.9	49.7	14.6
Maintain prior outcome	0	16	31	42	56	67	80		100.0	99.5	96.5	81.8	54.5	15.7
Common centres	0	15	29	40	54	67	80			100.0	97.5	82.3	54.5	15.7
Stable common centres														
Circle-arc equating	0	16	30	43	56	69	81			100.0	96.5	81.8	50.0	14.6
ISAWG	0	15	30	41	55	68	81			100.0	97.0	81.9	51.8	14.6

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 13: Summary of candidates included in common centre models (Swedish A: Literature SL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	248	195	32	41	9
2018	259	215	31	40	10
2019	199	169	29	23	6

Grade boundary marks differ only by a mark or two across the three years with a few exceptions (grade 7 stable common centres in 2018 and grade 3 ISAWG in 2017). There is no obvious pattern of any methodology appearing consistently generous or harsh; most methods shift from being on the lenient to severe side across the three sessions.

5.4. DP English A: Language and Literature HL (May sessions)

This subject is the example of a 'gradually growing subject', in that its cohort has continued to increase in size steadily over time[8]. This is highly likely to upset the current SRB setting methodology, and as a large subject there is significant value in finding a more suitable alternative that more validly accounts for changes in cohort ability over time.

Table 14: DP English A: Language and Literature HL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	42	55	67	80	100.0	100.0	99.7	94.0	68.4	29.6	4.7
Maintain prior outcome	0	12	27	40	54	65	78	100.0	100.0	99.7	95.7	71.0	34.6	6.6
Common centres	0	12	27	40	54	65	78	100.0	100.0	99.7	95.7	71.0	34.6	6.6
Stable common centres	0	12	27	40	54	66	79	100.0	100.0	99.7	95.7	71.0	32.0	5.5
Circle-arc equating	0	12	27	39	53	67	80	100.0	100.0	99.7	96.3	73.6	29.6	4.7
ISAWG	0	12	27	42	55	65	78	100.0	100.0	99.7	94.0	68.4	34.6	6.6

Table 15: DP English A: Language and Literature HL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	42	55	67	80	100.0	100.0	99.4	93.8	69.0	31.0	5.1
Maintain prior outcome	0	12	28	42	55	67	80		100.0	99.4	93.8	69.0	31.0	5.1
Common centres	0	12	28	42	55	67	80		100.0	99.4	93.8	69.0	31.0	5.1
Stable common centres	0	13	30	42	55	68	81	100.0	100.0	99.2	93.8	69.0	28.1	4.1
Circle-arc equating	0	12	28	42	55	67	80		100.0	99.4	93.8	69.0	31.0	5.1
ISAWG	0	12	28	42	55	67	80		100.0	99.4	93.8	69.0	31.0	5.1

[8] Though it is worth noting that this growth is still markedly more than the DP average. It was driven by a shift made by many schools from English A: Literature to this subject starting when this course was introduced in M13.

Table 16: DP English A: Language and Literature HL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	13	27	41	54	66	79	100.0	100.0	99.5	93.5	67.9	30.4	5.3
Maintain prior outcome	0	12	28	42	55	67	80	100.0	100.0	99.4	92.4	64.7	27.4	4.3
Common centres	0	12	28	42	55	67	80	100.0	100.0	99.4	92.4	64.7	27.4	4.3
Stable common centres	0	12	27	42	55	67	80	100.0	100.0	99.5	92.4	64.7	27.4	4.3
Circle-arc equating	0	12	28	42	55	67	80	100.0	100.0	99.4	92.4	64.7	27.4	4.3
ISAWG	0	12	28	42	55	67	80	100.0	100.0	99.4	92.4	64.7	27.4	4.3

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 17: Summary of candidates included in common centre models (English A: Language and Literature HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	20,928	18,906	862	5,948	212
2018	24,742	22,075	988	7,412	260
2019	27,727	24,660	1,099	8,755	302

In this subject, the pattern is mixed. In 2017, one might surmise that the existing grade thresholds are harsh – compared to the alternative approaches modelled here. In 2017, the 'prior outcome' and 'common centres' approaches would give grade 6 boundaries two marks lower than the existing approach. In turn, this would provide five percentage points higher attainment at grade 6. However, in 2019, that pattern is reversed. The actual boundaries are one mark more lenient than the prior outcomes, common centres, and stable common centres methods. That single mark gives rise to three percentage points more candidates achieving grade 6 on the existing boundaries, than on the boundaries that the three noted methods would provide. Stable common centres again appears the most varied method, with more deviation away from the other methods.

5.5. DP English A: Literature HL (TZ1, May sessions)

This subject is an example of a 'shrinking subject', and as noted above it does see a real terms reduction in cohort size over time, largely as it shifts from this subject to other English subject offerings. The issues of cohort instability discussed above for growing subjects may therefore emerge here too, just due to cohort shrinkage – though it is worth noting that the subject is historically large and the shrinkage is relatively small. In short, we would expect the current maintain prior outcomes SRBs to fail to be valid due to the changing ability of the cohort over time.

Table 18: DP English A: Literature HL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	33	45	57	69	80	100.0	100.0	99.0	90.5	60.0	19.1	2.4
Maintain prior outcome	0	16	32	45	57	69	80	100.0	100.0	99.1	90.5	60.0	19.1	2.4
Common centres	0	16	32	45	57	69	80	100.0	100.0	99.1	90.5	60.0	19.1	2.4
Stable common centres	0	16	31	43	56	68	80	100.0	100.0	99.3	92.8	63.0	21.5	2.4
Circle-arc equating	0	17	33	45	57	69	80	100.0	100.0	99.0	90.5	60.0	19.1	2.4
ISAWG	0	16	32	45	57	69	80	100.0	100.0	99.1	90.5	60.0	19.1	2.4

Table 19: DP English A: Literature HL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	33	45	57	69	80	100.0	100.0	98.8	89.3	56.9	17.4	2.1
Maintain prior outcome	0	17	33	45	57	69	80	100.0	100.0	98.8	89.3	56.9	17.4	2.1
Common centres	0	17	33	45	57	69	80	100.0	100.0	98.8	89.3	56.9	17.4	2.1
Stable common centres	0	16	32	45	57	69	80	100.0	100.0	99.0	89.3	56.9	17.4	2.1
Circle-arc equating	0	17	33	45	57	69	80	100.0	100.0	98.8	89.3	56.9	17.4	2.1
ISAWG	0	17	33	45	57	69	80	100.0	100.0	98.8	89.3	56.9	17.4	2.1

Table 20: DP English A: Literature HL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	33	44	56	68	79	100.0	100.0	98.2	88.4	55.3	15.8	1.7
Maintain prior outcome	0	17	33	45	57	69	80	100.0	100.0	98.2	86.6	51.6	13.5	1.3
Common centres	0	17	33	45	57	69	80	100.0	100.0	98.2	86.6	51.6	13.5	1.3
Stable common centres	0	17	33	45	57	69	80	100.0	100.0	98.2	86.6	51.6	13.5	1.3
Circle-arc equating	0	17	33	45	57	69	80	100.0	100.0	98.2	86.6	51.6	13.5	1.3
ISAWG	0	17	33	45	57	69	80	100.0	100.0	98.2	86.6	51.6	13.5	1.3

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 21: Summary of candidates included in common centre models (DP English A: Literature HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	36,549	35,973	847	15,844	290
2018	36,063	35,602	834	16,348	309
2019	34,376	33,958	821	15,434	297

Of all the examples commented upon above, DP English A Literature HL TZ1 perhaps provides the most consistent set of boundaries. There is barely a mark's difference between any grade boundaries in any of the three years, though across the years there is some variation as to whether this is more severe or lenient than the actual boundaries. One factor which may cause this is that language papers have relatively few marks available. This relative coarseness might mean that

the various models find it harder to 'slip' onto suggesting the next mark available as a grade boundary, especially when combined with a large cohort.

Stable common centres differs slightly from the rest in 2017 and 2018. One single circle arc boundary is one mark higher than most other methods in 2017.

5.6. DP Global Politics HL (May sessions)

This subject is the example of a 'significantly growing subject'. As opposed to English A: Language and Literature, its consistent growth over time has been much more sizeable (as a proportion of the entry size in 2016). It therefore serves as a context that is different from the above one in nuanced but key way. Notably in 2016 the subject was a pilot with a small number of schools taking part in a fairly tightly controlled pilot of the subject.

Table 22: DP Global Politics HL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	23	35	48	61	74	100.0	99.9	99.1	94.2	72.6	33.5	7.9
Maintain prior outcome	0	10	21	32	44	58	70	100.0	99.9	99.5	96.5	81.1	41.8	14.1
Common centres	0	9	21	32	44	57	69	100.0	99.9	99.5	96.5	81.1	43.4	16.0
Stable common centres														
Circle-arc equating	0	9	21	33	46	60	73	100.0	99.9	99.5	95.7	77.0	36.1	9.0
ISAWG	0	10	22	33	46	58	71	100.0	99.9	99.2	95.7	77.0	41.8	12.1

Table 23: DP Global Politics HL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	12	23	35	48	61	74		100.0	98.7	93.8	74.4	37.9	7.6
Maintain prior outcome	0	11	24	35	48	59	71		100.0	98.7	93.8	74.4	41.9	11.4
Common centres	0	11	24	35	47	59	71		100.0	98.7	93.8	77.0	41.9	11.4
Stable common centres														
Circle-arc equating	0	10	22	34	47	60	73		100.0	98.8	94.3	77.0	39.9	8.4
ISAWG	0	12	25	37	48	60	71		100.0	98.3	92.2	74.4	39.9	11.4

Table 24: DP Global Politics HL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	23	35	47	60	72	100.0	99.9	98.9	93.3	72.2	35.3	7.9
Maintain prior outcome	0	10	22	34	47	59	71	100.0	99.9	99.1	94.1	72.2	38.2	9.4
Common centres	0	9	21	34	47	59	71	100.0	99.9	99.3	94.1	72.2	38.2	9.4
Stable common centres	0	10	22	34	46	58	71	100.0	99.9	99.1	94.1	74.8	41.5	9.4
Circle-arc equating	0	10	22	34	47	60	73	100.0	99.9	99.1	94.1	72.2	35.3	6.5
ISAWG	0	11	23	35	48	59	71	100.0	99.9	98.9	93.3	69.7	38.2	9.4

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 25: Summary of candidates included in common centre models (DP Global Politics HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	1,172	361	25	134	8
2018	1,674	1,244	82	403	21
2019	2,255	1,662	119	482	36

To some extent, the story here is of the subject 'settling down' and the various methods being more able to provide usable results. In 2017 and 2018, 'stable common centres' is unable to provide boundaries due to the small cohort size in 2016 and 2017.

All sessions' results suggest that the existing boundary setting may have been somewhat severe, compared to possible alternatives, but most so in 2017 and least so in 2019. Broadly circle arc produces the most different boundaries (in all cases more severe) relative to the other methods.

5.7. DP ITGS HL (November sessions)

This subject is the example of a 'suddenly growing subject'. As opposed to the above two growing subjects, its growth was experienced across two sessions (2017 and 2018) with a huge influx of new candidates (two waves of schools from one country joined the subject in these sessions), meaning a different approach might be more appropriate than for the more consistently growing subjects above. Notably this change was more than reversed in 2019 when these schools subsequently dropped back out, with it falling to around 75 per cent of its size in 2017 (or 50 per cent of its 2018 size).

Table 26: DP ITGS HL (November 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	12	24	38	49	59	70	100.0	99.9	99.0	84.3	40.0	7.7	0.4
Maintain prior outcome	0	12	24	37	47	56	66	100.0	99.9	99.0	85.6	48.3	13.7	0.9
Common centres	0	12	27	38	48	57	66	100.0	99.9	98.0	84.3	43.9	11.7	0.9
Stable common centres														
Circle-arc equating	0	10	23	37	48	57	68		100.0	99.0	85.6	43.9	11.7	0.6
ISAWG	0	12	27	40	50	59	69	100.0	99.9	98.0	77.3	35.6	7.7	0.5

Table 27: DP ITGS HL (November 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	43	53	63	73	100.0	99.9	99.2	81.7	39.9	7.6	0.8
Maintain prior outcome	0	13	29	43	52	63	73	100.0	99.9	99.1	81.7	45.0	7.6	0.8
Common centres	0	14	30	43	53	62	73	100.0	99.9	98.9	81.7	39.9	9.0	0.8
Stable common centres	0	14	29	43	53	65	76	100.0	99.9	99.1	81.7	39.9	4.9	0.3
Circle-arc equating	0	13	29	43	53	63	73	100.0	99.9	99.1	81.7	39.9	7.6	0.8
ISAWG	0	15	31	43	53	63	72	100.0	99.9	98.6	81.7	39.9	7.6	0.8

Table 28: DP ITGS HL (November 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	13	26	41	52	61	72		100.0	98.1	80.1	34.9	7.5	0.7
Maintain prior outcome	0	11	24	40	51	63	74		100.0	98.8	82.7	39.4	4.6	0.4
Common centres	0	11	25	41	52	62	73		100.0	98.5	80.1	34.9	6.1	0.7
Stable common centres	0	10	22	40	51	63	75		100.0	99.3	82.7	39.4	4.6	0.3
Circle-arc equating	0	11	26	41	51	61	71		100.0	98.1	80.1	39.4	7.5	0.8
ISAWG	0	11	25	41	52	63	74		100.0	98.5	80.1	34.9	4.6	0.4

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 29: Summary of candidates included in common centre models (DP ITGS HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	1,084	571	42	160	11
2018	1,597	1,022	51	449	19
2019	742	638	46	188	12

Overall the various methods produce quite a range of results for this subject. In 2017 common centres is lenient at 7 and severe lower down whilst maintain prior outcome and circle arc produce similar results (circle arc's more severe grade 7 notwithstanding) and ISAWG is severe across the board. In 2018 things are more similar across the methods, though stable common centres deviates most. In 2019 circle arc is lenient at 7 and stable common centres lenient at 3.

