

**The International Association for Educational Assessment**

**International Standards for Educational Assessment Organisations**

**DRAFT**

**Background**

**Why is there a need for the standards?**

The international standards for educational assessment organisations (hereafter Standards) was developed as an recognition by IAEA members of how important they could be to these organisations. The Standards ought to contribute to an increase in the trust and confidence that stakeholders have in assessments being offered around the world. The philosophy underpinning the Standards is aligned to the vision of IAEA to contribute towards and facilitate assessment organisations in improving assessments and assessments systems.

**What is the purpose of the Standards?**

These Standards provides a set of benchmarks of good practice for educational assessment organisations. There are other assessment standards and frameworks that have been published. The IAEA Standards is a very practical framework for international operational contexts. That makes it unique.

**Who will read the Standards?**

The Standards is applicable to organisations directly involved in delivering an assessment system – from registration through assessment to certification. It is published on the IAEA’s website to be accessible to anyone engaged or interested in the business of educational assessment.

**What do the Standards say?**

The Standards covers those aspects of an organisation’s operation that would be expected to be part of its self-evaluation given the local context. It is intended to encourage self-reflection; to inspire educational assessment organisations to review and question taken-for-granted practices by challenging their assumptions.

**How is the Standards structured?**

Section 1 is about how educational assessment organisations can evaluate some of their own processes against the three key measurement standards: validity, reliability and fairness. The threats posed by not attending to the three standards: construct under-representation, construct irrelevance, and unreliability, can appear throughout the assessment cycle.

Section 2 comprises the assessment organisational standards. They are about the organisation itself and the people who contribute to the work of an educational assessment organisation.

Section 3 concerns the operational process standards. They concern the key operational processes that are used during an assessment cycle. We have divided those processes into four parts: examination development, examination administration, marking, and grading and reporting.

Section 4 provides a number of illustrative scenarios related to a number of aspects encountered in sections 2 and 3. Several key questions are raised, as examples, for the consideration of educational assessment organisations aiming to meet the standards. Assessment organisations are expected to review their operation in other aspects in a similar manner.

*The four sections are rather different in terms of their focus. It is perhaps not surprising then that each has a different presentation and style in this document.*

**How should the Standards be used?**

As the Standards would be applicable to assessment organisations of different sizes, with different levels of resources, working in different contexts and running different types of examinations, a balance has been struck in providing a set of benchmarks of good practice for educational assessment organisations. The Standards are not too prescriptive to be applicable to a wider range of assessment organisations.

The Standards is intended to be used and adapted by educational assessment organisations as they feel appropriate taking into consideration their own context. These educational assessment organisations can conduct a self-evaluation in critically reviewing their operations and examination processes to see whether they meet the suggested requirements laid out in the Standards. Through the self-evaluation, assessment organisations would be able to identify their strengths as well as areas for further improvement. In addition, organisations can share their experiences and practices with each other and thereby support each other’s development.

IAEA’s intention is that once the suggestions posed by the Standards have been sufficiently tested, the Standards will evolve into being a more prescriptive initiative, with recognition being granted by IAEA to member organisations which have been independently verified as meeting the Standards.

**Section 1: Measurement Standards**

**Introduction**

There are several assessment design and development criteria that all assessment organisations mustattend to if they are to claim that their assessments are of suitable quality. Quality is about achieving ‘fitness for purpose’ and achieving it consistently. An assessment organisation should have in place quality assurance procedures to ensure that its assessments are of appropriate quality. The organisation should continuously collect evidence to substantiate assessment claims that it makes, identify potential threats to assessment quality and take measures to minimise or eradicate those threats.

This first section of the Standards is about how assessment organisations can evaluate some of their own processes against well-established assessment design criteria. Successful assessments, that is, ones that achieve validity and fulfil a useful purpose in the contexts in which they are to be used (‘fit-for-purpose’), cannot be developed without due consideration being given to three key measurement standards: validity, reliability and fairness. These criteria are critical to the quality of an assessment in general and, therefore, need to assume an explicit role in the self-evaluation of an assessment organisation.

Assessment organisations – even those that have only limited resources – have a responsibility to demonstrate the quality of their assessments.

