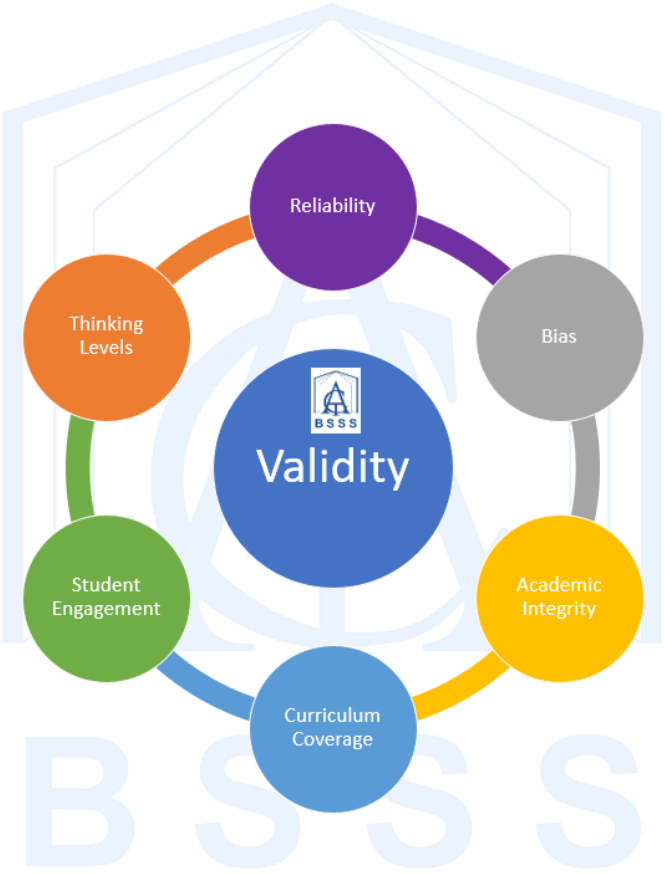
Quality Assessment Guidelines

Board Endorsed 2020



Contents

Introduction

Assessment forms an integral part of the ACT Senior Secondary System (see Figure 1). Developing quality assessment tasks is therefore important for the integrity of the Senior Secondary System. To support teachers and ensure quality, the assessment guidelines are based on contemporary research and are designed to develop a common understanding and language of how to develop assessment to meet the needs of the students. In addition, the guidelines will inform the work of the Board of Senior Secondary Studies (BSSS) in the areas of moderation and assessment.

The BSSS Quality Assessment Criteria (BSSS QAC) can be used for designing and reviewing tasks. The application of the BSSS QAC can be used to assess individual tasks and be used to assess tasks holistically across a unit. A web version of this tool is available here: https://sites.google.com/view/bsss-quality-assessment

The guidelines draw directly from the principles of assessment as articulated in the *Policy and Procedures* document:

Principles

That assessment will:

- *maintain and articulate standards that describe student achievement*
- *provide information about the level of students' skill, knowledge, and conceptual understandings*
- *discriminate between the students*
- *recognise the social and cultural contexts of students*
- *support teaching and learning goals through clear alignment with curriculum, pedagogy, and reporting*
- *involve a range and balance of types of assessment and modes of responding*
- *enhance professional and public confidence.*

Policy and Procedures, P.31

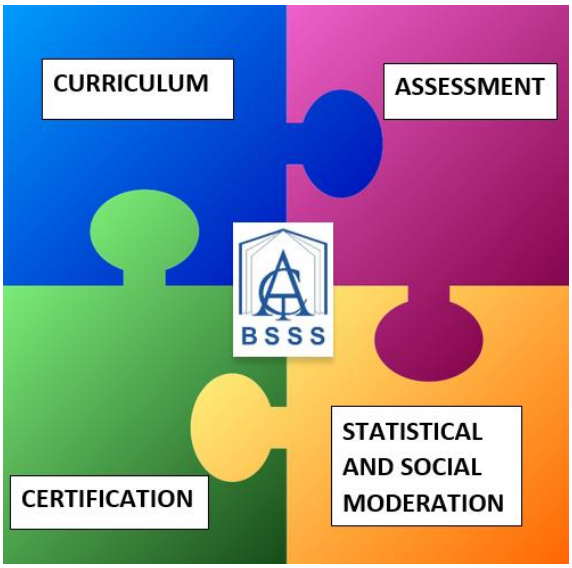


Figure 1: ACT Senior Secondary System

Literature Review

What is assessment?

The OECD defines assessment as "judgements on individual student progress and achievement of learning goals. It covers classroom-based assessments as well as large-scale, external assessments and examinations" (OECD, 2013). This encompasses a very broad range of purposes and practices, so in order to clarify what is being discussed, normally assessment is split into two categories *formative* and *summative* assessment based on the primary aim of the assessment (*for change* or *for grading*).

Formative assessment

Formative assessment is a key component of the teaching and learning process and is generally used to describe the methods used by teachers to monitor student learning at any point in time. This is for the purpose of providing feedback to the student of how to improve their performance in the future, and to assist the teacher in making decisions and changes for teaching and learning going forward (Black & Wiliam, 1996).

Although the word assessment is used, it shouldn't be considered synonymous with formal assessment as the vast majority of formative assessment conducted in classrooms lies outside of formal assessment regimes. Wiliam, one of the pioneers of much of the work related to formative assessment, has expressed regret using the word 'assessment' at all when describing this concept as it has led to some confusion, wishing he had called it 'Responsive Teaching' instead (2013).

Summative assessment

The purpose of summative assessment is to judge the extent of student learning for the purpose of grading, certification, or evaluation. The reason for doing this is to communicate to others about what a student knows and can do within, and often beyond, the classroom (Black & Wiliam, 1996).

An obvious key component for this to occur is that there is a shared meaning across different stakeholders such as students, teachers, parents, and employers. To have any real use, summative judgements need to have meaning outside of their immediate context and there needs to be confidence that a student receiving a particular summative judgement in one school will have received the same in a different school (Christodoulou, 2016).