Consistently with this notion of 'suddenly growing', 'stable common centres' seems to work least well here; possibly on the grounds that the centres are not especially stable. This method delivers no results for 2017, and gives the highest grade 7 threshold for 2018, and 2019.

Interestingly, most other methods are reasonably in accordance with the 2018 and 2019 boundaries, suggesting that despite the changing cohort the standard was maintained reasonably well.

5.8. DP Film HL (M19 only)

This subject is the example of a 'changing curriculum and assessment model', which in this case occurred between M18 to M19. This represented a wholesale revision of the subject, though the curriculum manager provided us with a mapping of old to new components as follows:

- Textual analysis: compare to old Film presentation
- Comparative study: compare to old Independent study
- HL Collaborative film project: compare to old Production portfolio
- Film Portfolio: treat as a completely new component

However, it was also noted that the old Independent study component was unusual in terms of the cohort's performance – so we would not necessarily expect it to predict outcomes in the Comparative study perfectly.

Because the subject is completely new, the conventional methodology for all of our approaches fails to function, as they rely on (in varying ways) benchmarking each component against outcomes in the reference year's instance of that component. Because this does not exist, a substitute must be used, or our methods fail to function correctly. There are several options for alternative benchmarks that can be used such instances:

a. All results from a specific other component in the reference year

b. All results from a specific subject in the reference year

c. All results from a specific group of subjects in the reference year

d. All results from all subjects in the reference year

Broadly speaking, from a-d these are less specific and more general comparators. We would normally prefer a more closely related comparator component or subject to benchmark against, rather than defaulting to 'the entire suite's results' (which effectively means setting the component's standard at the suite's average), but in instances of completely new subjects some of the more specific options may not be possible.

Luckily, in this subject's case, there is a clear mapping to a prior component for three of the components, meaning option a. can be used and the prior component's results used as the benchmark for three of the new components. For the fourth component however (the Film Portfolio) there is no old component equivalent to it, so we will be defaulting to option b. and using the prior Film specification's outcomes as a benchmark. This amounts to setting the standard for this component at the average of the three prior components' previous standards[9]. This mapping to derive a prior benchmark is technically workable for all approaches we modelled, and should be borne in mind when interpreting results.

Table 30: DP Film HL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	39	53	68	82	100.0	99.7	93.5	78.3	47.1	17.8	3.8
Maintain prior outcome	0	11	28	40	51	66	77	100.0	99.8	93.5	76.4	51.6	20.8	7.0
Common centres	0	11	28	38	50	65	77	100.0	99.8	93.5	79.9	54.3	22.1	7.0
Stable common centres	0	11	28	41	54	67	80	100.0	99.8	93.5	74.7	44.5	19.5	4.7
Circle-arc equating	0	12	28	40	54	65	78	100.0	99.8	93.5	76.4	44.5	22.1	6.3
ISAWG	0	11	28	40	51	65	77	100.0	99.8	93.3	76.2	51.5	22.1	7.0

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

[9] The prior Film specification had equal weightings between the three components.

Table 31: Summary of candidates included in common centre models (DP Film HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2019	2,534	2,223	252	412	37

The methods give relatively similar results, with most variation at grade 4 (common centres being lowest and stable common centres highest) whilst stable common centres gives the highest mark at grade 7, making it the most severe method modelled. This distinction between stable common centres and common centres is notable; the inclusion or exclusion of some centres in the prior outcome we attempt to maintain can clearly have a significant impact on the outcome of the model.

5.9. DP Sports Exercise & Health Science (SEHS) HL (M18 only)

This subject is the example of a completely new level of a subject with no similar prior standard to equate back to (as was the case for Film above). In SEHS' case the new specification was introduced at HL in M18, so this is the session of interest for this awarding context.

As discussed for Film above, our conventional methodology fails to function when there is no direct comparator for each component to benchmark the standard against – nor is there a previous subject to use as the benchmark. As such, in this instance we use an amalgam of several similar subjects' results in the reference year to benchmark the standard of this new subject, due to lack of a more closely related benchmark.

Because SEHS is a group 4 (science) subject, the three largest science subjects (Biology, Chemistry and Physics)' reference year outcomes were aggregated and used as the benchmark (candidates sitting multiple retained their multiple results, via some being modified so each was treated as if they were the result for a unique candidate). This means that SEHS' outcomes are dictated by a) how many of each grade these three subjects awarded in the reference year, and for some methods, b) how different the cohort in SEHS was to the cohort sitting these three subjects in the reference year. Ultimately, this means that SEHS' standard will be set at an average of the standard for the three largest sciences.

Again, we attempt to model each SRB setting method for SEHS, albeit with an unusual benchmark. Unlike for Film however, stable common centres is *a priori* unlikely to function in SEHS' case, as the number of entrants from a centre for Biology, Chemistry and Physics is likely to considerably exceed the number of entrants from the same centre for SEHS, meaning most centres are likely to fall foul of the method's check on centres having similar entry size in the reference and current sessions.

Table 32: DP SEHS HL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	16	31	42	53	65	76		100.0	96.4	88.4	63.3	34.3	12.4
Maintain prior outcome	0	18	37	48	58	69	79	100.0	99.7	92.2	74.7	50.8	25.3	8.1
Common centres	0	16	34	45	55	66	76		100.0	94.2	82.5	58.1	31.1	12.2
Stable common centres														
Circle-arc equating	0	17	35	47	58	69	80		100.0	93.3	76.9	50.8	25.3	6.7
ISAWG	0	19	38	50	61	71	81	100.0	99.7	91.7	69.3	42.8	21.5	4.7

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 33: Summary of candidates included in common centre models (DP SEHS HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2018	362	362	65	0	0

This – arguably somewhat extreme / difficult to model – example does show some limitations. More 'conventional' approaches ('prior outcomes' and 'common centres') appear to get closest to replicating the actual boundaries, even using the amalgam of Biology, Chemistry and Physics as the baseline. Circle arc equating and especially ISAWG are harsher – requiring four and five more marks to achieve the highest grade respectively. Notably the maintain prior outcomes method does produce similar results to the ISAWG, if slightly less severe.

However, in this case the unusual baseline is likely the result of this difference – relative to the cohort sitting the three large sciences, the ISAWG scores imply that the cohort sitting SEHS is less able, meaning they should be allocated fewer high grades in order for its standard to be comparable to the other sciences. Similarly for circle arc the mean mark for SEHS shifting relative to that of the other sciences is likely what causes the observed shift in grade boundaries.

5.10. MYP Mathematics (November sessions)

This subject is an example of 'a new cohort in an existing subject'. This is not too dissimilar to the changing entry examples above, and is therefore another instance where we would expect the current SRB setting procedure to fail to account for possible changes in the cohort's ability level.

A large reason for this subject's inclusion despite its somewhat similar awarding context is in order to have modelled an example of an MYP (as opposed to DP like all the other subjects) assessment – which are fundamentally different from DP in that they are based only on a single component. Note that because MYP is a different suite to the DP, we did not model an ISAWG approach (as to do so would require a full repeat of the ISAWG modelling for just this one subject). It is worth noting that the MYP was very new in N17 (the first MYP award being in 2015) and that the cohort for this subject has only two common centres. Further, there are retakes which vary in quantity from year to year in the November session. This may influence the stability of some methods, but renders it useful to model as somewhat of a volatile context.

Table 34: MYP Mathematics (November 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	22	33	45	56	68	100.0	87.6	61.6	43.8	25.2	15.7	6.6
Maintain prior outcome	0	5	11	19	34	48	63	100.0	96.3	87.6	69.4	40.5	21.5	9.1
Common centres	0	5	11	20	34	49	63	100.0	96.3	87.6	66.5	40.5	21.1	9.1
Stable common centres	0	6	12	21	35	48	62	100.0	95.5	84.7	63.2	38.4	21.5	10.3
Circle-arc equating	0	8	17	25	34	43	52	100.0	93.8	73.6	54.1	40.5	26.9	18.6

Table 35: MYP Mathematics (November 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	22	33	45	56	68	100.0	91.9	62.8	43.0	26.7	12.8	2.3
Maintain prior outcome	0	11	22	33	42	52	61	100.0	91.9	62.8	43.0	32.6	15.1	5.8
Common centres	0	16	33	43	49	54	60	100.0	79.1	43.0	31.4	22.1	14.0	8.1
Stable common centres	0	21	42	45	50	56	61	100.0	62.8	32.6	26.7	18.6	12.8	5.8
Circle-arc equating	0	11	22	33	45	56	68	100.0	91.9	62.8	43.0	26.7	12.8	2.3

Table 36: MYP Mathematics (November 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	12	24	37	48	59	70	100.0	98.0	76.5	47.7	30.2	17.4	7.4
Maintain prior outcome	0	15	31	40	53	65	78	100.0	96.0	60.4	42.3	22.1	12.8	2.0
Common centres	0	17	35	44	54	65	75	100.0	91.3	53.0	34.9	21.5	12.8	4.7
Stable common centres														
Circle-arc equating	0	14	28	40	51	63	74	100.0	97.3	69.1	42.3	25.5	14.1	4.7

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 37: Summary of candidates included in common centre models (MYP Mathematics)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	242	216	6	200	<5
2018	86	47	7	14	<5
2019	149	34	10	<10	<5

This subject shows a slightly unusual shift in that the modelled boundaries appear to be lower than the actual boundaries in 2017, whereas they are higher in 2019. 2018 is more mixed – modelled boundaries are higher overall, but the grade 7 boundary is a persistent exception. The gaps are substantial, too; eight marks for grade 7 between 'maintain prior outcomes' and actual in 2019, and seven for the same boundary in 2018.

This shift stems from the awarded standard shifting over time, as can be observed considering the 'maintain prior outcome' method with reference to the prior year's actual outcomes. In each case this method matches the prior outcome well – it's just the case that the actual outcomes do not. With the caveat that we lack substantive evidence to this point, the large differences between the actual and modelled boundaries over the three years may point to a challenge to the maintenance of standards over time in the actual boundaries for this subject.

In terms of method-to-method differences, in 2017 the main notable point is that circle arc results in marked boundary compression into the centre of the mark range – the 3 boundary is much higher and the 7 boundary much lower than other methods. In 2018, both common centres methods produce substantially different (more severe) lower grade boundaries than the other methods,

whilst circle arc produces the most severe grade 7 boundary. In 2019, circle arc has similar lower boundaries to maintain prior outcome, and a similar grade 7 boundary to common centres, but overall matches the actual boundaries most closely.

In this subject the substantial variation in results between methods means it is very challenging to suggest which standard is the 'correct' one. With the outcomes changing so much year on year the cohort's stability cannot be assured, and with relatively low entry sizes common centres becomes problematic.

5.11. DP Theatre HL (May sessions)

This subject is an example of 'verification model' subject. In other words, all components are coursework as opposed to examinations. Whilst, per the method section above, the IB usually holds coursework components' grades constant, which we reflect in our modelling to mimic these components' unchanging boundaries over time, in a purely coursework subject this approach must be adjusted. As such we treat each component 'as if it were an examination component' and model new boundaries for it with each method, deviating slightly from the 'normal' awarding process for this subject so as to include an entirely coursework subject in the analysis.

As such, it is relatively straightforward to apply all our other models for SRB setting to this awarding context. It is of particular interest whether the nature of this subject as a verification only model means any approaches work better or worse.

Table 38: DP Theatre HL (May 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	10	21	33	49	65	82	100.0	99.6	96.2	85.1	61.7	33.0	9.8
Maintain prior outcome	0	8	20	32	48	65	81	100.0	100.0	96.6	86.7	63.3	32.9	10.4
Common centres	0	7	19	31	47	64	81	100.0	100.0	97.2	87.7	64.7	34.6	10.4
Stable common centres	0	8	21	32	48	64	79	100.0	100.0	96.2	86.7	63.3	34.6	12.5
Circle-arc equating	0	8	21	33	48	66	82	100.0	100.0	96.2	85.1	63.3	31.1	9.8
ISAWG	0	8	20	32	48	64	79	100.0	100.0	96.6	86.6	63.2	34.6	12.5

Table 39: DP Theatre HL (May 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	10	21	33	49	65	82	100.0	99.9	96.6	85.9	61.4	34.0	10.1
Maintain prior outcome	0	10	22	33	50	65	82	100.0	99.9	96.1	85.9	59.8	33.9	10.1
Common centres	0	8	20	32	48	64	80	100.0	99.9	97.2	86.9	62.5	35.3	11.8
Stable common centres	0	8	20	31	47	62	78	100.0	99.9	97.2	87.7	63.9	37.9	14.4
Circle-arc equating	0	9	22	34	50	65	82	100.0	99.9	96.1	84.1	59.8	33.9	10.1
ISAWG	0	8	20	31	48	65	82	100.0	99.9	97.2	87.7	62.5	34.0	10.1

Table 40: DP Theatre HL (May 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	10	21	33	49	65	82	100.0	99.7	96.7	86.2	61.8	34.0	9.3
Maintain prior outcome	0	9	21	34	50	64	81	100.0	99.8	96.7	84.5	60.0	35.2	9.9
Common centres	0	9	21	33	49	64	81	100.0	99.8	96.7	86.2	61.8	35.2	9.9
Stable common centres	0	7	17	31	47	63	79	100.0	99.9	98.4	88.1	64.8	36.6	12.2
Circle-arc equating	0	9	21	33	49	65	81	100.0	99.8	96.7	86.2	61.8	34.0	9.9
ISAWG	0	10	22	35	50	65	81	100.0	99.7	96.2	83.4	60.0	34.0	9.9

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 41: Summary of candidates included in common centre models (DP Theatre HL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	2,538	2,213	406	515	88
2018	2,694	2,315	420	525	84
2019	2,602	2,296	435	507	101

There are few major differences between boundary positions in this subject. Stable common centres is generous in 2019 and 2018, and common centres generous in 2017. Maintain prior outcome and circle arc are very similar and result in the most severe boundaries of all the models across the years, though ISAWG is also quite severe in 2019.

Overall however, there is certainly again a trend, as in other subjects, that identical or more lenient boundaries are set across the board by most methods.

