

A framework for describing comparability between alternative assessments

Stuart Shaw Cambridge Assessment International Education, **Victoria Crisp** Research Division and **Sarah Hughes** Cambridge Assessment International Education

Introduction

The credibility of an awarding organisation is partly reliant upon the claims it makes about its assessments¹ and on the evidence it can provide to support such claims. Some such claims relate to comparability. For example, for syllabuses with options, such as the choice to conduct coursework or to take an alternative exam testing similar skills, there is a claim that overall candidates' results are comparable regardless of the choice made. This article describes the development and piloting of a framework that can be used, concurrently or retrospectively, to evaluate the comparability between different assessments that act as alternatives. The framework is structured around four types of assessment standards and is accompanied by a recording form for capturing declared comparability intentions and for evaluating how well these intentions have been achieved. The framework and recording form together are intended to:

- provide a structure for considering comparability in terms of four established assessment standards;
- afford an opportunity for test developers to consider their intentions with respect to the comparability claims they wish to make;
- provide a list of factors (within each assessment standard) that are likely to contribute to the comparability of two alternative assessments;
- give a structure for collecting a body of relevant information against these factors;
- prompt an evaluation (on the part of the test developer) of how effectively the claims have been met.

Developing the comparability framework

Concepts of comparability and standards

This work focused on the comparability of assessment standards – in other words, the application of the same standard across different assessments (Newton, 2007). However, what is meant by *assessment standards* requires specification. In an attempt to explore assessment standards, we reviewed the relevant literature, with a focus on standards as associated with comparability in examination systems similar to our own. The search involved scrutinising the literature generated by awarding organisations both within the UK and internationally and the general comparability literature. Drawing on this literature, we identified four types of standard for the purposes of this work: content, demand, marking and awarding.

- **Content standards** are about the value or relevance of the content of the assessment (Cambridge Assessment, 2010). They involve the appropriateness and coverage of the content specified to be assessed. They are also affected by how appropriate the specification or assessment criteria are and how well the questions are aligned to these. In addition, how well an assessment samples the content set out in the specification/syllabus is part of the 'content standards'.
- **Demand standards** are about the nature of knowledge, skills and understanding (KSU) required to successfully complete an assessment (Newton, 2005). This is evidenced in the degree of challenge in the questions and also relates to the level of accessibility of the assessment. The degree of challenge will be affected by the cognitive process(es) that students need to use to tackle the question. These are impacted on by the tools involved (e.g., paper, pencil/pen, notepad, calculator, ruler, computer, keyboard/mouse, computer screen, on-screen tools, response space on-screen) and the cognitive abilities of the candidate needed to answer the question. The tools provided will influence students' performance and experience of the assessment. For example, student familiarity with the tools they will need to use during their assessment (e.g., with the software platform used in an on-screen assessment) is likely to influence performance.

Content and demand standards are related but the distinction between them is useful. The content standard relates to the appropriateness and coverage of topics, whereas the demand standard relates to what the student is expected to do in relation to the topics. However, we recognise that there may be some overlap in terms of the topic and the demands it makes on the student.

- **Marking standards** are about how marks are assigned to reward the knowledge, understanding and skills shown in students' performances. Marking standards also relate to the degree of leniency or severity of marking (Pinot de Moira, Massey, Baird, & Morrissy, 2002; Cambridge Assessment International Education, 2017). Marking standards are inherent in the mark scheme, where the underlying knowledge, understanding and skills to be rewarded are defined. Marking standards are also affected by the marking processes, the compliance of marking processes with codes of practice, the accuracy of the marking, the competence of examiners and adequacy of any standardisation or moderation procedures. This type of standard relates to how well scores reflect the constructs that the assessment is intended to measure.
- **Awarding standards** are about the results that students achieve on an assessment (the assessment outcome, e.g., a grade) and about the kinds of performances that should receive a particular outcome

1. Note that the terms 'assessment' and 'test' are used interchangeably in this article.

(Coe, 2010; Baird, Cresswell, & Newton, 2000). In other words, these standards are about the scores that will receive a particular grade. When grading the assessment, the aim will (almost always) be to maintain the awarding standard applied in previous sessions. Awarding standards are affected by the procedures and policies in place to support grading and by the combination of technical and statistical evidence and professional judgement used in order to determine cut scores.