Formal assessment in the ACT

These guidelines have been designed for the development and evaluation of formal assessment items in the ACT's senior secondary context.

In the ACT senior secondary system, formal assessment outcomes are communicated through grades and, in T courses, scores. Grades are defined by Achievement Standards, and when used with the Specific Unit Goals and Content Descriptions, describe the level of understanding the student demonstrated and what the student can do in the unit. Scores communicate attainment of a student in comparison with others in their scaling group. Ensuring validity of unit scores in scaling groups that consist of different courses, meshing, is largely beyond the scope of this project but is explored in other BSSS training.

Although formal assessment's primary purpose is to ascertain summative judgements, the ACT's continuous assessment model encourages formal assessment to be used formatively as well, by providing feedback to students on areas of improvement, and assisting classroom teachers in evaluation of their teaching programs.

The *2019 BSSS Review of Assessment and Moderation* specified that formal assessment should be determined through three to five assessment items for a standard unit, or two to three assessment items for a half unit. This is to create consistency between the assessment and workload expectations of different subjects, ensure reliability, and balance the stress and anxiety concerns that can be raised by both too few and too many assessment items.

What are we assessing?

Assessment criteria vary across subject areas based on their context but in general are usually categorised as knowledge and understandings, and skills. The Achievement Standards contain the assessment criteria.

Models are continually being refined but current cognitive science suggests that knowledge and understanding are the result of the interaction of a limited working memory and seemingly endless long-term memory with Kirshner et al. (2006) describing learning as "a change in long-term memory". This long-term memory is seen as the single dominant structure of human cognition and is called upon in everything we interact with. The brain builds connections between facts and ideas and develops interwoven schemata of concepts so that working memory can quickly call upon and apply that knowledge to different situations (Kirshner et al., 2006; Nuthall, 2007). These connections and schemata are strengthened through use, with expertise developing over time. Experts can make judgements very quickly as 'snap judgements' and apply them to different situations as a result of these well-developed connections in long-term memory (Gladwell, 2005). For assessment to truly reflect the knowledge and understanding of the student, it is important to encourage this consolidation of concepts into long-term memory.

The ability to *do* something, skills, are often categorised as being assessable through direct observation and evidence. Even here, teachers need to bear in mind the working memory to long-term memory relationship, so as to not confuse performance while learning in one context as the generalised ability to apply the skill to different contexts (Wiliam, 2014).

Human knowledge is often divided into what are called domains, or in the school context, subjects. Wiliam (2014) makes the point that "the 'traditional' school subjects are not arbitrary divisions but are rather distinct ways of thinking about the world". In addition, both Christodoulou (2016) and Willingham (2019) argue that skills such as problem solving, and critical thinking do not exist in isolation at all but are dependent on large bodies of domain-specific knowledge held by the student. Understanding can be defined as the processing of all interactive elements simultaneously (Chen et al., 2017) or the ability to transfer knowledge to new and different situations (Wiggins & McTighe, 2011). The guidelines are flexible to accommodate interpretation in all learning areas.

The OECD defines assessment as "judgements on individual student progress and achievement of learning goals. It covers classroom-based assessments as well as large-scale, external assessments and examinations" (OECD, 2013).

Validity

It is impossible to directly assess the knowledge and understandings in the brain of a student. What teachers instead try to do is use carefully selected proxies (assessment tasks) to provide evidence in order to make valid inferences on the knowledge and understandings of the student (Christodoulou, 2016).

The idea of validity in assessment is a key lynchpin of all assessment tasks and inferences drawn from assessment data (Christodoulou, 2016). Three perspectives are considered in determining validity, "the form of the measure, the purpose of the assessment, and the population for which it is intended." (Dirksen, 2013). Masters (2013) argues that validity focuses on how fit for purpose the assessment is for the domain being assessed. Darr (2005a) notes that *"Judging validity cannot be reduced to a simple technical procedure. Nor is validity something that can be measured on an absolute scale. The validity of an assessment pertains to particular inferences and decisions made for a specific group of students."* (p.55). Inferences drawn from the data that assessment generates, is the foundation of the ACT system. Bennett (2011) argues that for an assessment to be valid, it should be supported with data that shows that different observers would draw the same inferences from the same evidence.

These guidelines focus on six areas of validity (see Figure 2).

Validity can be affected by six factors which form the core of the quality assessment guidelines:

- coverage of the curriculum
- reliability
- bias
- provision for a range of thinking levels
- student engagement
- academic integrity.

Masters (2013) argues that validity focuses on how fit for purpose the assessment is for the domain being assessed.