5.12. DP Chinese B SL (November sessions)

This subject is an example of a subject with 'skewed mark distributions'. The cohort entering this subject are often extremely high ability, leading to very negatively skewed mark distributions with very few candidates falling in the lower grades. This is a known challenge for some SRB-setting approaches, so is a key context to model in order to flag which are likely to fall down in such situations.

Table 42: DP Chinese B SL (November 2017) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	44	57	70	83				100.0	99.3	93.4	80.6
Maintain prior outcome	0	12	27	42	56	68	83				100.0	99.3	94.6	80.6
Common centres	0	12	27	42	56	68	83				100.0	99.3	94.6	80.6
Stable common centres	0	13	27	38	53	66	81				100.0	99.3	96.3	82.6
Circle-arc equating	0	12	27	43	57	70	84				100.0	99.3	93.4	77.6
ISAWG	0	13	29	43	56	68	81				100.0	99.3	94.6	82.6

Table 43: DP Chinese B SL (November 2018) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	28	46	60	72	86			100.0	99.9	98.3	93.5	82.0
Maintain prior outcome	0	13	28	46	59	72	85			100.0	99.9	98.4	93.5	82.6
Common centres	0	13	28	46	58	72	85			100.0	99.9	98.6	93.5	82.6
Stable common centres	0	11	25	38	52	65	79				100.0	99.7	97.6	89.1
Circle-arc equating	0	14	32	47	59	72	85			100.0	99.9	98.4	93.5	82.6
ISAWG	0	12	26	45	58	71	85				100.0	98.6	94.4	82.6

Table 44: DP Chinese B SL (November 2019) – subject level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	14	29	44	59	71	86			100.0	99.8	98.5	94.4	80.7
Maintain prior outcome	0	14	30	45	59	72	86			100.0	99.8	98.5	93.2	80.7
Common centres	0	14	29	44	58	71	85			100.0	99.8	98.6	94.4	82.0
Stable common centres	0	13	28	41	54	68	81			100.0	99.9	99.0	95.7	85.7
Circle-arc equating	0	13	28	46	59	72	85			100.0	99.8	98.5	93.2	82.0
ISAWG	0	15	33	47	59	72	84			100.0	99.8	98.5	93.2	83.2

Component-level results for this subject can be found here. A brief summary of the number of candidates sitting the subject each year, and the number of common centres (and candidates therein) can be found in the table below.

Table 45: Summary of candidates included in common centre models (DP Chinese B SL)

Year	No. cands	Common centres		Stable common centres	
		No. cands	No. centres	No. cands	No. centres
2017	1,073	1,064	35	717	12
2018	1,146	1,134	38	620	8
2019	1,235	1,198	40	712	10

This subject is remarkably stable given the low number of candidates in the lowest grades, with little difference between the methods. Stable common centres consistently requires learners to achieve slightly fewer marks in order to achieve the top grades – potentially because two particularly large common centres make up almost half the overall cohort, so they will have a substantial impact on the models. This difference is as much as six marks for grade 7 in 2018. ISAWG is arguably next most different, and is differently severe or lenient at different judgmental grades within the same year.

5.13. Summary of trends across subjects

With each individual subjects' results summarised, this section seeks to draw, across subjects, a brief summary of the trends observed for each SRB setting approach modelled. This is essentially the other way of 'cutting' the analysis; by method rather than by subject.

5.13.1. Maintain prior outcome

In many cases, the 'prior outcomes' this method seeks to mimic are actually quite different from the actual outcomes resulting from the grade boundaries set in practice. For example, in the case of Maths, Swedish A: Literature, English A: Literature and ITGS, where there was a year-on-year drop in the actual percentage of candidates attaining top marks, maintain standards is more lenient than the actual standard as it attempts to maintain the same percentage of pupils attaining each grade as in the previous year [10]. This indicates that the expert and teacher information feeding into the boundary setting process in practice has generally had a slight inflationary effect on the grade boundaries relative to just maintaining the prior outcome stringently, suggesting that their moving the boundaries away from the SRBs is down to a perception that the cohort is not unchanging over time, but is increasingly able.

Overall, this method is one of the most consistent with the actual grade boundaries, however.

5.13.2. Common centres

Whilst overall common centres is a method that tends to result in similar outcomes and grade boundaries to 'maintain prior outcome', it does sometimes deviate from it. This is not too surprising, as it uses a potentially small subset of the cohort as our presumed 'unchanging benchmark of ability'. This is evident in that the method can become clearly unstable or unusable entirely for small cohort subjects.

5.13.2.1. Stable common centres

Stable common centres is similar to common centres, but is slightly more extreme in that it is fairly often a large outlier method (i.e. the one with the most severe or most lenient boundaries). We attribute this to sampling error due to the method using a lower number of candidates relative to the other methods, though the only way to get to the bottom of it is likely another simulation study where we fabricate student performance based on generated student abilities and exam item difficulties and see how the method fares. Outlying results often occur in the smaller subjects, where the method quickly becomes unstable. Unlike ISAWG, it is not consistently outlying in either direction, which does cast some doubt on its reliability.

5.13.3. Circle arc equating

Circle arc proves interesting in that it is the most robust of the methods to very small sample sizes, being possible in cases where the others fail to operate due to lack of data. This is obviously a key benefit.

However, whilst it sometimes marries up with the other methods, it is also often the method most 'out of step' with the others. Whilst few of the differences in boundaries it suggests are large in absolute terms, there is still a clear trend that in several subjects circle arc is suggesting a boundary quite different to the other methods. Given this method lacks theoretical underpinning and is simply something that 'mimics equipercentile equating's behaviour' these intermittent deviations from the apparent trend is slightly alarming.

[10] One possible factor which may cause this is that the data we model is post-EUR which will generally result in an increase in marks for some candidates. Given higher marks, lower boundaries then need to be set to maintain prior outcomes.

5.13.4. ISAWG

ISAWG commonly produces extremely similar boundaries to many other methods, which is perhaps surprising given how different its computation is. As the only method with an 'external' indicator of performance (concurrent attainment) it is reassuring that it does not suggest radically different results – this implies that the other methods suggest broadly similar standards as a method accounting for concurrent attainment does.

Where slight differences from the other methods do emerge, it is interesting that these are not consistently lenient or severe (relative to other methods). Commonly they are actually severe at one end of the mark scale and lenient at the other, relative to other methods (but again this varies from subject to subject with some showing leniency at the top and severity at the bottom, and others the inverse). This suggests that ISAWG can result in somewhat of a different 'slope' to the equating relationship than other methods, potentially as a result of its prediction matrix approach taking account of candidates in each ability decile.

6. Discussion

The main aim of this research is to provide guidance for the IB on which methods are likely to be most suitable in which awarding contexts. There are two key stages here – firstly, what methods are possible in what contexts, and secondly, what methods are advisable/best suited to which contexts.

It is important to note that there is no 'correct' set of SRBs. SRBs do not care what the truth of the matter of candidate performance is, they just take a set of input data and process it in a pre-specified manner. A method where the SRBs deviate from those produced by the other methods may superficially appear questionable, but this is more likely to be down to variation in the methodology and other methods being more related to one another. This makes establishing the 'best' method for a given context challenging – what is easier to do is suggest in what circumstances a method may not be appropriate.

We will consider these questions generally initially, then reflect on what the most appropriate method of SRB setting is for each context.

6.1. What methods work in which contexts?

In terms of the first point, when it is possible to run each method can be seen summarised in the following table – red denotes instances the method failed to run entirely. Cohort sizes included in the subject level modelling are shown for reference.

Table 46: Summary of method useability

Subject	Year	No. cands	Method viability				
			Maintain prior outcome	Common centres	Stable common centres	Circle-arc equating	ISAWG
DP Mathematics SL	2017	23,168					
	2018	24,980					
	2019	26,031					
DP Armenian A: Literature SL	2017	27					
	2018	21					
	2019	29					
DP Swedish A: Literature SL	2017	248					
	2018	259					
	2019	199					
DP English A: Literature HL	2017	20,928					
	2018	24,742					
	2019	27,727					
DP English A: Literature HL	2017	36,549					
	2018	36,063					
	2019	34,376					
DP Global Politics HL	2017	1,172					
	2018	1,674					
	2019	2,255					
DP ITGS HL	2017	1,084					
	2018	1,597					
	2019	742					
DP Film HL	2019	2,534					
DP SEHS HL	2018	362					
MYP Mathematics	2017	242					
	2018	86					
	2019	149					
DP Theatre HL	2017	2,538					
	2018	2,694					
	2019	2,602					
DP Chinese B SL	2017	1,073					
	2018	1,146					
	2019	1,235					

The key relevant facts from this table can be summarised as follows:

- Stable common centres frequently fails to run – in about a third of cases modelled. With small cohort sizes or changes in entry it is more likely that there are insufficient stable common centres to permit the method to operate.
- Most other methods run in all cases bar Armenian A: Literature SL, which has entry sizes that are extremely small (under 30).
- Circle-arc equating continues to function even in this very small subject – every mark distribution has a mean, after all.

This could be taken to demonstrate that in the very smallest subjects, circle-arc equating is likely the best option given it is the only method that properly functions. However, is important to note that whilst it operates differently mechanically, this method essentially mimics an equipercentile

equating method (whereby the mark the Xth percentile of candidates achieve in each sitting are considered comparable). In turn, any equipercentile method aims, in effect, to maintain the previous standard – just by mapping each mark rather than just grade boundaries via this equating.

As such, there is perhaps marginal benefit over and above simply persisting with 'maintain prior standard' from applying circle arc equating in very small subjects (<30 to 50 candidates) – as it is not a method that solves that methods' issue of not accounting for increases or decreases in the cohort's ability over time. Arguably the main reason to institute it would be to provide a statistical SRB in cases such as Armenian A: Literature, where there are no candidates in the grade 3 to 4 region, preventing a strict 'maintain prior outcome' approach from setting SRBs for these grades. IB currently defaults back to previous years' grades in such circumstances, perhaps with an adjustment if there is a clear 'mean shift' in performance occurring.

We can therefore characterise the main benefit of shifting to circle-arc equating for very small subjects in terms of providing a codified and objective set of SRBs for grade boundaries that this might not be possible for via a 'maintain prior outcome' approach.

6.2. What methods are advisable in what contexts?

Aside from the very smallest subjects where circle arc appears the only viable choice, it is still 'all to play for' insofar as the other methods produce results in the remaining contexts – with the exception that stable common centres often does not function. Based on Table 46 it appears that this tends to be the case in subjects with fewer than around 500 entries (with the exception of the "significantly growing" cohort in global politics, where it is not possible with over 2,000 candidates). This is shown in the table below for information (along with how many candidates are in those stable common centres). Red text denotes cases for which the method could not be used.

Table 47: Summary of candidates included in common centre models

Subject	Year	No. cands	Common centres		Stable common centres	
			No. cands	No. centres	No. cands	No. centres
DP Mathematics SL	2017	23,168	21,543	1,069	8,054	330
	2018	24,980	24,114	1,182	8,886	346
	2019	26,031	24,980	1,267	9,962	386
DP Armenian A: Literature SL	2017	27	26	6	17	<5
	2018	21	17	5	<10	<5
	2019	29	26	5	<10	<5
DP Swedish A: Literature SL	2017	248	195	32	41	9
	2018	259	215	31	40	10
	2019	199	169	29	23	6
DP English A: Language and Literature HL	2017	20,928	18,906	862	5,948	212
	2018	24,742	22,075	988	7,412	260
	2019	27,727	24,660	1,099	8,755	302
DP English A: Literature HL	2017	36,549	35,973	847	15,844	290
	2018	36,063	35,602	834	16,348	309
	2019	34,376	33,958	821	15,434	297
DP Global Politics HL	2017	1,172	361	25	134	8
	2018	1,674	1,244	82	403	21
	2019	2,255	1,662	119	482	36
DP ITGS HL	2017	1,084	571	42	160	11
	2018	1,597	1,022	51	449	19
	2019	742	638	46	188	12
DP Film HL	2019	2,534	2,223	252	412	37
DP SEHS HL	2018	362	362	65	0	0
MYP Mathematics	2017	242	216	6	200	<5
	2018	86	47	7	14	<5
	2019	149	34	10	<10	<5
DP Theatre HL	2017	2,538	2,213	406	515	88
	2018	2,694	2,315	420	525	84
	2019	2,602	2,296	435	507	101
DP Chinese B SL	2017	1,073	1,064	35	717	12
	2018	1,146	1,134	38	620	8
	2019	1,235	1,198	40	712	10

It is apparent from the table that in some cases, common centres is nearly identical to 'maintain prior outcome' in that almost all candidates are from the same centres as were present in the reference year. This explains why the methods are so often closely aligned in terms of SRBs. It is only really the growing subjects where there is a substantial gulf between the number of candidates included in each approach.

Interestingly, the red highlighted cells show that there is almost always data for stable common centres in the cases the method fails. This is attributable to the fact that the method in its purest form does not function when there are not a range of candidates across all the judgemental grades, as this results in it being unable to allocate these key SRBs (the same reason maintain prior outcome failed for Armenian A: Literature). It would be possible to add in the current IB fall-back of maintaining the prior boundary in such cases in a live session, however[11].

It's also interesting to note that the number of centres included in stable common centres modelling can vary quite substantially from year to year, within the same subject. This perhaps violates an implicit assumption that there is a consistent group of centres within a subject which always achieve stable, similar results, that can be used as a baseline for such modelling. It may also indicate why stable common centres is often the most volatile-looking model in our results within a given subject, as the profile of centres included from year-to-year may not actually be particularly similar. It also suggests that it may be worth giving some further thought to the precise definition of a 'stable common centre' in the IB context.

Notably stable common centres, in some cases when it does function, produces quite different results to the other methods – suggesting that when there are very few such centres to use as a baseline, its results can be biased and not reflect the changing ability profile of the whole subject cohort adequately, as the centres included could quite easily be radically different to the overall cohort. This is somewhat of a 'wisdom of crowds' line of reasoning however, in that we are presuming that the method which produces most outlying results must be invalid in some manner, rather than all the other methods being problematic.