**Defining the primary standards of validity, reliability and fairness**

The extent to which the inferences which are made on the basis of the outcomes (assessment scores) of the assessment are meaningful, useful and appropriate is considered fundamental to its validity. Assessment is always located within a particular context, and validity is inextricably linked to the use of a specific assessment in a specific context, with a specific group of test takers.

While validity relates to whether the assessment is assessing what it is intended to assess and whether the inferences drawn from the results are justifiable, reliability relates to the stability and precision of the assessment (in other words, the non-random variation in a set of test scores). For example, whether on different occasions or using different markers similar student performances are marked in the same way. If validity is in some sense compromised, high reliability in the assessment will be ineffectual. If reliability is poor – and results are unstable – validity is compromised.

A fair assessment should not discriminate against sub-groups of test-takers (such as ethnical, racial, religious, social, geographical subgroups, test takers with disabilities, non-native speaker-test takers) or give an advantage to other sub-groups. It should also be fair to those who rely on results (such as employers and university admissions staff) by performing consistently and accurately assessing the knowledge, skills and abilities (the underlying constructs) being tested. For example, the content, format and conditions of a test should be suitable and accessible for all members of the intended population of test takers.

Thus, validity is primarily concerned with assessment score meaning; reliability is concerned with assessment score consistency; and fairness is concerned with comparable validity for relevant, identifiable sub-groups of interest. Valid tests are ones that take fairness as their focus.

**Identifying construct under-representation and construct irrelevance**

The credibility of quality assessments is reliant, to a large degree, upon a coherent understanding and articulation of the underlying latent constructs which they seek to assess. If the construct(s) is not well defined, then it will be difficult to support the claims an assessment organisation wishes to make about the usefulness of its assessments, including claims that the assessments do not suffer from construct-related threats such as *construct under-representation* and *construct irrelevance*.[[1]](#footnote-1) It is widely held that *construct under-representation* (under-sampling of the achievement domain) and *construct irrelevant variance* (erroneous inflation or deflation of assessment scores due to certain types of uncontrolled or systematic measurement errors) constitute the two most important risks in assessment design.

*Construct irrelevance*

Construct irrelevance means that the assessment measures too many variables, many of which are irrelevant to the intended construct. Construct irrelevant variance is viewed as a contaminant with respect to score interpretation. Examples of construct irrelevant variance undermining assessment score interpretations are when computer proficiency affects performance on a computerised mathematics test, or when familiarity with a particular item format (e.g., multiple-choice items) affects performance on a reading test. Assessment organisations should seek to reduce or eliminate ‘construct irrelevant variance’ wherever possible.

*Construct under-representation*

Construct under-representation indicates that the questions which are measured in the assessment fail to include important aspects of the construct. Therefore, the assessment scores are unlikely to reveal a test taker's true abilities within the construct which was indicated as having been measured by the assessment. For example, if the constructs being tested in a mathematics examination include three equally weighted topics, fewer items on any one of the three topics will lead to construct under-representation. Such an assessment would under-represent the domain of interest. ‘Construct under-representation’ should be guarded against in the development and evaluation of the assessment. Under such circumstances, the assessment results are unlikely to reveal a test taker’s abilities in respect of the construct.

**Identifying unreliability**

Apart from the importance of sufficient and relevant validity, the reliability of an assessment is equally important. If an assessment is given to two groups of test takers with equal ability under the same assessment conditions, for example, the results of the two assessments should be the same, or very similar. If the results are not the same, then the assessment manifests score unreliability. Factors that have a deleterious affect on assessment score reliability can lead to misinterpretations of the results, if not taken into account.

**Identifying unfairness**

Assessments should be fair and not biased in favour of one group of test takers over another. Bias can be thought of as the unwanted effects in assessment which give rise to unfairness. Bias is inextricably linked to both validity and reliability.

Bias occurs where there are systematic differences in assessment outcomes which are associated with being a member of a specific group – evident when people from two groups are in the same position in respect of the constructs which are the focus of the assessment, but have been given different scores. Aspects of *construct under-representation* and/or *construct* *irrelevance* can differentially affect the achievement of different groups of test takers and consequently the reliability and validity of interpretations and uses of assessment scores.