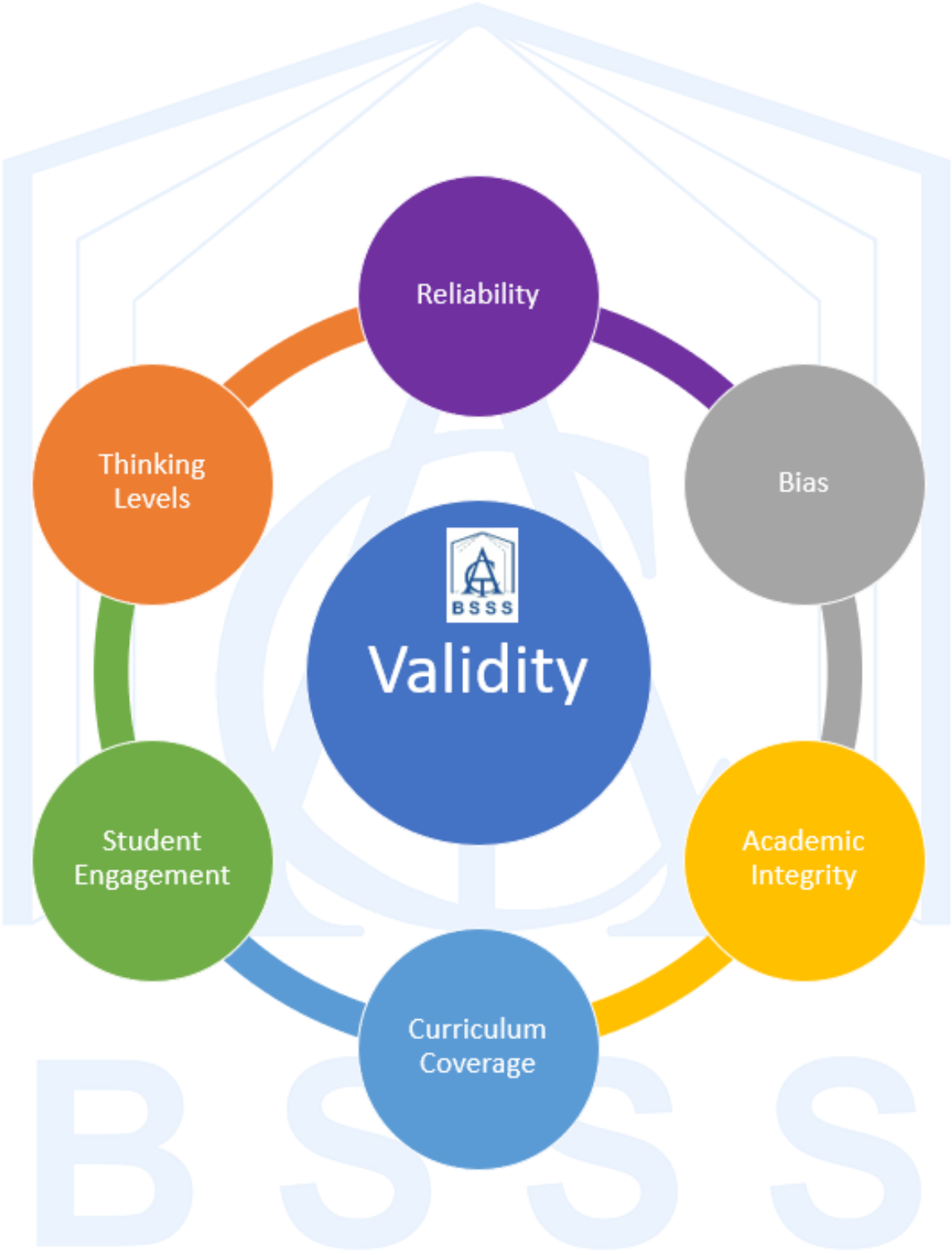


Figure 2: *Quality Assessment Criteria*

The Criteria

In order to ensure validity of assessment, consideration should be made of the following 6 criteria:

Coverage of BSSS Accredited Courses

Wiliam (2014) outlines two threats to validity: assessment which is 'too small' (construct under-representation) and fails to assess what it should, and assessment which is 'too big' (construct irrelevant variance) and assesses things which it should not. An example of both issues may be a video presentation assignment in a History class on a specific small historical aspect. Some teachers may look at the assignment and argue that the assignment is 'too small' only assessing a small part of the unit and others may argue that it is 'too big' assessing things it should not such as their video editing and presentation skills. This is not to say that this assessment should not take place. This assignment could provide a fantastic opportunity for students, but the teacher should try and address these concerns across the entirety of the unit assessment.

The domain of a subject's knowledge, skills and understandings is often impossibly large to assess in entirety. Even at the unit level there can often be goals or descriptions that could be interpreted and assessed in infinite ways. Due to this, assessment is almost always a construct under-representation but is then used to make inferences as to the students' performance in the construct as a whole. For these inference to be valid, teachers should ensure that appropriate breadth and depth is assessed (Christodolou, 2016).

A New Zealand meta-analysis review of the effects of curricula and assessment on pedagogical approaches (2005) shows that high stakes assessment can limit the classroom curriculum for students, particularly lower achievers, and minority students. It is easy for teachers to fall into the trap of assessing what is easy to assess and ignoring the assessment of more difficult to assess skills or content. Wiliam (2014) uses an example of the assessing of practicals in science. It had been shown previously that the skills in science practicals were highly correlated to the scores in science tests. However, when practical assessment was removed from the formal assessment program this correlation does not hold. It is important that assessment type and scope should not be allowed to distort curriculum delivery. (Carr, McGee, Jones, McKinley, Bell, Barr & Simpson, 2005).

Wiliam (2014) outlines two threats to validity: assessment which is 'too small' (construct under-representation) and fails to assess what it should and assessment which is 'too big' (construct irrelevant variance) and assesses things which it should not.

Reliability

To make valid inferences of student knowledge, skills and understandings in the domain, assessment measurements need to minimise the influence of non-relevant factors in the measurement. This is called reliability.

To understand what reliability means we need to understand that all assessment measurements (observed scores) have an error contained within them such that:

Observed Score = True Score + Error

The True Score in the above equation is not that we think a student's ability is predetermined or fixed but represents what that student would get on average if the task was given repeatedly, completed with appropriate 'memory wipe', or was given a multiple parallel assessment of the exact same difficulty on the same material (Bramley & Dhawan, 2010). Note that it is not possible to completely remove this error, while improving reliability of assessment means to aim to minimise this error to improve the stability of results there will always be variation. (Dirksen, 2013). Increased reliability increases our certainty that a student who receives an 80 in an assessment has a higher achievement than a student who receives a 70 for example.

Reliability can be thought of in terms of consistency:

- across time (would students receive the same result from the task if conditions were different?)
- across tasks (would students receive the same result from different tasks assessing this material?)
- and across markers (would students receive the same result from different markers?) (Christodolou, 2016; Darr, 2005b).

Within an assessment item such as a test, reliability can also be thought of as the consistency of a question compared to all the other questions in the task assessing the same material (Dirkson, 2013).