In any case, it is certainly the case that stable common centres is likely to only be workable in some IB contexts. Given the (presumed) aim of minimising non-essential divergence in SRB-setting approaches (to take this to extremes, we do not want to recommend a different approach for every single context for example), it may be justifiable to rule stable common centres out as an appropriate method for the IB on these grounds, as a third 'mid-sized subject' method would be required (in addition to circle-arc for very small subjects).

6.2.1. ISAWG vs common centres

Broadly this leaves common centres and ISAWG as the more plausible options (given that 'maintain prior outcome' is the baseline current method IB wishes to improve upon). Given this decision, it is worth restating the key assumptions of these methods for purposes of comparison.

- Common centres relies on the assumption that, as a whole, candidates in the same centres from year to year will achieve the same spread of grades. In other words, that there is no change in their ability from year to year. This means that it is only the movement of

[11] Though note that this can fall down if the reference year boundary is no longer outside the span of the boundaries the method can set! An edge case rule for such instances would be needed.

candidates *outside these common centres* that can permit an increase or decrease in the percentage of top grades awarded for the subject.

- ISAWG relies on an external indicator of performance (the ISAWG metric) to determine whether the cohort in the current year is more or less able than the cohort in the prior year, and determines how many top grades to award based on this. As such exactly how this ISAWG metric is derived is crucially important; given that under our method it uses 'stable common centres' to equate between years, the above assumption for common centres applies here too – i.e., that between years, the performance of candidates in these centres remains identical.
    - Given the implementation of ISAWG, this amounts to an assumption that we should give out more/fewer top grades in the current year *according to how much the average rank order of component performance achieved by candidates in the stable common centres shifts from year to year* – which can be caused by either the performance of the candidates in the common centres rising or falling (relative to the rest of the cohort) or by movement of candidates into or out of the suite.
- ISAWG also relies on the assumption that the relationship between general academic ability (the ISAWG score) and score on the component is consistent from one year to the next. In other words, that in both years there is a similar trend between the two. This means that if there is a sudden influx of candidates who are strong overall but particularly poor at a component, this relationship can be thrown off introducing bias into the method.
- ISAWG further relies on the assumption that the subject to subject (and component to component) relationship in performance is stable year-on-year. In other words, that a particular subject does not suddenly have a much stronger or weaker relationship with another from one year to the next. If this is not the case, then it will affect the validity of the equating of one year's score to the next.

Considering these will help us establish whether there are circumstances in which one or the other method is more or less appropriate.

6.2.1.1. Common centres

Because common centres only deals with grade outcomes (just for a subset of the cohort) a problem can arise when the candidates in the pool of common centres are not equally able from year to year. If they improve, then the method will not reward this; if they grow weaker the method will not make grading more severe to compensate. A set proportion of candidates in these centres will achieve each grade each year.

As such, in IB contexts where this is unlikely to be the case (perhaps new subjects centres are likely to be improving at delivering, or those with only a small number of common centres where year-on-year cohort volatility could result in differing ability profiles) common centres may be problematic.

6.2.1.2. ISAWG

Given the complex functioning of the ISAWG approach, it is useful to check what the average ISAWG score of the candidates in stable common centres is across the years to see if the assumption of stability over time holds. The below table shows, for the stable common centres used to equate the ISAWG score from series to series, what the mean and SD of ISAWG scores is. Note that in order to not have both years within an equating pair have identical values, the 'current year' values are pre-equating so that we can see the difference between the two prior to equating being carried out.

Table 48: Mean ISAWG scores for stable common centres (across whole DP suite)

Year pair	Year	Equating status	Mean	SD
2016-2017	2016	Unaffected	0.011	0.991
	2017	Pre-equating	0.017	0.989
2017-2018	2017	Unaffected	0.012	1.033
	2018	Pre-equating	0.029	0.981
2018-2019	2018	Unaffected	-0.012	1.060
	2019	Pre-equating	0.006	1.011

Table 49: Number of Schools and pupils in each equating step

Year Pair	Number of Schools	Reference Year Pupil Count	Current Year Pupil Count	Reference Pupil Count Percentage of Total	Current Year Pupil Count Percentage of Total
2016-2017	783	97,442	98,003	67%	63%
2017-2018	853	105,736	106,099	68%	65%
2018-2019	888	110,680	108,558	68%	66%

The key point to note from the table above is that within an equating pair, the 'current' year's candidates in these stable common centres always have a higher average ISAWG score than the reference year (and a lower SD of said scores). This reflects that the candidates in these stable common centres are creeping up the rank order of candidates, meaning ISAWG will equate these two non-identical rank orders of performance in the reference and current year (i.e. 50th percentile in one year to 55th in the second).

There are a few possible explanations for why this could be occurring. Firstly, it may be due to this cohort improving in absolute terms, which would have to happen in tandem with the remaining candidates in other centres not improving to impact the rank order. How probable it is that candidates in other centres improve less than those in the stable common centres is up for debate, of course. The other possibility is that the level of absolute performance in the stable common centres remains identical, but the other candidates have shifted down such that their rank order position has risen. An influx of many low ability candidates could result in this, as one example.

In order to disentangle what is occurring, we also need to consider, across the whole suite, what the (post-equating) mean ISAWG score is per learner. These scores are post-equating so are all on the same scale to allow for comparability.

Table 50: Mean ISAWG scores for whole DP cohort

Session	Overall cohort	
	Mean	SD
2016	-0.002	1.001
2017	-0.019	1.040
2018	-0.027	1.053
2019	-0.022	1.085

From the table, we can see that the average ISAWG score for the whole cohort drops from session to session with the exception of 2018-19, which rises slightly. The SDs continue to widen slightly from year to year as well. In light of the prior table's findings on stable common centre ISAWG scores, the drops in cohort ability from 2016-18 suggest that the 'non-stable common centre' candidates are on average weaker in 2017 than 2016, and weaker in 2018 than 2017.

In 2018-19 however, the overall cohort has grown slightly stronger, but the stable common centres having higher ISAWG scores means that whilst the cohort as a whole is more able, the stable common centres have improved relatively 'more', leading to them creeping up the rank order of candidates.

Whilst, based on the above tables, we can see the rank order changes in stable common centres in from year to year, it is entirely open to debate whether this represents a 'problem' or not due to failing to reward or penalise real changes in absolute performance within the matched cohort. Ultimately, it would take a comparison of performance on secure common tasks (i.e., a reference test) to establish this definitively. In the absence of this we can only state the observed change in rank order exists.

Returning to the question at hand though – whether the assumptions for ISAWG affect when it can be used – the assumption of stability in the stable common centres from year to year in ISAWG is extremely similar to that of stability in common centres in the basic common centres method. In each case, the candidates in a subset of the data are held constant in terms of the grades they are allocated – implicitly assuming therefore, that they are of comparable ability year-on-year. The main difference is that ISAWG does this at the whole suite level, thus being based on considerably more data, whilst the common centres approach is internal to a given component[12]. This being done at subject level is also what allows it to, at component level, factor in changes in ability to tweak the grades it doles out. It's also much more likely that across the schools sitting the entire suite and every component sat the assumption of stable ability from one year to the next holds.

Further, the IB contexts where this assumption might prove problematic for common centres (new subjects centres are likely to be improving at delivering, or those with only a small number of common centres where year-on-year cohort volatility could result in differing ability profiles) are at the subject level – ISAWG's approach of carrying out the equating at the whole suite level addresses these concerns largely caused by small quantities of data.

Our judgement would be that this renders ISAWG the most methodologically robust method of SRB-setting – at least in the 'end of session' scenario we were modelling here. It is interesting, therefore, that it produces extremely similar results to many other methods. This begs the question of whether, pragmatically, pursuing ISAWG is 'worth it' over and above a method such as common centres, which accounts for cohort ability shifts in a different way. However, given the fact that in IB's case common centres typically amount to 'nearly the entire cohort', a common centres approach is going to be identical to 'maintain prior outcome' except in contexts with substantial cohort change from year to year. Given that IB's aim was to consider moving away from this

[12] We would note, however, that it is definitely worth any further investigation into ISAWG considering in depth exactly what the most appropriate rules are to define the centres used as the 'stable common ones' to equate between series.

approach as it is problematic in cases of cohort ability change, this could also be considered an issue.

6.3. Context summary

6.3.1. Large, stable subjects

In these subjects, maintain prior outcome is a perfectly adequate method, because the cohort's ability is not changing over time. The slight issue is that when using this method, it would never be noticed if the cohort's ability does happen to change over time.

All other methods should work adequately too, though we noted that in practice circle arc equating does produce somewhat divergent results and it is likely less advisable as a result.

6.3.2. Small subjects

In IB's smallest subjects (<30-50 candidates), circle arc equating is the only method that functions properly. However, the ISAWG could be adjusted to function, if an approach similar to that adopted in SEHS (i.e., using another subject's data as the reference dataset) was utilised. Given this was not modelled here, we have no data on how such an approach would perform.

In slightly larger subjects of around 200-250 candidates however, common centres and ISAWG quickly become very stable and look to be just as viable options. Stable common centres remains problematic in subjects of this size, however, due to the very small number of candidates included in it. A key thing to note here is that even small subjects can still validly utilise most methods, as long as their cohorts are relatively stable.

6.3.3. Growing/shrinking subjects

Growing or shrinking subjects represent one of the greatest violations of the 'maintain prior outcomes' approach's assumptions – that the cohort will remain static in ability over time. As such other methodologies offer potential for a much more appropriate SRB-setting approach.

6.3.3.1. Gradual growth

In large subjects experiencing gradual growth or shrinkage such as the English Literature and English Language & Literature modelled, almost every method produces near identical results in our modelling. The exception is stable common centres, which via 'wisdom of crowds' logic is potentially therefore slightly suspect.

6.3.3.2. Sudden/significant growth

In subjects experiencing more sudden or significant growth stable common centres is prone to cease to function due to little data being included in it. Similarly, the pool of common centres is likely to be quite markedly reduced as shown in Table 47 for Global Politics and ITGS. This is arguably a point in favour of circle-arc and particularly ISAWG in these circumstances, as they respectively do not need to match candidates, and include all in their modelling. Circle arc does occasionally produce some unusual results though, whilst common centres and ISAWG are generally in close accordance. Overall ISAWG seems the forerunner here, with common centres a valid option if sufficient data remains (so depending on subject entry).

6.3.4. Changing assessment models/new subjects

In these contexts, the main challenge is actually not which method is most appropriate, but what the best 'prior' is to benchmark the subject or component against. Some external information is necessary to determine where to initially set the standard on the assessment, so that future sessions can proceed into the usual 'standards maintenance' approach. There are several options for alternative benchmarks that can be used such instances:

a. All results from a specific other component in the reference year

b. All results from a specific subject in the reference year

c. All results from a specific group of subjects in the reference year

d. All results from all subjects in the reference year

Broadly speaking, from a-d these are less specific and more general comparators. We would normally prefer a more closely related comparator component or subject to benchmark against, rather than defaulting to 'the entire suite's results' (which effectively means setting the component's standard at the suite's average), but in instances of completely new subjects some of the more specific options may not be possible.

Once a comparator has been selected all of our SRB setting methodologies become viable, accounting for other factors such as subject cohort size. Based on our results for SEHS and Film it is extremely difficult to identify which might be more or less appropriate solely due to the subject being new or different to the prior year.

6.3.5. Verification model

Notably, this context behaves extremely similarly to a stable moderately sized subject – as far as the various models are concerned, a coursework-based subject is identical to an exam-based one. Therefore, whether a subject is a verification model or not should not influence the choice of SRB-setting method.

6.3.6. Skewed distributions

In this context, all methods perform fairly similarly, though at the sparser parts of the mark distribution their minor differences do result in slightly higher boundary differences than in other contexts. As mentioned above it is difficult to say whether any methods are therefore 'wrong' or not. Broadly it is reassuring that there seems little obvious problem with any method in this context.

6.4. Conclusion

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 amount. 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.

This work also highlights a number of directions for future research:

- A consideration of the definition of a 'stable common centre' in the IB context.
- A consideration of the circumstances in which using a different subject as the baseline might be appropriate.
- Further work on ISAWG:
    - Modelling of 'in session' rather than 'end of session' ISAWG predictions to determine whether the method becomes more questionable with more limited information.
    - Modelling of the most appropriate subset of the cohort to include in the equating of ISAWG from one year to the next (or whether not to do this, and therefore maintain the exact same grade distribution on the suite level year-on-year).
    - Modelling of the minimum cohort sizes needed for the method to function reliably (i.e., in terms of populating prediction matrixes and creating SRBs).
    - Investigation of cases where the relationship between general ability (ISAWG) and component score is not stable from one year to the next.
    - Investigation of cases where the subject-to-subject relationship may not be stable from one year to the next.
    - Whether excluding particular components or subjects from the ISAWG may increase its stability.
    - Whether predicting outcomes for DP candidates only may improve ISAWG due to having a more reliable ability estimate for them.
    - Whether 'subject group' level ISAWG might produce more valid results.

7. References

AlphaPlus (2021). Statistical grade boundary setting approaches: Literature review for the IB. Manchester, U.K.: AlphaPlus Consultancy Ltd.

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.

Benton, T., & Sutch, T. (2014). *Analysis of use of Key Stage 2 data in GCSE predictions.* Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Eason, S. (2006). *Predicted Outcomes: Some Issues Related to Small Numbers of Candidates.* Guildford: AQA.

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

Johns, D., & Evans, A. (2019). *IMPLEMENTING ISAWG IN AWARDING.* [Conference presentation]. Ofqual Educational Assessment Seminar.

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.

8. Appendix 1: Component-level results

8.1. DP Mathematics SL (TZ2, May sessions)

Subject-level results for this subject can be found here.