**Establishing validity, reliability and fairness**

By identifying instances of *construct irrelevance*, *construct under-representation*, and *unreliability*, assessment organisations can enhance and assure the quality, integrity and fairness of their assessment processes. Ultimately, an assessment organisation should seek to develop assessments of the highest technical quality and as free from bias as possible. An assessment organisation should identify and prioritise areas relevant to its particular context and collect evidence deemed relevant and sufficient. The assessment organisation will then be in a better position to take action afterwards to mitigate identified concerns and to develop back-up arrangements and contingency plans. It will also be better placed to inform improved future practice.

**Section 2: Organisational Standards**

The success of an organisation depends on a number of important factors which include but are not limited to its strategies, policies, procedures, people, and quality assurance processes. In order to realise its goals and objectives in an effective and efficient way, an educational assessment organisation should ensure that it meets the following standards.

1. The organisation should have a **strategic plan** that highlights the strategic goals, core values and other relevant functions.

**Table 1** describes the organisational functions that have to be carried out by organisations that run large-scale assessments.

1. The organisation should have a **workforce** of appropriate size comprising sufficient competent academic, administrative and managerial staff to fulfill the key functions as shown in **Table 1**.

The organisation should have well-defined processes and criteria for staff recruitment and appraisals to ensure the efficient achievement of all goals and outcomes expected from them.

1. The organisation should have **quality assurance** and **risk management** processes that are regularly reviewed and updated.

The organisation is committed to rigorous quality assurance processes that are effective, transparent and acceptable to stakeholders. The organisation should also conduct periodic audits to ensure compliance of all assessment related activities.

The organisation should have a well-defined risk management plan that takes into consideration internal risks (e.g., high staff turnover, insufficient financial resources, inadequate infrastructure) and external risks (e.g., reduced governmental funding, competitive organisations, takeover). The organisation should also have strategies to prevent the risk from occurring (control) and strategies to contain the damage (mitigation).

1. The organisation should have effective **support** and **communication** mechanisms for a range of stakeholders.

The organisation should have robust mechanisms to support relevant stakeholders (such as teachers, students) to facilitate readiness for the assessments it offers.

The organisation should have strong communication processes to ensure that the relevant stakeholders are timely and effectively informed regarding the assessment expectations/requirements, student performance feedback, and score/grade interpretations.

1. The organisation should have **knowledge management** processes to facilitate knowledge sharing, transfer and retention, such that staff turnover does not affect the optimum functions of the organisation.
2. The organisation should have adequate processes for **staff development** in order that the key organisational functions are fulfilled.

The organisation should provide appropriate capacity development and support to ensure that its personnel can carry out all necessary functions effectively. It should monitor and evaluate this provision from time to time to make sure that it is effective and remains fit for its purpose.