It is both possible and reasonable for a pair of assessments to be comparable in terms of awarding standards but to not be comparable in terms of content standards, demand standards, or marking standards. For example, two optional assessments within a syllabus might test different topics, different demands, and be marked against a different marking scale but can be considered comparable in terms of awarding standards (though not in terms of the other standards) if the grading process ensures that the same grades are given for equally competent performances.

In deciding on this framework, we have gone beyond the traditional structure of content standards (defining what students should learn) and performance standards (the evidence types needed to demonstrate the content and the quality of student performance that is considered worth a particular grade) (see Linn, 1994). We are using content standards to refer to the content assessed (which will be a subset of the content to be learnt), as this is important in evaluating the comparability of two assessments that are alternatives within a qualification. Additionally, we have replaced performance standards with demand, marking and awarding standards. This provides a more detailed framework for use to support comparison between assessments.

Building the comparability framework and recording form based on the four comparability standards

The purpose of the comparability framework is to outline the criteria for comparability for the four types of standard described. The framework comprises four columns representing the four assessment standards. The ordering of the standards reflects their influence at different stages throughout the test design and testing process. Each standard is fronted by a conditional statement, for example: "If it is the intention that content standards are comparable across assessments, the following need to be fulfilled." What then follows is a list of factors that need to be the same across alternative assessments for there to be comparability with regard to that standard. By way of illustration:

- In the case of *demand* standards, one of the listed factors states that the range of kinds of questions or tasks should be the same across assessments. For example, there should be a similar balance of question types (e.g., MCQ, short answer, essay) on each of the assessments compared.
- In the case of *marking* standards, one of the listed factors states that the application of the mark scheme should be the same across assessments with markers complying with marking guidance and requirements for both assessments.

The comparability recording form provides opportunities to identify which comparability standards are intended as claims, space to record any differences between assessments for each of the standards, and an opportunity for making an overall judgement.

The intentions are likely to depend on the purpose of the assessment

in terms of how it relates to the qualification as a whole. If there is no intention for there to be comparability with regard to a particular standard then the relevant rows can be ignored. Where differences are identified, then any efforts made to address them can be recorded. Differences suggest potential threats to comparability. By addressing such threats, comparability between assessments can potentially be achieved. For example, in the case of comparing an on-screen and a paper-based assessment, if a certain skill cannot be assessed directly on-screen, efforts might be made to provide functionality that allows candidates to show their skills in this area in a comparable way.

Ultimately, it is necessary to determine whether comparability is achieved for each of the standards where it is intended. Whilst all differences are potential threats to comparability, it may be that not all of them are serious threats, and some threats may have been mitigated by efforts to address them (as recorded in the form). This is a judgement that needs to be made in light of the context of the qualification. For example, the omission of a particular subtopic on one of two alternative assessments might have a more or less serious effect on comparability depending on how important the subtopic is within the syllabus. Given the ways in which the differences are addressed, a judgement is necessary as to whether comparability between assessments is sufficient for them to be considered comparable alternatives within the same qualification for each of the standards where comparability is intended.

Piloting the comparability framework

We wished to explore whether those involved in creating assessments could use the framework and form in the way we intended, and whether they found it helpful. To do so, we conducted the pilot exercise described below.

Assessment contexts

The framework and recording form were piloted using two Cambridge Assessment International Education assessment contexts where there are two assessments that act as alternatives:

- On-screen and paper-based tests: Stage 8 Progression tests² in Science for 2018, Papers 1 and 2 (both available as either on-screen or paper-based).
- An Alternative to Practical exam paper and a Practical test: IGCSE³ Chemistry (0620, for June 2017).

Assessment materials

Materials specific to the relevant assessment context were used in the piloting. These were:

- For IGCSE Chemistry
 - IGCSE Chemistry (0620) Syllabus;
 - IGCSE Chemistry Practical test (June 2017), instructions and mark scheme;
 - IGCSE Chemistry Alternative to Practical (June 2017) and mark scheme.

2. Cambridge Primary/Lower Secondary Progression Tests are end-of-stage tests which are designed to measure learners' progress and identify their strengths and weaknesses.

3. The Cambridge International General Certificate of Secondary Education (IGCSE) is a general education qualification for 14 to 16 year olds, available in a range of subjects.