Reliability can really only be determined through the examination of results in the assessment but the factors that decrease error are well known. These include: standardising assessment conditions; designing suitable questions in terms of difficulty for the students involved; having questions that lead to a spread of scores; and having quality rubrics and marking schemes leading to consistent marking and moderation (Darr, 2005b; Masters, 2013).

To make valid inferences of student knowledge, skills and understandings in the domain, assessment measurements need to minimise the influence of non-relevant factors in the measurement. This is called reliability.

Bias Awareness

Bias in assessment is one which favours a student or students over others based on factors other than the key knowledge, skills, and understandings of the student in the unit. Bias plays a role in how inferences are drawn, and so to make assessment more principled, teachers need to recognise "that our characterisations of students are inferences and that, by their very nature, inferences are uncertain and also subject to unintentional biases." (Bennett, 2011, p.18). Bias can be evident in the construction of assessment tasks which means that teachers need to design assessment with, for example, gender, socio-economic and cultural considerations in mind in order to be able to make valid inferences from the data.

The most common way bias is caused by classroom teachers in assessment is through assumptions of background knowledge or the privileging of certain types of background knowledge (OECD, 2013). An individual assessment task may require a level of background knowledge to fully engage with, teachers should be aware of this and allow easy access to this information to lessen the impact of advantage or disadvantage and to not compound this advantage or disadvantage in other assessment items. *The Illinois Guiding Principles of Assessment* (2015) highlights the importance of classroom assessment practices being responsive to and respectful of the cultural and linguistic diversity of students and mentions unnecessary linguistic complexity as an example of bias. The NSW Centre for Education Statistics & Evaluation (2015) refers to assessment that does not "tacitly or explicitly privilege students from high socio-economic backgrounds" (p.6).

Under the Disability Standards for Education (2005) teachers are required to make reasonable adjustments to assessment for students with a disability. Reasonable adjustments are ones that maintain the assessment of a student against the Achievement Standards, unit goals and unit content descriptions of the unit while mitigating the effect of a disability on the assessment. Identifying the key knowledge, skills and understandings is an essential component to ensure that the validity of the assessment is maintained.

Formal assessment in senior secondary should assess the student's objective performance and not incorporate judgements of character, effort, behaviour or potential (Hanover Research, 2011). This can be difficult for some teachers. Teachers can, however, take steps to ensure these unconscious biases do not cloud their objective judgement such as transparent and explicit marking schemes and marking processes, deidentified student assessment, or having teachers not teaching the unit as markers of assessment (Stevens, Ructtinger, Liyanage & Crawford, 2017; Masters & Forster, 1996).

Calculating the bias in assessment can really only be determined through the analysis of assessment results.

Bias plays a role in how inferences are drawn, and so to make assessment more principled teachers need to recognise, "that our characterisations of students are inferences and that, by their very nature, inferences are uncertain and also subject to unintentional biases." Bennett (2011, p.18).

Levels of Thinking

There are a number of proposed theories for how students learn and how their thinking in concepts progresses. The most widely known general theoretical frameworks are Bloom's Taxonomy (1956), Anderson and Krathwohl's Taxonomy (Bloom's revised taxonomy) (2001) or SOLO Taxonomy (Biggs & Collis, 1982). These generally aim to describe phases of understanding and application, and the interconnectedness with other concepts or ideas.

Individual concepts from a domain can be mapped out to describe the sequence of how ideas and practices develop. These are generally called 'learning progressions' (Furtak, Morrison, and Kroog, 2014). The best developed learning progressions aim to be 'top-down' involving the views of content experts and 'bottom-up' by seeking to understand how student learning intersects with the content (Stevens, Ructtinger, Liyanage & Crawford, 2017). Ideally, they are linear and impossible for students to achieve higher elements without satisfying earlier elements. For this reason, learning progressions work best when focused on an appropriately small concept and are locally adapted to the students (Wiliam, 2014).

Providing assessment that assesses a range of thinking levels allows students access to the assessment task as well as the opportunity to develop and extend their thinking. Teachers are faced with increasing diversity in classrooms (Moon, 2005) and therefore using assessment tasks that have a range of thinking levels, from low to high, will allow for a spread of results. In addition, having a range of assessment tasks will allow students to demonstrate different thinking levels, skills and abilities, and different assessment tools such as group work, oral tests or debates can help to improve their learning (Murillo & Hidalgo, 2017).

Teachers are faced with increasing diversity in classrooms (Moon, 2005) and therefore using assessment tasks that have a range of thinking levels, from low to high, will allow for a spread of results.

All assessment tasks in the ACT are based on the Achievement Standards which cater for the needs of diverse learners. Rubrics which are developed for each task are specific and should use the verbs from the theoretical framework to define levels of achievement (Griffin, 2018).

Student Engagement

Students who are unmotivated to complete an assessment will not produce reliable or valid assessment results (Nuthall, 2007). Which means student engagement is an important aspect of a quality assessment.

Transparent and clear assessment instructions which describes what success looks like allows students to participate fairly in the assessment process and increases reliability (Wiliam, 2014). Students need to feel equipped to complete the task with the knowledge, understanding and skills gained from the classroom.

In addition, designing assessments that are embedded in contemporary issues and relevant to the students also improves engagement. Authentic tasks promote realistic problem-solving (Masters, 2014, Bae & Kokka, 2016) and allow students to think as an expert would in a discipline area. Bae and Kokka also outline how student autonomy can improve engagement, giving students decision-making opportunities in regard to their assessment. Collaborative opportunities are also often popular with students.

A student's engagement with assessment is not just affected by these factors. Indeed family, peer and internal pressures can have a greater impact on a student's motivation than the formal assessment requirements (Nuthall, 2007). Schools, leaders, and classroom teachers need to promote positive student wellbeing, ensuring that students feel supported with their needs.

Bae and Kokka (2016) also outline how student autonomy can improve engagement, giving students decision-making opportunities in regard to their assessment. Collaborative opportunities are also often popular with students.