**Table 1. Key Functions of Assessment Organisations**

|  |  |  |
| --- | --- | --- |
| **Functions1** | **People2** | **Professional requirements** |
| Creating, planning and implementing the strategic direction of the organisation.  Authorising the issue of students’ examination results. | Chief Executive/ Director | The Chief Executive should possess entrepreneurship and leadership skills, is visionary and is responsible for the overall operations of the organisation. S/he can provide strategic leadership, establish/direct department and organisation goals, strategies, polices, operations, budgets and procedures.  S/he should have the ability to develop partnerships with government and non-government organisations and with all stakeholders from school community.  S/he should be able to transform the organisation to meet emerging and future needs. |
| Overseeing day-to-day operations, question paper, documents and reports, scheduling and production, planning and logistics, printing and distribution of papers, rules and regulation, information technology, finance and [human resource](https://www.investopedia.com/terms/h/humanresources.asp) development. | Chief Operating Officer | S/he should be able to lead and direct the management of the examination administration teams, including the participation in training and development, management of complex processes with inflexible deadlines.  S/he should have experience of document management systems and production of printed materials, budgeting and financial management skills.  S/he should manage liaison and coordination with IT, Human and Financial Resources Directorates. |
| Employing appropriate assessment technology and statistical tools in managing, producing and presenting statistical/technical data and recommendations to chief/senior executive and other users, including relevant approving committees, in order to inform decisions about assessment and to ensure the quality, accuracy, consistency and fairness of assessment results and compliance with the requirements stipulated in the specifications and regulations. | Chief Psychometrician | S/he should possess knowledge of statistical and measurement techniques, research concepts and methods, statistical tools and methodologies, particularly in the field of psychometrics. |
| Managing one or more stages of the examination process to ensure that it is carried out in accordance with the test specification and in accordance with the organisation’s policies and procedures. | Subject Specialists | S/he should possess the related knowledge, competency and experience in subject assessment/examination, examination administration, data analysis, research. |
| Using data for research to inform decisions and policy making. | Researchers | S/he should have good research skills and be able to communicate research outcomes in a meaningful way. |
| **Examination Paper Development** | | |
| Develop good quality items and their mark schemes aligned to the curriculum, syllabus and examination specification.  Due consideration should be given to e.g.   * Accuracy of content * Social and cultural considerations * Sensitive topics * Maintaining standard of the examination paper * Comparability of optional questions * Attention to diagrams/illustrations, rubrics and layout of questions * Curriculum coverage over time | Item Writers/Paper Setters | S/he should have subject expertise appropriate in the relevant subject of the examination and at the relevant level. Related teaching and/or assessment experience of the expected performance of students at that level. Adequate language proficiency in the language in which examinations are written. |
| Evaluating and providing constructive, expert feedback on the question paper and mark scheme (*e.g., ensure that there is no overlap between questions and that the question paper and mark scheme contain no errors*). | Revisers/ Moderators | S/he should have the subject expertise and experience appropriate to the relevant level and in the relevant subject of the examination. Sufficient knowledge about the syllabus, student context and examination specifications. Related teaching and/or assessment experience of the expected performance of students at that level. Able to transform a draft test item into a more effective test item. Adequate language proficiency in the language in which examinations are written. |
| Independent check to ensure that the question papers will function as intended by putting themselves in the candidate’s place and working through the question paper independent of the mark scheme. | Assessors/Vetters | S/he should have subject expertise appropriate to the level of the examination. Experience of the likely performance of students at that level. Adequate language proficiency in the language in which examinations are written. |
| Managing a full grammatical, typographical, layout and formatting proof check of all live assessment materials, as part of the question paper production process. Providing editorial advice and training on the conventions of clear and accessible language in the production of assessment materials. | Proof readers | S/he should possess proofreading/editing skills and adequate language proficiency in the language in which examinations are written. |
| **Examination Script Marking** | | |
| Supervising all Team Leaders and monitoring the standard of marking of all teams, finalising mark scheme in order to provide all markers with the means to mark answer scripts consistently and reliably. Participating in grading processes to ensure quality of the assessment and preparing reports after marking. | Chief Examiners/Principal Examiners | S/he should have the subject expertise appropriate to the relevant level and in the relevant subject of the examination. Related teaching and/or assessment experience of the likely performance of students at that level. Competent in identifying benchmark scripts as typical exemplars of scoring rubrics. Able to guide and supervise assistant examiners. Adequate language proficiency in the language in which examinations are written. |
| Supervising teams of markers, monitoring the standard of marking and the accuracy and consistency of all markers during the marking process. | Team Leaders | S/he should have the subject expertise appropriate to the relevant level and in the relevant subject of the examination. Related teaching and/or assessment experience of the likely performance of students at that level. Ability to manage a team. Adequate language proficiency in the language in which examinations are written. |
| Marking answer scripts in accordance with the agreed mark scheme and marking procedures. | Examiners/Markers | S/he should have the subject expertise appropriate to the relevant level and in the relevant subject of the examination. Related teaching and/or assessment experience of the likely performance of students at that level. Adequate language proficiency in the language in which examinations are written. |

**1** It is understood that some functions may be undertaken by other parties or outsourced in some organisations. The assessment organisation should monitor or coordinate with the relevant parties in demonstrating how quality of the services or deliverables are ensured.