8.1.1. Paper one

Table 51: DP Mathematics SL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	20	39	51	59	68	76	100.0	98.8	87.8	70.9	55.5	35.6	19.1
Actual SRBs			43	54			78							
Zone of uncertainty			39-40	49-51			76							
Maintain prior outcome	0	21	43	54	62	70	78	100.0	98.6	83.1	65.6	49.0	31.5	15.1
Common centres	0	21	42	53	61	70	78	100.0	98.6	84.5	67.3	51.1	31.5	15.1
Stable common centres	0	20	41	53	61	70	78	100.0	98.8	85.7	67.3	51.1	31.5	15.1
Circle-arc equating	0	20	41	52	60	69	77	100.0	98.8	85.7	69.2	53.4	33.6	17.0
ISAWG	0	21	42	53	61	69	77	100.0	98.6	84.5	67.3	51.1	33.6	17.0

Table 52: DP Mathematics SL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	16	31	44	53	63	72	100.0	96.9	86.4	68.0	51.0	31.8	17.2
Actual SRBs			30	43			71							
Zone of uncertainty			33-34	44-45			72-74							
Maintain prior outcome	0	15	30	42	52	61	71	100.0	97.4	87.4	71.3	53.0	35.5	18.6
Common centres	0	14	29	42	52	61	71	100.0	97.7	88.3	71.3	53.0	35.5	18.6
Stable common centres	0	14	28	41	51	60	70	100.0	97.7	89.2	72.9	55.1	37.3	20.3
Circle-arc equating	0	16	32	44	53	62	71	100.0	96.9	85.4	68.0	51.0	33.6	18.6
ISAWG	0	14	29	42	51	61	70	100.0	97.7	88.3	71.3	55.1	35.5	20.3

Table 53: DP Mathematics SL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	15	30	43	53	63	73	100.0	96.1	85.1	68.3	51.5	32.1	14.4
Actual SRBs			29	43			71							
Zone of uncertainty			30-31	43-44			73-75							
Maintain prior outcome	0	14	29	43	52	62	71	100.0	96.6	86.1	68.3	53.3	34.1	17.5
Common centres	0	14	28	43	52	62	71	100.0	96.6	87.0	68.3	53.3	34.1	17.5
Stable common centres	0	13	26	41	51	60	70	100.0	97.0	88.9	71.2	55.1	38.1	19.3
Circle-arc equating	0	15	30	43	52	62	71	100.0	96.1	85.1	68.3	53.3	34.1	17.5
ISAWG	0	14	28	43	52	62	71	100.0	96.6	87.0	68.3	53.3	34.1	17.5

8.1.2. Paper two

Table 54: DP Mathematics SL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	36	46	55	64	73	100.0	95.8	79.7	62.9	44.8	26.0	10.2
Actual SRBs			35	45			73							
Zone of uncertainty			36-37	44-46			73-74							
Maintain prior outcome	0	17	35	45	54	63	72	100.0	96.3	81.0	64.8	47.0	28.0	11.5
Common centres	0	17	35	44	53	63	72	100.0	96.3	81.0	66.6	49.1	28.0	11.5
Stable common centres	0	16	33	43	53	62	72	100.0	96.7	83.6	68.5	49.1	30.0	11.5
Circle-arc equating	0	18	36	45	54	63	72	100.0	95.8	79.7	64.8	47.0	28.0	11.5
ISAWG	0	17	35	44	53	63	72	100.0	96.3	81.0	66.6	49.1	28.0	11.5

Table 55: DP Mathematics SL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	18	35	45	55	66	76	100.0	96.6	82.1	63.8	43.4	23.9	9.9
Actual SRBs			36	45			76							
Zone of uncertainty			35-36	45-46			76-77							
Maintain prior outcome	0	18	37	45	55	66	76	100.0	96.6	78.9	63.8	43.4	23.9	9.9
Common centres	0	18	36	45	55	66	76	100.0	96.6	80.6	63.8	43.4	23.9	9.9
Stable common centres	0	18	36	45	55	65	75	100.0	96.6	80.6	63.8	43.4	25.4	11.1
Circle-arc equating	0	18	37	47	56	64	73	100.0	96.6	78.9	59.7	41.4	27.1	13.6
ISAWG	0	18	36	45	55	65	75	100.0	96.6	80.6	63.8	43.4	25.4	11.1

Table 56: DP Mathematics SL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	16	31	40	50	60	70	100.0	95.9	80.4	64.0	43.7	25.5	10.9
Actual SRBs			30	40			71							
Zone of uncertainty			31-32	40-41			70-71							
Maintain prior outcome	0	15	30	40	50	61	71	100.0	96.3	82.0	64.0	43.7	23.7	9.8
Common centres	0	15	30	40	50	61	71	100.0	96.3	82.0	64.0	43.7	23.7	9.8
Stable common centres	0	14	28	38	48	59	69	100.0	96.8	84.8	68.1	47.8	27.1	12.0
Circle-arc equating	0	15	31	40	51	62	73	100.0	96.3	80.4	64.0	41.8	21.9	7.8
ISAWG	0	15	30	40	50	60	70	100.0	96.3	82.0	64.0	43.7	25.5	10.9

8.2. DP Armenian A: Literature SL (May sessions)

Subject-level results for this subject can be found here.

8.2.1. Paper one

Table 57: DP Armenian A: Literature SL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	6	9	12	15	17	19				100.0	85.2	55.6	14.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6-8	9-11			17-19							
Maintain prior outcome	0	7	15	18	19	19	20		100.0	85.2	44.4	14.8	14.8	3.7
Common centres														
Stable common centres														
Circle-arc equating	0	7	15	17	18	19	20		100.0	85.2	55.6	44.4	14.8	3.7
ISAWG	0	8	16	18	19	19	20		100.0	74.1	44.4	14.8	14.8	3.7

Table 58: DP Armenian A: Literature SL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	6	9	12	15	17	19				100.0	85.7	47.6	28.6
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6-8	9-11			17-19							
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	4	9	12	14	17	19				100.0	95.2	47.6	28.6
ISAWG														

Table 59: DP Armenian A: Literature SL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	6	9	12	15	17	20		100.0	93.1	89.7	79.3	58.6	17.2
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6-8	9-11			17-19							
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	4	8	11	14	16	19		100.0	96.6	93.1	82.8	79.3	24.1
ISAWG														

8.2.2. Paper two

Table 60: DP Armenian A: Literature SL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	10	14	17	20	24			100.0	88.9	66.7	37.0	7.4
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	6	13	16	18	21	23		100.0	92.6	77.8	51.9	25.9	14.8
Common centres														
Stable common centres														
Circle-arc equating	0	6	12	16	19	21	24		100.0	92.6	77.8	51.9	25.9	7.4
ISAWG	0	7	15	19	21	22	24		100.0	81.5	51.9	25.9	18.5	7.4

Table 61: DP Armenian A: Literature SL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	10	14	17	20	24				100.0	81.0	42.9	23.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	6	12	16	19	21	24			100.0	81.0	61.9	42.9	23.8
ISAWG														

Table 62: DP Armenian A: Literature SL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	10	14	18	21	25	100.0	93.1	89.7	89.7	75.9	41.4	6.9
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome														
Common centres														
Stable common centres														
Circle-arc equating	0	4	9	13	17	20	24	100.0	96.6	93.1	89.7	89.7	44.8	24.1
ISAWG														

8.3. DP Swedish A: Literature SL (May sessions)

Subject-level results for this subject can be found here.

8.3.1. Paper one

Table 63: DP Swedish A: Literature SL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	9	12	15	18			100.0	94.7	67.2	31.6	11.3
Actual SRBs			8	9			18							
Zone of uncertainty			7-8	8-9			18-19							
Maintain prior outcome	0	4	8	9	12	15	18		100.0	99.2	94.7	67.2	31.6	11.3
Common centres	0	4	8	9	12	16	19		100.0	99.2	94.7	67.2	24.3	6.1
Stable common centres	0	4	7	9	12	15	18			100.0	94.7	67.2	31.6	11.3
Circle-arc equating	0	3	7	9	12	15	18			100.0	94.7	67.2	31.6	11.3
ISAWG	0	4	8	9	12	15	18		100.0	99.2	94.7	67.2	31.6	11.3

Table 64: DP Swedish A: Literature SL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	12	15	18		100.0	99.6	94.6	67.6	32.4	11.6
Actual SRBs			5	9			18							
Zone of uncertainty			6	9			18							
Maintain prior outcome	0	4	7	9	12	15	18		100.0	99.2	94.6	67.6	32.4	11.6
Common centres	0	4	7	9	12	14	17		100.0	99.2	94.6	67.6	47.1	19.7
Stable common centres	0	4	7	9	12	16	19		100.0	99.2	94.6	67.6	25.9	5.4
Circle-arc equating	0	3	7	9	12	15	18		100.0	99.2	94.6	67.6	32.4	11.6
ISAWG	0	2	5	9	12	15	18			100.0	94.6	67.6	32.4	11.6

Table 65: DP Swedish A: Literature SL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	12	15	18			100.0	92.5	68.8	31.2	8.5
Actual SRBs			6	8			17							
Zone of uncertainty			n/a	n/a			17-18							
Maintain prior outcome	0	3	7	9	12	14	17			100.0	92.5	68.8	43.2	14.6
Common centres	0	3	7	8	11	14	17			100.0	97.0	80.9	43.2	14.6
Stable common centres														
Circle-arc equating	0	3	6	9	12	15	18			100.0	92.5	68.8	31.2	8.5
ISAWG	0	3	7	8	11	15	18			100.0	97.0	80.9	31.2	8.5

8.3.2. Paper two

Table 66: DP Swedish A: Literature SL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	10	12	15	19	22	100.0	98.8	95.2	87.5	66.5	35.9	15.7
Actual SRBs			10	12			22							
Zone of uncertainty			9-10	12			21-22							
Maintain prior outcome	0	5	10	12	15	19	22	100.0	98.8	95.2	87.5	66.5	35.9	15.7
Common centres	0	4	9	12	15	19	22	100.0	99.6	96.4	87.5	66.5	35.9	15.7
Stable common centres	0	5	9	12	15	19	22	100.0	98.8	96.4	87.5	66.5	35.9	15.7
Circle-arc equating	0	5	10	13	16	18	21	100.0	98.8	95.2	81.0	61.3	44.0	21.4
ISAWG	0	5	11	12	15	19	22	100.0	98.8	93.5	87.5	66.5	35.9	15.7

Table 67: DP Swedish A: Literature SL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	11	14	18	21		100.0	94.2	84.6	62.2	31.3	15.8
Actual SRBs			8	10			21							
Zone of uncertainty			8	11			20-21							
Maintain prior outcome	0	4	8	10	14	17	21		100.0	94.2	89.6	62.2	38.6	15.8
Common centres	0	3	7	10	13	17	20		100.0	96.9	89.6	69.9	38.6	19.7
Stable common centres	0	3	7	12	16	20	24		100.0	96.9	77.6	47.9	19.7	5.0
Circle-arc equating	0	4	9	11	14	18	21		100.0	93.1	84.6	62.2	31.3	15.8
ISAWG	0	3	7	10	14	17	21		100.0	96.9	89.6	62.2	38.6	15.8

Table 68: DP Swedish A: Literature SL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	11	14	18	21	100.0	97.5	91.0	81.4	65.3	36.2	17.6
Actual SRBs			8	10			21							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	4	8	10	14	17	21	100.0	98.5	94.5	85.9	65.3	41.2	17.6
Common centres	0	3	6	9	13	17	21	100.0	99.0	96.0	91.0	70.4	41.2	17.6
Stable common centres														
Circle-arc equating	0	4	8	11	14	18	21	100.0	98.5	94.5	81.4	65.3	36.2	17.6
ISAWG	0	3	7	10	14	17	21	100.0	99.0	95.0	85.9	65.3	41.2	17.6

8.4. DP English A: Language and Literature HL (May sessions)

Subject-level results for this subject can be found here.

8.4.1. Paper one

Table 69: DP English A: Language and Literature HL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	5	8	11	14	17	100.0	99.7	98.0	84.6	52.2	22.5	5.4
Actual SRBs			4	7			16							
Zone of uncertainty			5	8-9			17							
Maintain prior outcome	0	2	4	7	10	13	16	100.0	99.9	99.2	91.4	64.4	30.0	10.0
Common centres	0	2	4	7	10	13	16	100.0	99.9	99.2	91.4	64.4	30.0	10.0
Stable common centres	0	2	4	7	10	13	16	100.0	99.9	99.2	91.4	64.4	30.0	10.0
Circle-arc equating	0	2	4	7	10	14	17	100.0	99.9	99.2	91.4	64.4	22.5	5.4
ISAWG	0	2	4	8	11	13	16	100.0	99.9	99.2	84.6	52.2	30.0	10.0

Table 70: DP English A: Language and Literature HL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	5	8	11	14	17	100.0	99.8	98.4	86.9	55.4	23.1	5.2
Actual SRBs			5	8			17							
Zone of uncertainty			5	7-8			17							
Maintain prior outcome	0	2	5	8	11	14	17	100.0	99.9	98.4	86.9	55.4	23.1	5.2
Common centres	0	2	5	8	11	14	17	100.0	99.9	98.4	86.9	55.4	23.1	5.2
Stable common centres	0	3	6	8	11	14	17	100.0	99.8	96.6	86.9	55.4	23.1	5.2
Circle-arc equating	0	2	5	8	11	14	17	100.0	99.9	98.4	86.9	55.4	23.1	5.2
ISAWG	0	2	5	8	11	14	17	100.0	99.9	98.4	86.9	55.4	23.1	5.2

Table 71: DP English A: Language and Literature HL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	5	8	11	14	17	100.0	99.7	97.9	85.2	54.9	24.7	6.3
Actual SRBs			5	8			17							
Zone of uncertainty			5	7-8			17							
Maintain prior outcome	0	2	5	8	11	14	17	100.0	99.8	97.9	85.2	54.9	24.7	6.3
Common centres	0	2	5	8	11	14	17	100.0	99.8	97.9	85.2	54.9	24.7	6.3
Stable common centres	0	2	4	8	11	14	17	100.0	99.8	99.2	85.2	54.9	24.7	6.3
Circle-arc equating	0	2	5	8	11	14	17	100.0	99.8	97.9	85.2	54.9	24.7	6.3
ISAWG	0	2	5	8	11	14	17	100.0	99.8	97.9	85.2	54.9	24.7	6.3