- Stage 8 Science Progression tests
 - Cambridge Lower Secondary Science Curriculum;
 - Paper-based Stage 8 Progression tests (2018, Papers 1 and 2) and mark scheme;
 - Links to the on-screen versions of the Stage 8 Progression tests (2018, Papers 1 and 2).

Participants and procedure

For each assessment context, an expert was recruited who was known to have a familiarity with, and expertise in, the selected context (in terms of their setting and marking experience). Both experts were asked to:

- Read a report that was provided in order to familiarise themselves with the comparability framework, recording form and guidance on how it was envisaged these could be used.
- Re-familiarise themselves with the target assessment materials (as provided).
- Complete the comparability recording form that accompanies the framework with appropriate details. Participants were asked to refer to the assessment materials themselves and to use their knowledge of how the assessments were created, marked and graded. (If there were parts of the process with which they had no or little experience for these assessments, they were asked to leave the relevant boxes blank.)
- Complete a questionnaire in order to provide feedback on use of the framework and recording form, including thoughts on how it could be used in the future.

Feedback from participants

Feedback from the two experts on use of the framework and form, as provided in their questionnaire responses, are summarised below in terms of salient themes. The feedback led to changes to the framework and form. The reader may find it useful to refer to Tables 1 and 2, which show the revised framework and form, when specific points within them are mentioned in this section.

Comprehensibility of the comparability framework and recording form:

- In the main, the framework was largely understood by participants. This was aided by reading the report about the framework. Similar preliminary reading would be required for future users who might be unfamiliar with some of its concepts (e.g., the different 'standards').
- The standard relating to 'demand' proved to be the most challenging to comprehend. Despite the challenges, one participant acknowledged that the 'demand' standard has the potential to extend the thinking of the user beyond merely 'content' comparison and encourage consideration of the cognitive processes students need to employ to tackle a question.
- The bulleted points, 'cognitive processes' and 'range of kinds of questions' (both relating to demand standards) were undoubtedly thought-provoking to the participants. Their comments suggest that these themes required the most thought as they reflect key differences between alternative assessments that are difficult to avoid (e.g., risk of inaccurate results during a practical, effects of working on-screen on the cognitive processes used).
- There was some perceived overlap between certain bullet points within the comparability framework. This suggests an inability on

the part of the user to reliably distinguish between concepts in the framework. However, perceived areas of commonality may be more attributable to a lack of understanding of lexical terms ('domain' and 'topic', for example) than to truly indistinguishable categories. (The perceived areas of overlap have been addressed with revisions to wording, see later.)

- One participant reported challenges in how to use two of the columns within the comparability framework. In the draft of the framework used in the pilot, column 4 asked for differences between the two alternative assessments to be recorded and column 5 was to be used to record how these differences had been addressed. The participant felt that in terms of the paper production and marking processes "the differences have already been addressed as far as possible – and so to identify a difference and then state how they have been addressed is difficult." The participant made some suggestions for revisions to these columns.

Usefulness and usability of the comparability framework and recording form:

- The framework and form were considered useful by the participants especially in terms of providing criteria for assessing comparability between tests that will be treated as equivalent. For example, the criteria should ensure that the focus does not rely too heavily on test content without considering other elements of comparability. In addition, the importance of having the same senior examiners (or at least an overlap of senior examiners) involved in marking two optional tests is reinforced by the completion process (as is the need to maintain question similarity across test forms). However, some features of the framework and elements within the recording form were deemed to be beyond the control of the participants (such as standardisation methods and quality assurance). This is not necessarily problematic in terms of it being possible for users to complete the form but emphasises that some users may be better placed than others to address certain differences between assessments in order to improve comparability.
- Making a judgement in the final column in the recording form (*For the standards where comparability is intended, are you satisfied that there is sufficient comparability?*) appeared to present minimal problems to participants. However, despite differences acknowledged in other columns in the form, participants answered 'yes' in the final column for all of the standards. Given their extensive involvement in the qualifications, it is possible that participant's responses may be somewhat skewed. Alternatively, it may be that the differences are genuinely seen as trivial and not thought to compromise comparability.
- The participants felt that the completed recording form can provide evidence to support the stated (intentional) claims of comparability made by the test developer.
- Participant comments suggested that using the comparability framework and completing the recording form did not provide new insights for those involved in the qualifications. However, as mentioned earlier, participants reported that it provides a set of criteria for considering comparability issues and avoids certain concerns being over- or under-emphasised. Therefore, using the form to systematically consider and record information relating to comparability will still be valuable.