**2** Different organisations often use different terminology to describe the same role, for example, examiner and marker. The terminology used in this document should not be taken as having any particular status.

In an organisation, two or more functions may be carried out by separate people. In another organisation, one person may cover those different functions. This document makes no assumptions about whether different functions as described here should or should not be conducted by the same person. One person may discharge more than one function provided that the role does not produce a conflict of interest, such as that between the roles of reviser and setter for the same question paper.

**Section 3: Operational Process Standards**

**Introduction**

It is essential for an assessment organisation to ensure the comprehensiveness and rigor of its operational processes for validity assessment decisions and credible certifications/ qualifications. These processes can be broadly categorised into four major themes which reflect the key operations of an assessment cycle:

Part 1: Examination Development

Part 2: Examination Administration

Part 3: Marking

Part 4: Grading and Reporting

An educational assessment organisation should have arrangements in place that address all the expectations described below in order to secure the objective of optimising validity in its examinations.

**Part 1: Examination Development**

* Examination specification/ syllabus

The examination specification/ syllabus should align with curricular outcomes, that is, it should facilitate the development of an examination that is fit for purpose.

The examination specification/ syllabus should make clear what candidates are expected to learn, make clear what candidates will be assessed on and include a method of assessment that is appropriate for assessing the intended outcomes.

The assessment claims made for the examination should be clearly defined and justified. These claims should be related to the use of the outcomes.

* Question Papers and Mark Schemes

Confidentiality and security should be ensured throughout to preserve the integrity of the question paper development process.

An examination blueprint should facilitate translation of the examination specification/ syllabus into a question paper and mark scheme.

All questions and mark schemes should be written according to the blueprint by qualified personnel. Review, trialling and/or pre-testing of questions should be carried out where appropriate. There should be a policy regarding the purpose and scale of the re-use of questions. All question papers and mark schemes should be subject to a rigorous review process by qualified personnel.

Specific arrangements for candidates with special requirements should be made during examination paper and mark scheme development to ensure fairness for all candidates.

There should be provisions for examination arrangements for candidates with special requirements to ensure fairness for all candidates.

**Part 2: Examination Administration**

* Distribution of paper-based and/ or computer-based question papers

Confidentiality and security should be ensured throughout the printing, packaging, distribution and storage processes to preserve the integrity of the question paper development process. Standard operating procedures are required for all the processes involved.

* Registration

To ensure that accurate data is incorporated in the system for the smooth conduct of examinations, a quality enrolment/ registration process should ensure that candidates’ details can be verified. Standard operating procedures should be developed so that comprehensive details of the students/schools (consistent with the laws of the country of operation) are captured. Data security must be ensured for example, to prevent unauthorised changes of data, unauthorised sharing, leakage of data. The registration process could be online or use hard copies (subject to availability of technology). Verification process should be incorporated for quality assurance purposes.

* Access arrangements

Access arrangements refer to any change from normal facilities provided during the examinations to those candidates who might otherwise not be able to attempt to demonstrate their academic accomplishment.

The assessment organisation should have a set of regulations for access arrangements put in place in advance of the examinations. Examples for the set of arrangements include additional time allowance, separate seating arrangements, special furniture, have a reader or scribe, examination papers in a large font size and technological aids.

The organisation may also have a set of regulations to meet the needs for access arrangements on the day of examinations. These could be based on strong evidence of temporary injury, illness, or other indisposition at the time of the examination and may involve the application of, for example, tariff marks.

* Assessment Centres

A process should be in place to establish identity to ensure that the registered student is the one appearing for the particular examination.

Another process should be established to ensure that the assessment centre has suitable security arrangement in place which would not compromise the integrity of the examination during its conduct. This may include requirements around the use of trained/experienced invigilators, employing inspection teams for monitoring and having CCTV. In particular, a process should be in place that should include the use of job descriptions, a proper selection process and appropriate briefing or training of staff such as invigilators and exam centre supervisors.

The assessment centre staff should keep all the handled materials secured, so they must sign a confidentiality statement.