Academic Integrity

Academic integrity is the assurance that student work is the genuine product of the student being assessed. Academic integrity is of the utmost importance for ensuring that results allow valid inferences to be made about student achievement.

Assessment tasks that utilise 'test conditions' that prevent communication between students is a common approach for appropriate tasks. The test conditions should be clearly communicated to students to remove the possibility of ambiguity or confusion. Maintaining test security and ensuring tasks are not reused will further assist in academic integrity.

Teachers can build academic integrity into their assessments through: designing a wide range of assessment types; changing tasks regularly; using a recent or local context rather than a general context; incorporating classroom experiences that outside agents would not be privy to; including personal reflection/opinion; using interdependent tasks and drafting or evidence of planning, check points and clear tracking (Charles Sturt University, 2020, University of Waterloo, n.d., University of Tasmania, 2018).

Academic integrity is the assurance that student work is the genuine product of the student being assessed.

Process for Developing Quality Assessment Tasks

Process of developing a quality assessment

Before designing individual tasks, it is worth spending some time considering how the formal assessment suite for the unit as a whole will work. Coverage of the unit's curriculum is usually split between the tasks in some form, and individual tasks may lack in one criterion which is then made up for in other tasks, for instance. Prior planning should remove these shortfalls at the unit level.

When developing assessment tasks, it is important to work through a process. Wiliam, 2014, discusses the importance of starting with a consideration of what we want students to be able to do.

Step 1: What do we want students to show evidence of? (what are the big ideas in the unit?)

Step 2: What evidence do we need to collect to support these claims?

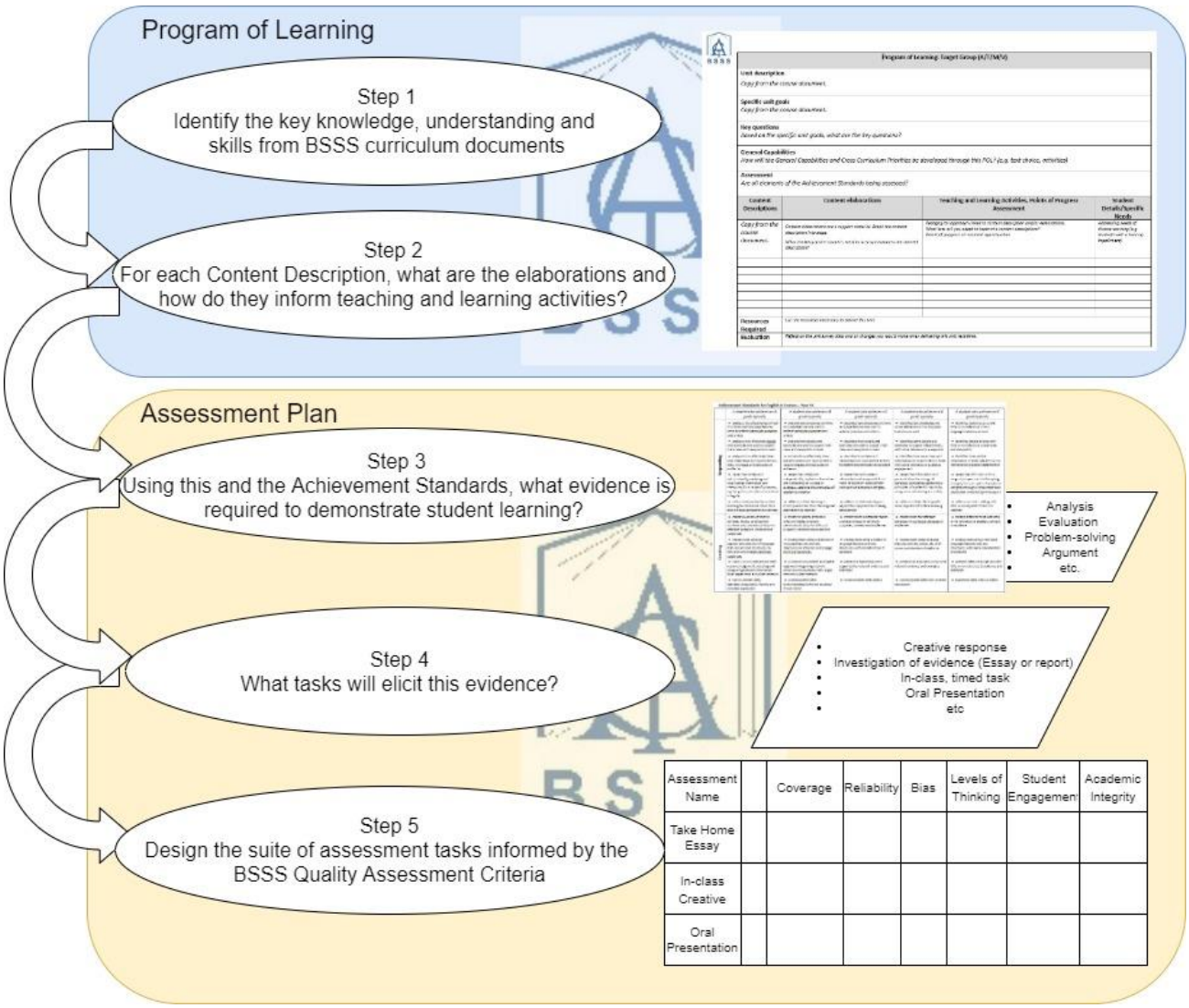
Step 3: What task will allow this evidence to be collected?

These three steps form the starting point when designing assessment and are supported by the work of Wiggins and McTighe (2011) in their *Understanding By Design* work. Wiggins and McTighe advocate that teachers should: identify desired results, determine acceptable evidence, and then plan instruction to facilitate this.

Quality assessment is developed individually and collaboratively. Where assessment is developed individually it should be peer reviewed. Assessment tasks should be considered through the lens of the student who will be taking the assessment.