8.4.2. Paper two

Table 72: DP English A: Language and Literature HL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	11	14	17	20	100.0	99.6	93.7	80.4	52.5	24.3	9.4
Actual SRBs			9	11			19							
Zone of uncertainty			9	11			20							
Maintain prior outcome	0	4	9	11	14	16	19	100.0	99.8	93.7	80.4	52.5	31.3	13.1
Common centres	0	4	9	11	14	16	19	100.0	99.8	93.7	80.4	52.5	31.3	13.1
Stable common centres	0	4	9	11	14	17	20	100.0	99.8	93.7	80.4	52.5	24.3	9.4
Circle-arc equating	0	4	9	10	13	17	20	100.0	99.8	93.7	88.2	62.8	24.3	9.4
ISAWG	0	4	9	11	14	16	19	100.0	99.8	93.7	80.4	52.5	31.3	13.1

Table 73: DP English A: Language and Literature HL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	11	14	17	20	100.0	99.4	92.7	79.9	54.2	27.7	10.8
Actual SRBs			9	11			20							
Zone of uncertainty			8-9	11-12			20							
Maintain prior outcome	0	4	9	11	14	17	20	100.0	99.6	92.7	79.9	54.2	27.7	10.8
Common centres	0	4	9	11	14	17	20	100.0	99.6	92.7	79.9	54.2	27.7	10.8
Stable common centres	0	4	9	11	14	18	21	100.0	99.6	92.7	79.9	54.2	21.3	6.9
Circle-arc equating	0	4	9	11	14	17	20	100.0	99.6	92.7	79.9	54.2	27.7	10.8
ISAWG	0	4	9	11	14	17	20	100.0	99.6	92.7	79.9	54.2	27.7	10.8

Table 74: DP English A: Language and Literature HL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	10	13	16	19	100.0	99.6	96.2	87.9	63.5	33.9	14.6
Actual SRBs			9	11			20							
Zone of uncertainty			8-9	10-11			19-20							
Maintain prior outcome	0	4	9	11	14	17	20	100.0	99.6	93.3	80.3	54.3	27.1	10.0
Common centres	0	4	9	11	14	17	20	100.0	99.6	93.3	80.3	54.3	27.1	10.0
Stable common centres	0	4	9	11	14	17	20	100.0	99.6	93.3	80.3	54.3	27.1	10.0
Circle-arc equating	0	4	9	11	14	17	20	100.0	99.6	93.3	80.3	54.3	27.1	10.0
ISAWG	0	4	9	11	14	17	20	100.0	99.6	93.3	80.3	54.3	27.1	10.0

8.5. DP Global Politics HL (May sessions)

Subject-level results for this subject can be found here.

8.5.1. Paper one

Table 75: DP Global Politics HL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	13	16	20	100.0	99.9	99.5	98.9	94.8	83.2	45.6
Actual SRBs			6	8			21							
Zone of uncertainty			5-7	8-10			20-21							
Maintain prior outcome	0	2	5	6	11	16	21		100.0	99.7	99.5	97.7	83.2	34.0
Common centres	0	2	5	7	11	16	20		100.0	99.7	99.3	97.7	83.2	45.6
Stable common centres														
Circle-arc equating	0	3	7	8	12	17	21	100.0	99.9	99.3	99.0	96.5	76.2	34.0
ISAWG	0	3	6	7	12	16	21	100.0	99.9	99.5	99.3	96.5	83.2	34.0

Table 76: DP Global Politics HL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	10	14	17	21		100.0	99.8	98.5	93.8	79.0	34.9
Actual SRBs			8	10			20							
Zone of uncertainty			7	10-11			20-21							
Maintain prior outcome	0	4	8	9	13	16	20		100.0	99.4	99.2	96.1	86.0	47.8
Common centres	0	4	8	9	13	16	20		100.0	99.4	99.2	96.1	86.0	47.8
Stable common centres														
Circle-arc equating	0	3	6	9	13	16	20		100.0	99.8	99.2	96.1	86.0	47.8
ISAWG	0	4	9	11	14	17	20		100.0	99.2	98.3	93.8	79.0	47.8

Table 77: DP Global Politics HL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	12	16	19	100.0	99.9	99.2	96.7	88.1	61.5	33.9
Actual SRBs			4	7			19							
Zone of uncertainty			7-8	10-11			20-21							
Maintain prior outcome	0	2	4	7	11	15	19	100.0	99.9	99.7	98.8	92.3	69.0	33.9
Common centres	0	1	3	7	11	15	19	100.0	100.0	99.9	98.8	92.3	69.0	33.9
Stable common centres	0	1	3	7	11	15	19	100.0	100.0	99.9	98.8	92.3	69.0	33.9
Circle-arc equating	0	2	5	8	12	15	19	100.0	99.9	99.4	98.0	88.1	69.0	33.9
ISAWG	0	2	4	8	12	15	19	100.0	99.9	99.7	98.0	88.1	69.0	33.9

8.5.2. Paper two

Table 78: DP Global Politics HL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	17	23	32	42	51	100.0	97.7	88.6	73.9	48.1	22.5	8.2
Actual SRBs			17	22			42							
Zone of uncertainty			17-19	25-29			53-55							
Maintain prior outcome	0	8	16	21	28	36	43	100.0	98.4	90.2	79.4	60.2	37.0	20.6
Common centres	0	7	15	21	28	35	42	100.0	99.0	91.9	79.4	60.2	40.2	22.5
Stable common centres														
Circle-arc equating	0	6	13	21	30	38	47	100.0	99.4	94.2	79.4	54.0	32.1	13.0
ISAWG	0	8	16	22	29	37	44	100.0	98.4	90.2	76.5	56.9	34.2	18.6

Table 79: DP Global Politics HL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	17	22	31	41	50	100.0	96.8	88.4	76.9	47.4	17.3	4.0
Actual SRBs			17	23			46							
Zone of uncertainty			16-18	22-23			49-51							
Maintain prior outcome	0	8	17	23	31	38	46	100.0	97.7	88.4	73.9	47.4	25.6	9.1
Common centres	0	8	16	22	30	38	46	100.0	97.7	89.9	76.9	51.0	25.6	9.1
Stable common centres														
Circle-arc equating	0	8	16	21	30	40	49	100.0	97.7	89.9	79.0	51.0	20.1	5.1
ISAWG	0	9	18	23	31	38	46	100.0	96.8	86.6	73.9	47.4	25.6	9.1

Table 80: DP Global Politics HL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	17	22	31	40	49	100.0	98.5	92.2	81.6	50.7	18.7	2.9
Actual SRBs			19	24			48							
Zone of uncertainty			17-18	22-23			50							
Maintain prior outcome	0	9	19	24	32	40	48	100.0	98.5	88.6	75.9	46.8	18.7	3.5
Common centres	0	9	19	24	32	40	48	100.0	98.5	88.6	75.9	46.8	18.7	3.5
Stable common centres	0	10	20	23	31	39	47	100.0	97.9	86.6	79.0	50.7	21.3	5.0
Circle-arc equating	0	8	17	22	32	41	51	100.0	98.7	92.2	81.6	46.8	15.3	1.5
ISAWG	0	10	20	25	33	40	48	100.0	97.9	86.6	73.2	43.5	18.7	3.5

8.6. DP ITGS HL (November sessions)

Subject-level results for this subject can be found here.

8.6.1. Paper one

Table 81: DP ITGS HL (November 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	14	21	27	33	39	100.0	98.3	82.5	41.5	14.0	2.4	0.7
Actual SRBs			14	21			34							
Zone of uncertainty			14	21			39							
Maintain prior outcome	0	5	10	17	23	28	34	100.0	99.6	94.0	67.7	31.2	11.1	1.8
Common centres	0	5	11	18	23	29	34	100.0	99.6	91.4	62.4	31.2	8.1	1.8
Stable common centres														
Circle-arc equating	0	5	11	17	23	28	34	100.0	99.6	91.4	67.7	31.2	11.1	1.8
ISAWG	0	6	12	19	25	30	36	100.0	99.4	89.0	55.0	21.6	6.5	1.1

Table 82: DP ITGS HL (November 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	17	25	30	35	40	100.0	99.1	86.5	42.5	16.9	4.6	0.9
Actual SRBs			18	25			42							
Zone of uncertainty			17-18	24-25			40-41							
Maintain prior outcome	0	9	18	25	30	36	41	100.0	99.1	82.6	42.5	16.9	3.4	0.8
Common centres	0	9	18	25	31	36	42	100.0	99.1	82.6	42.5	13.1	3.4	0.6
Stable common centres	0	9	18	25	32	40	47	100.0	99.1	82.6	42.5	9.9	0.9	0.3
Circle-arc equating	0	8	17	25	31	36	42	100.0	99.6	86.5	42.5	13.1	3.4	0.6
ISAWG	0	9	19	26	31	35	40	100.0	99.1	78.1	36.5	13.1	4.6	0.9

Table 83: DP ITGS HL (November 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	8	16	23	29	34	40	100.0	97.4	76.1	41.2	16.8	5.8	1.2
Actual SRBs			13	23			41							
Zone of uncertainty			16	23-24			40							
Maintain prior outcome	0	6	13	23	29	35	41	100.0	99.1	86.7	41.2	16.8	4.6	0.9
Common centres	0	6	13	23	29	35	41	100.0	99.1	86.7	41.2	16.8	4.6	0.9
Stable common centres	0	6	13	23	30	36	43	100.0	99.1	86.7	41.2	14.4	3.8	0.5
Circle-arc equating	0	7	15	23	28	33	38	100.0	98.2	78.8	41.2	19.7	8.2	2.4
ISAWG	0	7	14	23	29	36	42	100.0	98.2	83.6	41.2	16.8	3.8	0.7

8.6.2. Paper two

Table 84: DP ITGS HL (November 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	12	16	19	100.0	99.8	97.6	88.5	64.2	15.1	3.0
Actual SRBs			6	9			18							
Zone of uncertainty			6	9			19							
Maintain prior outcome	0	4	8	11	14	16	19	100.0	99.8	92.9	73.6	37.9	15.1	3.0
Common centres	0	4	9	11	14	16	19	100.0	99.8	88.5	73.6	37.9	15.1	3.0
Stable common centres														
Circle-arc equating	0	3	7	11	14	17	20	100.0	99.8	95.7	73.6	37.9	9.8	1.3
ISAWG	0	4	9	12	15	17	20	100.0	99.8	88.5	64.2	24.4	9.8	1.3

Table 85: DP ITGS HL (November 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	10	13	17	20	100.0	99.7	98.2	90.0	56.9	13.3	2.2
Actual SRBs			7	10			20							
Zone of uncertainty			7	10			20							
Maintain prior outcome	0	3	7	10	13	17	20	100.0	99.9	98.2	90.0	56.9	13.3	2.2
Common centres	0	4	8	10	13	16	19	100.0	99.7	96.9	90.0	56.9	20.9	4.3
Stable common centres	0	4	8	10	13	16	19	100.0	99.7	96.9	90.0	56.9	20.9	4.3
Circle-arc equating	0	3	7	10	13	17	20	100.0	99.9	98.2	90.0	56.9	13.3	2.2
ISAWG	0	4	8	10	13	17	20	100.0	99.7	96.9	90.0	56.9	13.3	2.2

Table 86: DP ITGS HL (November 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	10	13	17	20	100.0	99.9	98.2	89.9	60.0	13.6	1.5
Actual SRBs			7	10			20							
Zone of uncertainty			7-8	10-11			16							
Maintain prior outcome	0	3	7	10	13	17	20	100.0	99.9	98.2	89.9	60.0	13.6	1.5
Common centres	0	3	7	10	13	16	19	100.0	99.9	98.2	89.9	60.0	22.1	4.0
Stable common centres	0	2	5	10	13	17	20	100.0	99.9	99.3	89.9	60.0	13.6	1.5
Circle-arc equating	0	3	7	10	13	17	20	100.0	99.9	98.2	89.9	60.0	13.6	1.5
ISAWG	0	3	7	10	13	17	20	100.0	99.9	98.2	89.9	60.0	13.6	1.5

8.6.3. Paper three

Table 87: DP ITGS HL (November 2017) paper three – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	15	18	20	23	100.0	99.7	98.0	70.3	34.7	18.1	4.7
Actual SRBs			10	15			23							
Zone of uncertainty			9	15			23							
Maintain prior outcome	0	5	10	15	17	20	22	100.0	99.7	96.7	70.3	46.0	18.1	8.1
Common centres	0	5	11	16	18	20	22	100.0	99.7	94.2	59.4	34.7	18.1	8.1
Stable common centres														
Circle-arc equating	0	4	9	15	18	20	23	100.0	99.8	98.0	70.3	34.7	18.1	4.7
ISAWG	0	5	11	16	18	21	23	100.0	99.7	94.2	59.4	34.7	12.6	4.7

Table 88: DP ITGS HL (November 2018) paper three – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	10	17	20	22	25	100.0	99.8	98.9	72.0	39.0	19.6	2.8
Actual SRBs			11	17			24							
Zone of uncertainty			10-11	16-17			24							
Maintain prior outcome	0	5	11	17	19	22	24	100.0	99.8	98.3	72.0	50.7	19.6	6.1
Common centres	0	5	11	17	19	22	24	100.0	99.8	98.3	72.0	50.7	19.6	6.1
Stable common centres	0	5	10	17	19	22	24	100.0	99.8	98.9	72.0	50.7	19.6	6.1
Circle-arc equating	0	5	11	17	19	22	24	100.0	99.8	98.3	72.0	50.7	19.6	6.1
ISAWG	0	6	12	17	19	22	24	100.0	99.8	96.9	72.0	50.7	19.6	6.1

Table 89: DP ITGS HL (November 2019) paper three – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	16	19	21	24	100.0	99.9	98.5	66.3	36.4	20.8	5.4
Actual SRBs			8	16			25	100.0	99.9	98.5	66.3	36.4	20.8	5.4
Zone of uncertainty			9	16			24-25							
Maintain prior outcome	0	4	8	15	18	22	25		100.0	99.1	74.4	46.1	15.0	2.4
Common centres	0	4	9	16	19	22	25		100.0	98.5	66.3	36.4	15.0	2.4
Stable common centres	0	4	8	15	18	22	25		100.0	99.1	74.4	46.1	15.0	2.4
Circle-arc equating	0	4	9	16	19	21	24		100.0	98.5	66.3	36.4	20.8	5.4
ISAWG	0	4	9	16	19	22	25		100.0	98.5	66.3	36.4	15.0	2.4

8.7. DP Film HL (M19 only)

Subject-level results for this subject can be found here.