* Collection of scripts (packaging/despatch)

Confidentiality and security should be ensured throughout the process used to transport scripts (and the student work from school-based assessments) between schools, examiners/moderators and the assessment organisation.

**Part 3: Marking**

This part of Section 2 should be seen to cover marking carried out on paper or on screen, by humans or by machines. The objective in all cases is to mark accurately and in a timely fashion while ensuring that the quality of marking is appropriate. Confidentiality and security should be ensured throughout the whole marking process.

* Planning for marking

The assessment organisation should schedule marking to ensure that it will be of high quality. In identifying or recruiting markers it should accurately estimate the number of markers (both those it will contract and/or those working in schools) that will be required and publish a code of conduct for its markers. It should ensure that it has a backup plan covering both anticipated and unanticipated circumstances.

* Before marking

The marking standard should be established through agreement of the senior markers. All markers must understand and conform to the agreed marking process. To ensure standardisation, all markers will have to apply the mark scheme in the same way as decided by the senior markers.

* During marking

There should be on-going quality checks of marking during the process to ensure that markers are marking to the agreed standard.

* After marking

The senior markers should produce a report on student performance on the question paper. Both the markers themselves and the marking process should be evaluated.

Marking should be adapted as appropriate for school-based assessment.

The marking data created then leads naturally into the grading process in Part 4.

**Part 4: Grading and reporting**

* Grading

The purpose of the grading process is to maintain standards in established examinations and to set standards in new examinations. The assessment organisation should have a grading procedure that covers how recommendations from appropriately qualified personnel about the position of cut scores are made and how approval of final cut scores and grades are made by a senior panel or person. This may involve several steps, is likely to use multiple sources of evidence and generate documentary evidence justifying the decisions made.

* Reporting

Reporting should ensure that the outcomes of the examination provide credible certification of individuals and feedback to the wider system. The assessment organisation should have documentation that covers both the interpretation and use of assessment outcomes as well as guidance for stakeholders on how to interpret and use assessment outcomes.

There should be procedures to ensure the accurate issue of results to individuals, processes (including security features of certificates) to prevent forgery etc, and procedures for dealing with re-marking and appeals against grades.

Assessment organisations should provide support including feedback to schools and teachers so that they can improve future teaching and examination preparation. There should also be appropriate feedback to other relevant stakeholders (e.g., governments, universities, employers).

**Section 4: Meeting the Standards**

The standards described in Sections 1, 2 and 3 of this document set out the IAEA approach to international educational assessment and represent a set of test design criteria and guiding principles against which assessment organisations can reflect upon their own practice. The Standards also provide guidelines for legitimising the three primary measurement qualities of validity, reliability and fairness – essential for ensuring assessment quality. Providing a benchmark for best practice, the Standards serves as a foundation for which assessment organisations can not only build upon their own assessment processes but also appraise those processes.

Although the Standards is not designed to be enforceable by IAEA, its content coheres with generally accepted professional standards adhered to by assessment organisations and users.

In this section, a number of scenarios that relate to areas encountered in previous sections is given here to illustrate how assessment organisations can conduct their self-evaluation. For the Organisational Standards and Operational Process Standards areas, several key questions are raised for the consideration of assessment organisations. Assessment organisations are expected to review their whole assessment operation in a similar manner.

Compliance (or otherwise) with the Standards (through related evaluative questions, including but not limited to, those given here) should be based on careful consideration of local assessment contexts and other relevant factors.

**Organisation Standards**

* Aspect: Strategic Planning. The organisation has a well-defined strategic plan in place that highlights the strategic goals, core values and other relevant functions.
  + Does the organisation review and update its strategic plan in a regular manner so as to cater for the latest changes in the needs of stakeholders and the society?
  + Are departmental goals and objectives aligned to the overall vision and mission of the organisation with a view to facilitate achievement of the organisational goals and objectives?
  + Can departmental goals and objectives be translated into individual goals and objectives?
  + Is annual staff appraisal based on individual goals and objectives?
* Aspect: Knowledge management. The organisation has knowledge management processes in place to facilitate knowledge sharing, transfer and retention.
  + Are there appropriate mechanisms in place to facilitate knowledge sharing, transfer and retention among staff?
  + Are there appropriate documentation and staff development processes in place, such that staff turnover does not affect the quality of the assessments delivered by the organisation?