In addition, strategies for moderating and meshing student results must be planned.

Quality assessment is developed individually and collaboratively.


Program of Learning
Step 1
Identify the key knowledge, understanding and skills from BSSS curriculum documents
Step 2
For each Content Description, what are the elaborations and how do they inform teaching and learning activities?
Assessment Plan
Step 3
Using this and the Achievement Standards, what evidence is required to demonstrate student learning?
Analysis
Evaluation
Problem-solving
Argument
etc.
Step 4
What tasks will elicit this evidence?
Creative response
Investigation of evidence (Essay or report)
In-class, timed task
Oral Presentation
etc
Step 5
Design the suite of assessment tasks informed by the BSSS Quality Assessment Criteria
Assessment Name
Coverage
Reliability
Bias
Levels of Thinking
Student Engagement
Academic Integrity
Take Home Essay
In-class Creative
Oral Presentation

Criteria considerations for the designing the suite of assessment

When planning the suite of assessment items for a unit, teachers should consider the validity of the assessment by applying the six criteria and answering this question: **Is the suite of tasks fit for purpose?**

Coverage of BSSS Accredited Courses

Do the assessment tasks cover the curriculum I am aiming to cover?

1. **Outstanding Coverage of BSSS Accredited Courses**
    - The suite of assessment tasks is strategically planned for alignment with Achievement Standards, unit goals and content descriptors. Assessments are not too big: assessing irrelevant content or criteria; nor too small: missing important content or criteria.
    - Assessment is strategically planned so that it does not distort the intent of the curriculum, aligning with and developing the skills evident in course goals.
    - The general capabilities and cross curriculum priorities are seamlessly integrated into the suite of assessment tasks.

2. **High Coverage of BSSS Accredited Courses**
    - The suite of assessment tasks is thoughtfully planned. Assessments are not too big: assessing irrelevant criteria; nor too small: missing important content or criteria.
    - Assessment does not distort the intent of the curriculum.
    - The general capabilities and cross curriculum priorities are integrated into the suite of assessment tasks.

3. **Satisfactory Coverage of BSSS Accredited Courses**
    - Assessment tasks are appropriately planned. Assessments are not too big: assessing irrelevant criteria; nor too small: missing important criteria.
    - Assessment does not distort the intent of the curriculum.
    - Assessment tasks provide opportunities to engage with the general capabilities and cross curriculum priorities.

4. **Minimal Coverage of BSSS Accredited Courses**
    - Assessment tasks require refinement. Assessments are uneven. Some tasks are either too big: assessing irrelevant criteria; or too small: missing important criteria.
    - Assessment distorts the intent of the curriculum.
    - Assessment tasks provide minimal opportunity to engage with the general capabilities and cross curriculum priorities.

5. **No Coverage of BSSS Accredited Courses**
    - Assessment tasks are unplanned. Assessments are uneven. Some tasks are either too big: assessing irrelevant criteria; or too small: missing important criteria.
    - Assessment extensively distorts the intent of the curriculum.
    - Assessment tasks provide little or no opportunity to engage with the general capabilities and cross curriculum priorities.

Reliability

Are the assessment tasks reliable? Do they have a clear marking schemes and rubric?

1. **Outstanding Reliability**
    - Assessment tasks and conditions are strategically designed to remove all sources of non-relevant variation in measurements.
    - It is a high priority that assessment conditions are clear, consistent, and enforced.
    - In units with multiple classes and teachers, consistency around messaging and assistance is ensured through embedded practice.
    - Marking schemes and rubrics are clear and unambiguous to ensure consistency in student and marker interpretation.
    - Consistency of marking is ensured through a range of moderation processes such as single marker of task or sub-task, double marking or sample double marking, utilising sample scripts/responses for all grade levels, or comprehensive in-school moderation or marking calibration activities.
    - Instructions/questions are clear and unambiguous to student interpretation.

2. **High Reliability**
    - Assessment tasks and conditions are thoughtfully designed to remove sources of non-relevant variation in measurements.
    - Assessment conditions are clear and do not advantage or disadvantage individual students.
    - In units with multiple classes and teachers, consistency around messaging and assistance is considered.
    - Marking schemes and rubrics are clear and aim to reduce marker variation.
    - Consistency of marking is considered through processes such as single marker of task or sub-task, double marking or sample double marking, access to sample scripts/responses, or in-school moderation or marking calibration activities.

3. **Satisfactory Reliability**
    - Assessment tasks and conditions are designed with some consideration of reducing sources of non-relevant variation in measurements.
    - The assessment and assessment conditions are discussed in units with multiple classes and teachers.
    - There is a marking scheme developed for the task and applied in marking.
    - Different markers discuss marking of concern to ensure consistency and have access to an answer key or sample answers.

4. **Minimal Reliability**
    - Assessment tasks are designed with minimal consideration of reducing sources of non-relevant variation in measurements.
    - Assessment conditions could be interpreted differently by different students.
    - There is minimal discussion about the assessment between teachers of the same unit.
    - The marking scheme is underdeveloped and requires interpretation.
    - Different markers barely discuss the marking, or the answer key or sample answers are underdeveloped or incomplete.

5. **No Reliability**
    - Performance in the assessment tasks is largely determined by sources of non-relevant variation.
    - Assessment conditions are not clearly stipulated to students and could be interpreted very differently by different students.
    - There is minimal discussion about the assessment between teachers of the same unit. There is no clear marking scheme.
    - There is no answer key or similar provided. Different markers do not discuss marking.

Bias Awareness

Are there any inherent biases (gender, socio-economic status, culture, disability) evident in the assessment tasks? How will marking and moderating of the assessment ensure objectivity?