8.7.1. Comparative study

Table 90: DP Film HL (May 2019) comparative study – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	14	19	23	28	100.0	97.5	88.4	65.9	40.5	27.2	12.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	6	13	16	20	25	29	100.0	96.5	71.0	54.9	36.8	22.5	9.3
Common centres	0	6	13	16	20	24	28	100.0	96.5	71.0	54.9	36.8	24.9	12.8
Stable common centres	0	7	14	17	21	24	28	100.0	94.5	65.9	50.2	33.3	24.9	12.8
Circle-arc equating	0	6	13	17	21	24	28	100.0	96.5	71.0	50.2	33.3	24.9	12.8
ISAWG	0	6	13	16	20	24	28	100.0	96.4	70.9	54.8	36.7	24.8	12.8

8.7.2. Film portfolio

Table 91: DP Film HL (May 2019) film portfolio – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	10	14	17	21	100.0	99.1	94.7	77.6	42.5	19.5	4.0
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	4	9	11	13	16	18	100.0	99.1	84.0	68.9	51.9	25.8	14.1
Common centres	0	4	9	11	13	16	18	100.0	99.1	84.0	68.9	51.9	25.8	14.1
Stable common centres	0	4	8	11	14	17	20	100.0	99.1	89.8	68.9	42.5	19.5	6.9
Circle-arc equating	0	4	8	11	14	17	20	100.0	99.1	89.8	68.9	42.5	19.5	6.9
ISAWG	0	4	9	11	13	16	18	100.0	99.1	84.0	68.9	51.9	25.8	14.1

8.7.3. Textual analysis

Table 92: DP Film HL (May 2019) textual analysis – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	11	15	19	23	100.0	94.8	76.8	65.4	43.5	27.4	13.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	4	8	9	13	18	22	100.0	97.6	82.4	76.8	54.0	31.5	17.4
Common centres	0	4	8	9	13	18	22	100.0	97.6	82.4	76.8	54.0	31.5	17.4
Stable common centres	0	4	9	10	14	19	23	100.0	97.6	76.8	70.7	49.2	27.4	13.8
Circle-arc equating	0	4	8	9	14	18	23	100.0	97.6	82.4	76.8	49.2	31.5	13.8
ISAWG	0	4	8	9	13	18	22	100.0	97.6	82.4	76.8	54.0	31.5	17.4

8.7.4. Collaborative film project

Table 93: DP Film HL (May 2019) collaborative film project – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	10	13	17	20	100.0	98.1	85.0	70.3	42.9	16.6	4.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	2	5	10	13	16	19	100.0	99.6	95.7	70.3	42.9	21.8	8.0
Common centres	0	2	5	9	12	16	19	100.0	99.6	95.7	77.8	52.8	21.8	8.0
Stable common centres	0	2	5	10	13	16	19	100.0	99.6	95.7	70.3	42.9	21.8	8.0
Circle-arc equating	0	3	6	10	13	15	18	100.0	99.1	92.8	70.3	42.9	28.2	12.0
ISAWG	0	2	5	10	13	16	19	100.0	99.6	95.7	70.3	42.9	21.8	8.0

8.8. DP SEHS HL (M18 only)

Subject-level results for this subject can be found here.

8.8.1. Paper 1 (MCQ)

Table 94: DP SEHS HL (May 2018) paper 1 (MCQ) – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	17	23	27	31	35		100.0	97.0	83.4	66.3	40.9	14.6
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			n/a	n/a			n/a							
Maintain prior outcome	0	11	22	26	29	33	36	100.0	99.7	87.3	71.5	55.5	25.7	9.1
Common centres	0	10	21	25	28	32	35	100.0	99.7	90.3	75.4	61.9	34.3	14.6
Stable common centres														
Circle-arc equating														
ISAWG	0	11	23	27	30	34	37	100.0	99.7	83.4	66.3	47.5	19.6	5.0

8.8.2. Paper two

Table 95: DP SEHS HL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	12	23	30	41	51	62	100.0	99.4	90.9	81.2	59.9	36.2	13.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			23-24	30-31			62-63							
Maintain prior outcome	0	13	27	37	46	56	65	100.0	99.2	86.7	69.6	47.8	23.5	10.8
Common centres	0	11	23	33	42	52	61	100.0	99.4	90.9	75.4	58.0	33.4	14.4
Stable common centres														
Circle-arc equating														
ISAWG	0	14	28	39	49	58	68	100.0	99.2	84.8	64.6	39.8	19.9	6.4

8.8.3. Paper three

Table 96: DP SEHS HL (May 2018) paper three – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	9	18	22	27	33	38		100.0	94.2	85.0	63.9	39.2	19.2
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			17-18	22-23			38-41							
Maintain prior outcome	0	10	21	26	31	36	41		100.0	86.7	69.4	46.4	28.1	10.8
Common centres	0	9	19	24	29	34	39		100.0	92.2	76.9	54.2	35.6	15.0
Stable common centres														
Circle-arc equating														
ISAWG	0	11	22	27	32	37	42	100.0	99.7	85.0	63.9	41.4	23.1	6.7

8.9. MYP Mathematics (November sessions)

Subject-level results for this subject can be found here.

8.9.1. Onscreen examination

Table 97: MYP Mathematics (November 2017) onscreen examination – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	22	33	45	56	68	100.0	87.6	61.6	43.8	25.2	15.7	6.6
Actual SRBs			12	20			68							
Zone of uncertainty			20-24	31-36			66-75							
Maintain prior outcome	0	5	11	19	34	48	63	100.0	96.3	87.6	69.4	40.5	21.5	9.1
Common centres	0	5	11	20	34	49	63	100.0	96.3	87.6	66.5	40.5	21.1	9.1
Stable common centres	0	6	12	21	35	48	62	100.0	95.5	84.7	63.2	38.4	21.5	10.3
Circle-arc equating	0	8	17	25	34	43	52	100.0	93.8	73.6	54.1	40.5	26.9	18.6

Table 98: MYP Mathematics (November 2018) onscreen examination – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	11	22	33	45	56	68	100.0	91.9	62.8	43.0	26.7	12.8	2.3
Actual SRBs			22	33			61							
Zone of uncertainty			21-22	32-33			67-68							
Maintain prior outcome	0	11	22	33	42	52	61	100.0	91.9	62.8	43.0	32.6	15.1	5.8
Common centres	0	16	33	43	49	54	60	100.0	79.1	43.0	31.4	22.1	14.0	8.1
Stable common centres	0	21	42	45	50	56	61	100.0	62.8	32.6	26.7	18.6	12.8	5.8
Circle-arc equating	0	11	22	33	45	56	68	100.0	91.9	62.8	43.0	26.7	12.8	2.3

Table 99: MYP Mathematics (November 2019) onscreen examination – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	12	24	37	48	59	70	100.0	98.0	76.5	47.7	30.2	17.4	7.4
Actual SRBs			28	35			75							
Zone of uncertainty			22-27	33-38			68-76							
Maintain prior outcome	0	15	31	40	53	65	78	100.0	96.0	60.4	42.3	22.1	12.8	2.0
Common centres	0	17	35	44	54	65	75	100.0	91.3	53.0	34.9	21.5	12.8	4.7
Stable common centres														
Circle-arc equating	0	14	28	40	51	63	74	100.0	97.3	69.1	42.3	25.5	14.1	4.7

8.10. DP Theatre HL (May sessions)

Subject-level results for this subject can be found here.

8.10.1. Collaborative project

Table 100: DP Theatre HL (May 2017) collaborative project – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	13	18	22	27	100.0	98.9	93.3	83.9	64.0	44.5	20.2
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			9	13			27							
Maintain prior outcome	0	5	10	15	19	24	28	100.0	98.9	91.3	77.0	59.7	34.8	15.7
Common centres	0	4	9	14	19	23	28	100.0	99.3	93.3	80.7	59.7	39.7	15.7
Stable common centres	0	5	11	15	19	24	28	100.0	98.9	89.3	77.0	59.7	34.8	15.7
Circle-arc equating	0	5	11	15	19	24	28	100.0	98.9	89.3	77.0	59.7	34.8	15.7
ISAWG	0	5	10	15	19	24	28	100.0	98.9	91.3	77.0	59.7	34.8	15.7

Table 101: DP Theatre HL (May 2018) collaborative project – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	13	18	22	27	100.0	99.1	94.8	84.3	63.0	44.5	19.5
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			9	13			27							
Maintain prior outcome	0	5	10	13	18	22	27	100.0	99.1	92.9	84.3	63.0	44.5	19.5
Common centres	0	4	9	12	17	21	26	100.0	99.3	94.8	87.3	68.0	49.2	23.7
Stable common centres	0	4	9	13	17	21	25	100.0	99.3	94.8	84.3	68.0	49.2	29.0
Circle-arc equating	0	4	9	13	18	22	27	100.0	99.3	94.8	84.3	63.0	44.5	19.5
ISAWG	0	4	9	12	17	22	27	100.0	99.3	94.8	87.3	68.0	44.5	19.5

Table 102: DP Theatre HL (May 2019) collaborative project – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	9	13	18	22	27	100.0	98.7	94.7	84.9	66.2	46.9	20.9
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			9	13			27							
Maintain prior outcome	0	4	9	13	18	22	27	100.0	99.2	94.7	84.9	66.2	46.9	20.9
Common centres	0	4	9	13	18	22	27	100.0	99.2	94.7	84.9	66.2	46.9	20.9
Stable common centres	0	3	7	12	17	22	27	100.0	99.4	97.2	88.5	70.2	46.9	20.9
Circle-arc equating	0	4	9	13	18	22	27	100.0	99.2	94.7	84.9	66.2	46.9	20.9
ISAWG	0	4	9	14	18	23	27	100.0	99.2	94.7	81.3	66.2	41.0	21.0

8.10.2. Directors' notebook

Table 103: DP Theatre HL (May 2017) director's notebook – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	15	21	27	100.0	98.3	90.7	77.9	50.4	27.8	11.7
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6	9			27							
Maintain prior outcome	0	3	6	7	14	20	27	100.0	98.3	90.7	86.4	54.3	31.2	11.7
Common centres	0	2	5	7	13	20	26	100.0	99.5	93.9	86.4	59.1	31.2	13.9
Stable common centres	0	3	6	8	14	20	26	100.0	98.3	90.7	82.2	54.3	31.2	13.9
Circle-arc equating	0	3	6	8	14	21	27	100.0	98.3	90.7	82.2	54.3	27.8	11.7
ISAWG	0	3	6	8	14	20	26	100.0	98.3	90.7	82.2	54.3	31.2	13.9

Table 104: DP Theatre HL (May 2018) director's notebook – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	15	21	27	100.0	98.0	90.9	78.8	51.8	26.0	10.7
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6	9			27							
Maintain prior outcome	0	3	6	9	15	21	27	100.0	98.0	90.9	78.8	51.8	26.0	10.7
Common centres	0	3	6	9	14	20	25	100.0	98.0	90.9	78.8	56.4	29.7	15.0
Stable common centres	0	3	6	9	14	20	25	100.0	98.0	90.9	78.8	56.4	29.7	15.0
Circle-arc equating	0	3	6	9	15	21	27	100.0	98.0	90.9	78.8	51.8	26.0	10.7
ISAWG	0	3	6	9	15	20	26	100.0	98.0	90.9	78.8	51.8	29.7	12.8

Table 105: DP Theatre HL (May 2019) director's notebook – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	9	15	21	27	100.0	97.7	90.2	76.7	47.3	24.1	10.0
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6	9			27							
Maintain prior outcome	0	3	6	9	15	20	26	100.0	97.7	90.2	76.7	47.3	27.9	11.1
Common centres	0	3	6	8	14	20	26	100.0	97.7	90.2	81.8	52.4	27.9	11.1
Stable common centres	0	2	5	8	14	19	25	100.0	98.8	93.4	81.8	52.4	31.1	12.8
Circle-arc equating	0	3	6	9	15	21	27	100.0	97.7	90.2	76.7	47.3	24.1	10.0
ISAWG	0	3	6	9	15	20	26	100.0	97.7	90.2	76.7	47.3	27.9	11.1

8.10.3. Research presentation

Table 106: DP Theatre HL (May 2017) research presentation – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	10	15	21	26	100.0	98.2	91.4	76.3	52.7	25.0	9.3
Actual SRBs			5	9			24							
Zone of uncertainty			7	12			27							
Maintain prior outcome	0	2	5	9	14	19	24	100.0	99.3	94.2	80.4	58.8	32.6	15.1
Common centres	0	2	5	9	14	19	24	100.0	99.3	94.2	80.4	58.8	32.6	15.1
Stable common centres	0	2	5	9	14	19	24	100.0	99.3	94.2	80.4	58.8	32.6	15.1
Circle-arc equating	0	2	5	10	15	20	25	100.0	99.3	94.2	76.3	52.7	28.8	11.9
ISAWG	0	2	5	9	14	19	24	100.0	99.3	94.1	80.4	58.8	32.5	15.1