**Operational Process Standards**

***Examination development***

* + Aspect: Question papers and mark schemes. An examination blueprint should facilitate translation of the examination specification/syllabus into a question paper and mark scheme.
* Do the tasks adequately sample the constructs that are set out as important within the examination specification/syllabus?
* Do the tasks elicit performances that reflect the intended constructs?
* Are the constructs sampled representative of competence in the wider subject domain?
* \*Does the blueprint include content not in the syllabus? Does the blueprint contain context not familiar to some students?
* \*Does the blueprint under or over represent parts of the syllabus? Does the blueprint suggest the inclusion of too many difficult or too many easy questions?
* \*Does the blueprint suggest that tests are either too short or too long to optimise reliability? Is the blueprint so narrow that it limits the production of multiple comparable tests?
* Are the task, test and scoring specifications well defined enabling construction of parallel test forms?
* Are the constructs assessed relevant to the wider subject domain beyond the qualification syllabus?
* Can the constructs relevant to the subject be identified?
* Is it possible to design assessment tasks that require these constructs?
* Does task performance vary according to relevant constructs and is it not affected by irrelevant constructs?

***Examination administration***

* Aspect: Hiring and training of invigilators. To include proper selection process, job descriptions and briefing/ training.
* Is evidence present to assure invigilators to be free of being involved in unethical behaviour that would compromise the reliability or relevance of student performance?
* Does sufficient training and proper agreements on competent/ethical invigilator behaviour assure the standardised conduct of examination and so assure the relevance/ reliability of student performance?
* Aspect: Administrative conditions. Ensuring standardised examination conditions across all contexts.
* Are the administrative conditions under which tasks are set appropriate?

***Marking***

* Aspect: Before marking: establishing the standard. Agreement by the senior markers on the standard.
  + \*Does the mark scheme penalise students for mistakes made outside of the construct being assessed?
  + Does the mark scheme credit all possible correct student answers?
  + (Context: Four-mark question with mark scheme that provides acceptable answers.) How far can the mark scheme be stretched to accommodate student answers whilst remaining relevant?
  + Are the rules, guidance and procedures for scoring responses appropriate for providing evidence of intended constructs?
  + \*To what extent do the senior examiners differ in their interpretation of the agreed mark scheme?
* Aspect: During marking: ensuring application of examination standard.
  + Are examiners applying the mark scheme consistently both within themselves and across other examiners?
  + Are the rules for scoring responses consistently and accurately applied?
* Aspect: Post marking: post exam score adjustments.
  + Are the scores/grades dependable measures of the intended constructs?
  + Are the scaling, equating, aggregation and awarding procedures appropriate for differentiating performance in relation to intended constructs?

***Grading and Reporting***

* Aspect: The grading procedure will cover how recommendations from appropriately qualified personnel about the position of cut scores are made and how approval of final cut scores and grades are made by a senior panel or person.
* \*Has relevant data been overridden by perceptions in making decisions?
* \*Has misrepresented data for a cohort been used to influence the recommendations (e.g., inappropriate use of sub-group data)?
* \*Is anyone participating in the production of the recommendations who is inappropriately qualified or not authorised by the procedures?
* Do the test tasks provide a representative sample of performance?
* Is there a sufficient number of tasks included in the test to provide stable estimates of test performances?

* Aspect: Reporting test scores in a meaningful way to a range of stakeholders.
* Is guidance in place so that stakeholders know what scores/grades mean and how the outcomes should be used?
* Is the meaning of test scores/grades clearly interpretable by stakeholders who have a legitimate interest in the use of those scores i.e., admissions officers, test takers, teachers, employers?

\* Ideally, the desired response to these questions is ‘NO’.

1. According to Messick (1989) “Tests are imperfect measures of constructs because they either leave out something that should be included…or else include something that should be left out, or both”. (p.34). Messick, S (1989). Validity. In R. Linn (Ed.) *Educational Measurement* (pp. 13-103). New York: Macmillan. [↑](#footnote-ref-1)