1. **Outstanding Bias Awareness**
    - Assessment tasks are strategically designed to be sensitive and empowering for all students, catering for the diverse needs of gender, socio-economic status, disabilities and/or cultures.
    - The suite of assessment does not marginalise or favour a student or group of students, or advantage or disadvantage certain background knowledge or ways of thinking.
    - Intercultural understanding and consideration of alternate points of view is promoted in task design.
    - Highly considered modifications are made to assessment ensuring participation of diverse learners in a fair and equitable way.
    - Marking bias is strategically planned for and marking is based on evidence which utilises comprehensive practices to avoid bias.

2. **High Bias Awareness**
    - Assessment tasks are designed that promote the diverse needs of gender, socio-economic status, disabilities and/or cultures that do not marginalise or favour a student or group of students.
    - Considered modifications are made to assessment which ensures the participation of diverse learners in a fair way.
    - Marking bias is planned for and marking is based on evidence which utilises practices to avoid bias.

3. **Satisfactory Bias Awareness**
    - Assessment tasks are designed to meet the needs of the dominant culture, socio-economic group, or gender, with evidence of minor alterations for genders, socio-economic status and/or cultures.
    - Straightforward modifications are made to assessment which ensures the participation of diverse learners.
    - Teachers are aware of marking bias and marking is generally based on evidence of learning rather than the personality of the student.

4. **Minimal Bias Awareness**
    - Assessment tasks are designed to meet the needs of the dominant culture, socio-economic group, or gender.
    - Simple modifications are made to assessment which ensures the participation of some diverse learners.
    - Teachers have limited awareness of marking bias and marking can be biased towards the personality of the student.

5. **No Bias Awareness**
    - Assessment tasks are openly skewed to favour or marginalise a student or group of students.
    - No modifications are made to assessment.
    - Teachers have no understanding of marking bias and no effort is made to separate the personality of the student from the marking.

Levels of Thinking

Have I catered for a range of ability by providing opportunities for a range of responses and thinking levels, using the necessary cognitive verbs?

1. **Outstanding Levels of Thinking**
    - Comprehensive assessment tasks are designed that allow students to engage at progressively higher cognitive demands. Students are supported through a range of thinking levels and verbs clearly articulate cognitive requirements using a theoretical framework (such as Bloom or SOLO taxonomy) to underpin the thinking involved.
    - The suite of assessments demonstrates that there are high expectations for all learners at all levels of learning and opportunities for extending all learners are strategically planned for.
    - Higher order thinking skills such as creativity, problem solving, abstract thinking, and synthesis of ideas across concepts and domains are included and assessed in all tasks.
    - The suite of assessment tasks is flexible and varied, strategically utilising a range of assessment modes.

2. **High Levels of Thinking**
    - Clear assessment tasks are designed that allow students to engage at progressively higher cognitive demands. Students are presented with a range of thinking levels and verbs articulate cognitive requirements using a theoretical framework (such as Bloom or SOLO taxonomy) to underpin the thinking involved.
    - The suite of assessments demonstrates that there are expectations for all learners at all levels of learning and opportunities for extending all learners are planned for.
    - Higher order thinking skills such as creativity, problem solving, abstract thinking, and synthesis of ideas across concepts and domains are included and assessed in tasks.
    - The suite of assessment tasks is flexible and varied, covering a range of assessment modes.

3. **Satisfactory Levels of Thinking**
    - Assessment tasks are designed around the thinking progression of the Achievement Standard. Students are presented with a range of thinking levels and verbs articulate cognitive requirements.
    - The suite of assessments demonstrates that there are some expectations for most learners to extend their thinking at all levels of learning.
    - Questions allow for application of knowledge, understanding and skills and opportunity to demonstrate higher order thinking skills within the learning domain.
    - Assessment demonstrates some assessment modes.

4. **Minimal Levels of Thinking**
    - Assessment tasks are limited for the top students who are unable to show the extent of their thinking.
    - The suite of assessments demonstrates that there are minimal expectations for learners to extend their thinking.
    - Questions are restricted and narrow in conception with minimal opportunity for the student to demonstrate higher order thinking skills or application of knowledge, understanding and skills.
    - Assessment is generally in one mode with some small changes to make each task different.

5. **No Levels of Thinking**
    - Assessment tasks are one dimensional and do not encourage a range of thinking levels.
    - The suite of assessments demonstrates that there are no expectations for learners to extend their thinking.
    - Assessment is in one mode only.

Student Engagement

Are the assessment tasks accessible and inclusive? Are they contemporary and relevant for student needs?

1. **Outstanding Student Engagement**
    - The suite of assessment tasks are strategically planned to engage students through a range of methods such as connection to student lived experiences, interests, or prior knowledge; student autonomy; real world problems; contemporary issues; collaboration opportunities; or resemblance to activities conducted by real practitioners.
    - Success and what is being asked of the student in all aspects of the tasks are clear to from the instructions and marking schemes.
    - There are extensive supports in place to ensure student wellbeing and engagement.

2. **High Student Engagement**
    - The suite of assessment tasks aims to engage students through a range of methods such as connection to student lived experiences, interests, or prior knowledge; student autonomy; real world problems; contemporary issues; collaboration opportunities; or resemblance to activities conducted by real practitioners.
    - Success and what is being asked of the student in the tasks is clear from the instructions and marking schemes.
    - There are supports in place to ensure student wellbeing and engagement.

3. **Satisfactory Student Engagement**
    - The suite of assessment tasks appropriately aims to engage students. This may include minimally utilising methods such as connection to student lived experiences, interests, or prior knowledge; student autonomy; real world problems; contemporary issues; collaboration opportunities; or resemblance to activities conducted by real practitioners.
    - Students have a rough idea what success in the tasks will look like and the path there.
    - There are adequate supports in place to ensure student wellbeing and engagement.

4. **Minimal Student Engagement**
    - Not all assessment tasks aim to engage students. There is only token utilisation of methods such as connection to student lived experiences, interests, or prior knowledge; student autonomy; real world problems; contemporary issues; collaboration opportunities; or resemblance to activities conducted by real practitioners.
    - Students are rather unclear what success in the tasks will look like or there is unclarity surrounding what the student is expected to do in areas of the assessment tasks.
    - There are limited supports in place to ensure student wellbeing and engagement.