Table 107: DP Theatre HL (May 2018) research presentation – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	10	15	21	26	100.0	98.0	92.4	77.7	57.2	30.6	12.4
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6	10			26							
Maintain prior outcome	0	3	6	10	16	21	27	100.0	98.0	92.4	77.7	52.7	30.6	9.5
Common centres	0	3	6	10	16	21	27	100.0	98.0	92.4	77.7	52.7	30.6	9.5
Stable common centres	0	3	6	9	15	20	26	100.0	98.0	92.4	81.9	57.2	35.1	12.4
Circle-arc equating	0	3	7	11	16	21	26	100.0	98.0	89.7	74.5	52.7	30.6	12.4
ISAWG	0	3	6	10	16	21	27	100.0	98.0	92.4	77.7	52.7	30.6	9.5

Table 108: DP Theatre HL (May 2019) research presentation – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	3	6	10	15	21	26	100.0	99.0	96.1	87.6	69.8	40.3	17.8
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			6	10			26							
Maintain prior outcome	0	4	8	13	18	22	27	100.0	98.3	92.5	78.0	55.3	35.7	13.9
Common centres	0	4	8	13	18	22	27	100.0	98.3	92.5	78.0	55.3	35.7	13.9
Stable common centres	0	3	7	12	17	22	27	100.0	99.0	94.7	80.8	60.5	35.7	13.9
Circle-arc equating	0	4	8	12	17	22	27	100.0	98.3	92.5	80.8	60.5	35.7	13.9
ISAWG	0	4	8	13	18	22	27	100.0	98.3	92.5	78.0	55.3	35.7	13.9

8.10.4. Solo theatre piece

Table 109: DP Theatre HL (May 2017) solo theatre piece – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	12	17	22	27	100.0	98.7	91.9	79.3	59.0	32.8	12.5
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			8	12			27							
Maintain prior outcome	0	3	7	11	16	22	27	100.0	99.2	94.3	82.7	63.3	32.8	12.5
Common centres	0	3	7	11	16	22	27	100.0	99.2	94.3	82.7	63.3	32.8	12.5
Stable common centres	0	3	7	11	16	21	26	100.0	99.2	94.3	82.7	63.3	38.0	16.5
Circle-arc equating	0	3	7	11	16	22	27	100.0	99.2	94.3	82.7	63.3	32.8	12.5
ISAWG	0	3	7	11	16	21	26	100.0	99.2	94.3	82.6	63.2	38.0	16.5

Table 110: DP Theatre HL (May 2018) solo theatre piece – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	12	17	22	27	100.0	98.7	91.3	78.7	56.3	32.4	12.7
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			8	12			27							
Maintain prior outcome	0	4	8	12	17	22	27	100.0	98.7	91.3	78.7	56.3	32.4	12.7
Common centres	0	3	7	12	17	22	27	100.0	99.2	94.5	78.7	56.3	32.4	12.7
Stable common centres	0	3	7	11	16	21	26	100.0	99.2	94.5	82.6	61.4	36.9	16.5
Circle-arc equating	0	4	8	12	17	22	27	100.0	98.7	91.3	78.7	56.3	32.4	12.7
ISAWG	0	3	7	11	16	22	27	100.0	99.2	94.5	82.6	61.4	32.4	12.7

Table 111: DP Theatre HL (May 2019) solo theatre piece – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	12	17	22	27	100.0	98.1	89.6	74.5	49.6	26.1	9.6
Actual SRBs			n/a	n/a			n/a							
Zone of uncertainty			8	12			27							
Maintain prior outcome	0	3	7	11	16	21	26	100.0	99.1	92.2	78.6	55.5	30.1	12.5
Common centres	0	3	7	11	16	21	26	100.0	99.1	92.2	78.6	55.5	30.1	12.5
Stable common centres	0	3	6	10	15	20	25	100.0	99.1	95.4	82.9	60.2	34.7	15.1
Circle-arc equating	0	3	7	11	16	21	26	100.0	99.1	92.2	78.6	55.5	30.1	12.5
ISAWG	0	4	8	11	16	21	26	100.0	98.1	89.6	78.6	55.5	30.1	12.5

8.11. DP English A: Literature HL (TZ1, May sessions)

Subject-level results for this subject can be found here.

8.11.1. Paper one

Table 112: DP English A: Literature HL (May 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	9	12	14	17	100.0	99.5	94.2	81.0	43.4	21.1	4.4
Actual SRBs			7	9			17							
Zone of uncertainty			7	9			17							
Maintain prior outcome	0	3	7	9	12	14	17	100.0	99.8	94.2	81.0	43.4	21.1	4.4
Common centres	0	3	7	9	12	14	17	100.0	99.8	94.2	81.0	43.4	21.1	4.4
Stable common centres	0	3	6	8	11	14	17	100.0	99.8	97.3	89.7	54.9	21.1	4.4
Circle-arc equating	0	3	7	9	12	14	17	100.0	99.8	94.2	81.0	43.4	21.1	4.4
ISAWG	0	3	7	9	12	14	17	100.0	99.8	94.2	81.0	43.4	21.1	4.4

Table 113: DP English A: Literature HL (May 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	9	12	14	17	100.0	99.5	94.5	80.4	41.5	18.5	3.4
Actual SRBs			7	9			17							
Zone of uncertainty			6-7	9			17							
Maintain prior outcome	0	3	7	9	12	14	17	100.0	99.6	94.5	80.4	41.5	18.5	3.4
Common centres	0	3	7	9	12	14	17	100.0	99.6	94.5	80.4	41.5	18.5	3.4
Stable common centres	0	3	7	9	12	14	17	100.0	99.6	94.5	80.4	41.5	18.5	3.4
Circle-arc equating	0	3	7	9	12	14	17	100.0	99.6	94.5	80.4	41.5	18.5	3.4
ISAWG	0	3	7	9	12	14	17	100.0	99.6	94.5	80.4	41.5	18.5	3.4

Table 114: DP English A: Literature HL (May 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	7	9	12	14	17	100.0	99.6	96.0	84.1	47.7	23.6	4.3
Actual SRBs			7	9			17							
Zone of uncertainty			7	9-10			17							
Maintain prior outcome	0	3	7	9	12	14	17	100.0	99.7	96.0	84.1	47.7	23.6	4.3
Common centres	0	3	7	9	12	14	17	100.0	99.7	96.0	84.1	47.7	23.6	4.3
Stable common centres	0	3	7	9	12	14	17	100.0	99.7	96.0	84.1	47.7	23.6	4.3
Circle-arc equating	0	3	7	9	12	14	17	100.0	99.7	96.0	84.1	47.7	23.6	4.3
ISAWG	0	3	7	9	12	14	17	100.0	99.7	96.0	84.1	47.7	23.6	4.3

8.11.2. Paper two

Table 115: DP English A: Literature HL (May 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	12	14	17	19	100.0	99.7	96.7	78.2	60.0	31.0	15.6
Actual SRBs			7	12			19							
Zone of uncertainty			7-8	12			19							
Maintain prior outcome	0	3	7	12	14	17	19	100.0	99.9	98.2	78.2	60.0	31.0	15.6
Common centres	0	3	7	12	14	17	19	100.0	99.9	98.2	78.2	60.0	31.0	15.6
Stable common centres	0	3	7	11	14	16	19	100.0	99.9	98.2	84.8	60.0	39.3	15.6
Circle-arc equating	0	4	8	12	14	17	19	100.0	99.7	96.7	78.2	60.0	31.0	15.6
ISAWG	0	3	7	12	14	17	19	100.0	99.9	98.2	78.2	60.0	31.0	15.6

Table 116: DP English A: Literature HL (May 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	12	14	17	19	100.0	99.6	96.4	76.4	58.2	30.4	16.3
Actual SRBs			8	12			19							
Zone of uncertainty			8	11			19							
Maintain prior outcome	0	4	8	12	14	17	19	100.0	99.6	96.4	76.4	58.2	30.4	16.3
Common centres	0	4	8	12	14	17	19	100.0	99.6	96.4	76.4	58.2	30.4	16.3
Stable common centres	0	3	7	12	14	17	19	100.0	99.7	98.1	76.4	58.2	30.4	16.3
Circle-arc equating	0	4	8	12	14	17	19	100.0	99.6	96.4	76.4	58.2	30.4	16.3
ISAWG	0	4	8	12	14	17	19	100.0	99.6	96.4	76.4	58.2	30.4	16.3

Table 117: DP English A: Literature HL (May 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	4	8	11	13	16	18	100.0	99.4	95.9	84.7	70.4	40.1	22.8
Actual SRBs			8	12			19							
Zone of uncertainty			8	11-12			19							
Maintain prior outcome	0	4	8	12	14	17	19	100.0	99.4	95.9	78.3	61.0	31.3	15.7
Common centres	0	4	8	12	14	17	19	100.0	99.4	95.9	78.3	61.0	31.3	15.7
Stable common centres	0	4	8	12	14	17	19	100.0	99.4	95.9	78.3	61.0	31.3	15.7
Circle-arc equating	0	4	8	12	14	17	19	100.0	99.4	95.9	78.3	61.0	31.3	15.7
ISAWG	0	4	8	12	14	17	19	100.0	99.4	95.9	78.3	61.0	31.3	15.7

8.12. DP Chinese B SL (November sessions)

Subject-level results for this subject can be found here.

8.12.1. Paper one

Table 118: DP Chinese B SL (November 2017) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	13	19	25	31	37	100.0	99.8	99.3	95.8	92.2	86.5	78.3
Actual SRBs			12	16			37							
Zone of uncertainty			12-13	19			37							
Maintain prior outcome	0	6	12	16	23	30	37		100.0	99.6	97.7	93.3	87.8	78.3
Common centres	0	6	12	16	23	30	37		100.0	99.6	97.7	93.3	87.8	78.3
Stable common centres	0	5	11	14	21	28	35		100.0	99.7	99.0	94.4	89.7	82.1
Circle-arc equating	0	6	13	18	25	31	38		100.0	99.3	96.5	92.2	86.5	74.8
ISAWG	0	6	13	17	23	30	36		100.0	99.3	96.9	93.3	87.8	80.3

Table 119: DP Chinese B SL (November 2018) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	13	19	26	32	39		100.0	98.8	95.6	90.9	85.7	73.2
Actual SRBs			11	19			37							
Zone of uncertainty			12-13	18-20			38-39							
Maintain prior outcome	0	5	11	19	25	31	37		100.0	99.3	95.6	92.0	86.6	78.7
Common centres	0	5	11	18	24	31	37		100.0	99.3	96.3	92.5	86.6	78.7
Stable common centres	0	4	8	12	18	23	29		100.0	99.9	99.1	96.3	93.0	88.6
Circle-arc equating	0	6	13	18	24	31	37		100.0	98.8	96.3	92.5	86.6	78.7
ISAWG	0	4	9	18	24	31	37		100.0	99.7	96.3	92.5	86.6	78.7

Table 120: DP Chinese B SL (November 2019) paper one – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	7	14	19	26	32	39		100.0	99.0	96.6	90.6	84.6	72.6
Actual SRBs			15	21			39							
Zone of uncertainty			13-15	18-19			38-39							
Maintain prior outcome	0	7	15	21	27	33	39		100.0	98.8	95.3	89.7	83.5	72.6
Common centres	0	7	14	20	26	32	38		100.0	99.0	96.1	90.6	84.6	74.9
Stable common centres	0	7	14	17	22	28	33		100.0	99.0	97.6	94.1	88.7	83.5
Circle-arc equating	0	7	14	20	26	33	39		100.0	99.0	96.1	90.6	83.5	72.6
ISAWG	0	8	17	22	27	33	38	100.0	99.9	97.6	94.1	89.7	83.5	74.9

8.12.2. Paper two

Table 121: DP Chinese B SL (November 2017) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	10	13	15	18	20		100.0	99.3	97.2	93.0	80.6	65.0
Actual SRBs			9	12			20							
Zone of uncertainty			9-10	12-14			21-22							
Maintain prior outcome	0	4	9	12	15	17	20		100.0	99.6	98.2	93.0	85.6	65.0
Common centres	0	4	9	12	15	17	20		100.0	99.6	98.2	93.0	85.6	65.0
Stable common centres	0	5	10	10	13	16	19		100.0	99.3	99.3	97.2	90.0	73.2
Circle-arc equating	0	4	8	12	15	18	21		100.0	99.9	98.2	93.0	80.6	55.7
ISAWG	0	5	10	13	15	17	19		100.0	99.3	97.2	93.0	85.6	73.2

Table 122: DP Chinese B SL (November 2018) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	10	15	17	20	22		100.0	99.5	97.3	94.2	83.9	62.7
Actual SRBs			10	15			22							
Zone of uncertainty			10-12	13-15			20-22							
Maintain prior outcome	0	5	11	15	17	20	22		100.0	99.3	97.3	94.2	83.9	62.7
Common centres	0	5	11	15	17	20	22		100.0	99.3	97.3	94.2	83.9	62.7
Stable common centres	0	4	9	11	14	18	21		100.0	99.8	99.3	97.8	92.1	76.4
Circle-arc equating	0	6	13	16	18	20	22		100.0	98.4	96.2	92.1	83.9	62.7
ISAWG	0	5	10	14	17	19	22		100.0	99.5	97.8	94.2	88.9	62.7

Table 123: DP Chinese B SL (November 2019) paper two – component level results

SRB-setting method	Boundary positions							Cumulative grade outcomes (%)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Actual boundaries	0	5	10	13	16	19	22	100.0	99.9	99.5	97.2	90.0	79.9	59.0
Actual SRBs			10	13			22							
Zone of uncertainty			10	13-14			22-23							
Maintain prior outcome	0	5	10	13	16	19	22	100.0	99.9	99.5	97.2	90.0	79.9	59.0
Common centres	0	5	10	12	15	19	22	100.0	99.9	99.5	98.0	92.6	79.9	59.0
Stable common centres	0	4	9	11	14	18	21	100.0	99.9	99.8	99.0	95.5	84.0	68.7
Circle-arc equating	0	4	9	14	16	19	21	100.0	99.9	99.8	95.5	90.0	79.9	68.7
ISAWG	0	6	12	14	16	19	21	100.0	99.9	98.0	95.5	90.0	79.9	68.7

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

www.alphaplus.co.uk
john.winkley@alphaplus.co.uk

+44 (0) 161 249 9249