- As a tool for retrospectively evaluating the comparability of two alternative assessments, both participants considered the framework and form valuable (though this perception was subsequently caveated by one participant who argued that any form of retrospective analysis might be considered somewhat tardy).
- Participants felt that the framework could be used beyond the contexts in which it was piloted, wherever parallel routes to certification exist.
- Participants reported that the framework would be useful throughout the test development process but would be most helpful at the setting stage. The framework and recording form might be used at different times during test construction and by different assessment personnel.
- Participants felt that the recording form could constitute an additional source of comparability evidence (alongside existing evidence such as specification grids and statistical data), providing that completion of recording forms does not degenerate into a mechanical checklist exercise.

Frequency of application of the comparability framework and recording form:

- The application of the framework and recording form is not perceived as being necessary every time that parallel assessments are created and used.
- Participants felt that the comparability framework and recording form could be useful when syllabuses are reviewed, and the first time an alternative assessment is created (to parallel an existing assessment).

Users of the comparability framework and recording form:

- One participant reported that the comparability framework and recording form should be thought of as "an organic document that is amended and changed during the life-time of the test."
- Participant responses suggested that a range of personnel with roles within the (re)development of an assessment should be engaged in using the comparability framework and form at different stages throughout the assessment process, for example:
 - Revisers⁴: could be tasked at the revising stage with some responsibility for completing parts of the form when checking for comparability;
 - QPEC (Question Paper Evaluation Committee) personnel: could complete parts of the form when reviewing the assessment materials;
 - Principal Examiners⁵: could use the form when considering what grade thresholds to recommend to the grading team;
 - Assessment Managers⁶: should have a responsibility for declaring the intended comparability claims (i.e., whether the assessments are intended to be comparable with regard to each of content, demands, marking and awarding) and for final evaluation of whether there is sufficient comparability for each dimension where comparability was intended.

4. After questions have been drafted by a setter, revisers provide constructive, expert feedback, checking that the question paper and mark scheme match the syllabus, contain accurate content, are of appropriate demand, and avoid construct-irrelevant effects.

5. Principal Examiners oversee the marking of student responses and are responsible for standards in the marking of examination scripts.

6. Assessment Managers oversee all stages of the creation and use of the assessments for a particular syllabus. They are responsible for standards in a particular examination and over time.

Changes to the structure and content of the recording form

As mentioned earlier, feedback from the pilot participants led to some revisions to the comparability framework and recording form, as detailed below. The revised framework and form are shown in Tables 1 and 2.

- There was potential overlap (giving rise to possible ambiguity) between 'topic' and 'subtopic' in the content standards section of the original form. These two categories were conflated into one row as simply 'subject topics'.
- Reference to *Assessment Objectives* was added within the point relating to 'knowledge, understanding and skills' in the demand standards (this already appeared in the framework itself).
- Reference to 'range of kinds of questions' in demand standards was changed to 'range of kinds of questions/tasks' to ensure this point encompasses a wide variety of assessment task types.
- Due to potential overlap, the categories 'standardisation methods' and 'quality assurance processes' in marking standards were merged into one row to read 'standardisation methods and any other quality assurance processes'.
- One issue that was raised during the pilot was how, when conducting a retrospective comparability evaluation, it was difficult to identify differences between tests (original column 4, *What are the differences between tests, if any, in terms of these features?*) and then how those differences had been addressed (original column 5, *How have the differences been addressed (if they have been)?*), as differences that had been addressed might not be observable in the final materials. As a consequence, the original column 5 was removed. Column 4 was retained and a note was added that actions to minimise differences could also be recorded in this column. This should allow the form to be appropriate for both concurrent and retrospective evaluations.

Conclusions

The comparability framework constitutes a structure for considering four comparability standards when developing an alternate assessment. The comparability recording form affords a means for capturing comparability intentions and for evaluating whether those intentions have been achieved.

A number of issues emerged from both the developmental work on the framework and subsequent piloting:

- The value and application of the framework and recording form should extend beyond the two kinds of contexts with which they were piloted (paper-based and computer-based comparisons and Alternative to Practical and Practical tests) and may include a number of other contexts where there are optional assessments within a qualification.
- There are a number of circumstances in which an evaluation of the comparability of parallel routes might be desirable. For example, where a new assessment is being introduced as a parallel to an existing assessment; where the comparability of two alternative assessments has been queried; or where a qualification containing parallel optional assessments is undergoing routine review with a view to redevelopment. There are two options for how a comparability review using the framework and form can be

Table 1: Revised version of comparability framework

Comparability of:

Content standards	Demand standards	Marking standards	Awarding standards
<p>If it is the intention that content standards are comparable across tests, the following need to be fulfilled:</p> <ul style="list-style-type: none"> subject domains are the same across tests; subject topics are the same across tests; whole test content coverage is the same across tests. 	<p>If it is the intention that demand standards are comparable across tests, the following need to be fulfilled:</p> <ul style="list-style-type: none"> knowledge, understanding and skills (e.g., Assessment Objectives) assessed are the same across tests; the range of kinds of questions or tasks are the same across tests (e.g., similar balance of MCQ, short answer, essay); the test environment does not affect the nature of the teaching and learning; the test environment is easy to use and students have been given sufficient opportunity for familiarisation with the test environment; the cognitive processes (as supported by tools) are the same across tests as far as we can tell; the possible effects of any differences in response format are carefully considered (e.g., for on-screen tests, the effects of typing rather than writing on paper, or of using a drop-down list rather than circling a response on paper). 	<p>If it is the intention that marking standards are comparable across tests, the following need to be fulfilled:</p> <ul style="list-style-type: none"> the mark schemes reward the same knowledge, skills and understanding; the application of the mark scheme is the same across tests with markers complying with marking guidance and requirements for both tests; the way that student responses are presented to markers needs to give equal opportunity for accurate marking across tests; marker competence/accuracy is the same across tests (ideally, the same specific markers are used for both tests); markers are standardised appropriately for both tests and appropriate quality assurance processes are used for both tests; auto-marking (if used) and human marking are both sufficiently accurate and reward intended constructs (only relevant if comparing an on-screen test to a paper-based test). 	<p>If it is the intention that awarding standards are comparable across tests, the following need to be fulfilled:</p> <ul style="list-style-type: none"> awarding is conducted separately for different tests with potentially different grade thresholds (thus ensuring comparability of awarding standards between tests even if there are differences in content, demand or marking standard); the awarding process is the same across tests (e.g., use of judgemental and statistical evidence, methods of recording awarding decisions); sufficient data is available to compare across tests (e.g., entry sizes, benchmark centres, syllabus pairs, knowledge of the characteristics of the candidates entering for each test); awarding standards are maintained over time across tests.

Table 2: Revised version of comparability recording form

Comparability recording form: a structure for describing comparability across tests

Completed by (name)..... (job role)..... Date.....

Assessment name and code.....

1. Standard	2. Is it intended that there should be comparability between tests in terms of each standard?	3. Comparability features – these should be the same across tests if comparability between tests is intended for that standard	4. What are the differences between tests, if any, in terms of these features? (Notes can also be included on actions taken to minimise differences)	5. For the standards where comparability is intended, are you satisfied that there is sufficient comparability?
Content standards		Subject domains*		
		Subject topics*		
		Whole test coverage		
Demand standards		Knowledge, understanding and skills (e.g., Assessment Objectives)		
		Range of kinds of questions/tasks		
		Teaching and learning		
		Test environment ease of use and opportunity for familiarisation		
		Cognitive processes		
Marking standards		Response format		
		Mark schemes		
		Application of the mark scheme		
		The way that student responses are presented to markers		
		Marker competence/accuracy		
Awarding standards		Standardisation methods and any other quality assurance processes		
		Any auto-marking is sufficiently accurate and rewards intended constructs (only relevant if comparing an on-screen test to a paper-based test)		
		Awarding conducted separately for different modes		
		Awarding process		
	Sufficient data is available			
	Awarding standards are maintained over time			

*For example, for a Physics assessment subject domain would refer to areas such as 'electricity and magnetism' and subject topic to aspects such as 'electric circuits'.

conducted, the choice of which will be influenced by the circumstances of the evaluation. The two options are:

- *Concurrent* – During the development of the assessments for a particular examination session (i.e., a particular administration of the assessment), those involved use the comparability framework and form at intervals to guide aspects of the assessment design and to monitor success in achieving comparability. The form can be updated alongside the papers' development, administration, marking and grading, thus providing an audit trail and record of efforts made to achieve comparability.
- *Retrospective* – After the development, administration, marking and grading of the assessments for a particular session, those who were involved use the framework and recording form to review the comparability of the tests based on relevant documents and their own experience of involvement in parts of the process.
- The inherent value of the form is in its potential to capture substantive qualitative features of comparison (and not simply a checklist set of yes/no responses). Therefore, thoughtful consideration of the assessments needs to be encouraged when the framework and form are used.
- There is enough evidence from the pilot and preliminary (albeit tentative) evaluations that the comparability process provided by the framework and recording form could be used to enhance the professional development of examiners, conveying as it does the need to consider and apply several comparability standards.
- Information marshalled in support (or otherwise) of 'content', 'demand' and 'marking' standards might inform the awarding process.
- The Assessment Manager (person responsible for the assessment) is likely to be best placed to have overall responsibility for a comparability evaluation, beginning the process of form completion themselves and then passing the form to other relevant personnel as needed. Whilst there was some variation in the personnel that our pilot participants suggested as appropriate to complete each part of the framework, and there might sometimes be reasons for varying who is involved, some commonalities emerged allowing us to suggest the general pattern in Table 3 (note that the suggestions given here are specific to Cambridge Assessment International Education and may not necessarily generalise to other awarding bodies).

The comparability process outlined here affords a greater level of granularity of reporting for awarding bodies when making comparability claims regarding alternate options within the same syllabus. Not only can claims of comparability be made at a general level (qualification and subject), they can be made in light of specific standards of comparability making clear to stakeholders which of the four assessment standards are applicable. Importantly, standards for which comparability cannot be claimed (intentionally or otherwise) can be identified and described in greater detail than is currently reported.

The framework and form provide a tool that can be used to evaluate the comparability claims made regarding alternative assessments. The resulting evidence may provide support to the argument for the comparability of the parallel tests or provide insights that can inform adjustments to ensure comparability. Whilst the development and piloting of this tool has focused on general qualification contexts, the comparability framework and form might equally be applicable to vocational and technical qualifications.

References

- Baird, J., Cresswell, M., & Newton, P.E. (2000). Would the real gold standard please step forward? *Research Papers in Education*, 15(2), 213–29.
- Cambridge Assessment. (2010). *Exam standards: the big debate – Report and recommendations*. Cambridge: Cambridge Assessment. Retrieved from: <http://www.cambridgeassessment.org.uk/Images/125765-exam-standards-report-and-recommendations.pdf>
- Cambridge Assessment International Education. (2017). *Code of practice*. Cambridge: CAIE. Retrieved from: <http://www.cambridgeinternational.org/images/416992-code-of-practice.pdf>
- Coe, R. (2010). Understanding comparability of examination standards. *Research Papers in Education*, 25(3), 271–284.
- Linn, R.L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher*, 23(9), 4–14.
- Newton, P.E. (2005). Examination standards and the limits of linking. *Assessment in Education: Principles, Policy and Practice*, 12(2), 105–23.
- Newton, P.E. (2007). Contextualising the comparability of examination standards. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Pinot de Moira, A., Massey, C., Baird, J.A., & Morrissy, M. (2002). Marking consistency over time. *Research in Education*, 67(1), 79–87.

Table 3: Proposal for appropriate personnel to complete the comparability recording form*

1. Standard	2. Is it intended that there should be comparability between tests in terms of each standard?	4. What are the differences between tests, if any, in terms of these features? (Notes can also be included on actions taken to minimise differences)	5. For the standards where comparability is intended, are you satisfied that there is sufficient comparability?
Content standards	Assessment Manager	Question Setter ⁷ and Reviser	Assessment Manager
Demand standards	Assessment Manager	Question Setter and Reviser	Assessment Manager
Marking standards	Assessment Manager	Principal Examiner	Assessment Manager
Awarding standards	Assessment Manager	Principal Examiner and awarding team	Assessment Manager

*Note that column numbers match those in Table 2.

7. Question Setters set and develop a draft paper and mark scheme, paying attention to matching the syllabus, accuracy of content, appropriateness of demand, and avoidance of construct-irrelevant effects.