5. **No Student Engagement**
    - The suite of assessment tasks do not aim to engage students.
    - Success in the tasks are unclear. There is a lack of clarity of what the student is expected to do in the tasks or how to begin.
    - There are no supports in place to ensure student wellbeing and engagement.

Academic Integrity

Is it easy for students to access someone else's work and claim it as their own? How can I prevent this?

1. **Outstanding Academic Integrity**
    - Students are required to engage in genuine deep learning at a level of challenge appropriate to the student and tasks make provision for sense making or knowledge construction. Assessment is designed to ensure authenticity from students and requires individualised responses.

- Students are well educated as to what constitutes academic integrity as evident in a highly considered Program of Learning. Expectations in regard to plagiarism, assistance by others and referencing are made clear to students. Students declare that work is their own.
- Assessment tasks are not reused.

2. High Academic Integrity

- Academic integrity is discussed with students with expectations with respect to academic integrity and the consequences of cheating or plagiarising made clear.
- Students are informed as to what constitutes academic integrity as evident in a considered Program of Learning.
- Assessment is designed to encourage original thinking from students and require individualised responses that will be different. Expectations in regard to plagiarism, assistance by others and referencing are conveyed to students. Students declare that work is their own.
- Assessment tasks are not wholly reused with important aspects changed.

3. Satisfactory Academic Integrity

- Academic integrity is discussed with students in a general sense as evident in an appropriate Program of Learning.
- Assessment is designed so that a majority of the assessment encourages original thinking from students or requires individualised responses. Expectations in regard to plagiarism, assistance by others and referencing are referred to in the assessment task description. A statement is included in the task description stating that submitted work is declared as own work.
- Assessment tasks have some aspects changed from year to year.

4. Minimal Academic Integrity

- Minimal evidence of academic integrity processes in place.
- Assessment allows for the possibility of identical responses from students. Expectations in regard to plagiarism and referencing are inconsistent or applied inconsistently. The assessment task does not state that submitted work is declared as own work.
- Assessment tasks are largely the same from year to year.

5. No Academic Integrity

- Academic integrity is not mentioned in any documentation.
- Assessment requires identical responses from students.
- Expectations in regard to plagiarism and referencing are not addressed. Students do not make a declaration of own work.
- Assessment tasks are the same from year to year.

Table of criteria

When planning the suite of assessment items for a unit, teachers should consider the validity of the assessment by applying the six criteria and answering this question: **Is the suite of tasks fit for purpose?**

Criteria Consideration	Y/N
Coverage of BSSS Accredited Courses Do the assessment tasks cover the curriculum I am aiming to cover?	
The suite of assessment tasks is aligned with Achievement Standards, unit goals and content descriptions. Assessments are not too big: assessing irrelevant content or criteria; nor too small: missing important content or criteria.	
Assessment is strategically planned so that it does not distort the intent of the curriculum, aligning with and developing the skills evident in course goals.	
The general capabilities and cross curriculum priorities are seamlessly integrated into the suite of assessment tasks.	
Reliability Is this suite of assessment reliable? Do they have a clear marking schemes and rubrics?	
Assessment tasks and conditions are strategically designed to remove sources of non-relevant variation in measurements.	
Assessment conditions are clear, consistent, and enforced.	
In units with multiple classes and teachers, consistency around messaging and assistance is ensured through embedded practice.	
Assessment instructions and marking schemes are clear and unambiguous to ensure consistency in student and marker interpretation.	
Consistency of marking is ensured through a range of moderation processes such as: • Single marker of task or sub-task • Double marking or sample double marking • Marking calibration activities • Sample scripts/responses for all grade levels • Comprehensive in-school moderation discussions	
Bias Awareness Are there any inherent biases (gender, socio-economic status, culture, disability) evident in the suite of assessment?	
The suite of assessment tasks is strategically designed to be sensitive and empowering for all students, catering for the diverse needs of gender, socio-economic status, disabilities and/or cultures.	
The suite of assessment does not marginalise or favour a student or group of students, or advantage or disadvantage certain background knowledge or ways of thinking.	
Intercultural understanding and consideration of alternate points of view is promoted in task design.	
Marking bias is strategically planned for with comprehensive practices to mark objectively and avoid bias.	

Required reasonable adjustments are highly considered to ensure participation of diverse learners in a fair and equitable way.	
Levels of Thinking Does the suite of assessment allow for the full range of levels of thinking to be demonstrated?	**Y/N**
Comprehensive assessment tasks are designed that allow students to engage at progressively higher cognitive demands. Students are supported through a range of thinking levels and verbs clearly articulate cognitive requirements using a theoretical framework (such as Bloom or SOLO taxonomy) to underpin the thinking involved.	
The suite of assessments demonstrates that there are high expectations for all learners at all levels of learning and opportunities for extending all learners are strategically planned for.	
Higher order thinking skills such as creativity, problem solving, abstract thinking, and synthesis of ideas across concepts and domains are included in all tasks.	
The suite of assessment tasks is flexible and varied, strategically utilising a range of assessment modes.	
Student Engagement Is the suite of assessment tasks accessible and inclusive? Are they contemporary and relevant for student needs?	
The suite of assessment tasks is strategically planned to engage students through methods such as: • Connection to student lived experiences, interests, or prior knowledge. • Student autonomy • Real world problems • Contemporary issues • Collaboration opportunities • Resemblance to activities conducted by real practitioners.	
Success in the tasks is clear to the student from the instructions and marking schemes.	
There are extensive supports in place to ensure student wellbeing and engagement.	
Academic Integrity Is it easy for students to access someone else's work and claim it as their own? How can I prevent this?	
The suite of assessment ensures high levels of academic integrity requiring authenticity from students and individualised responses.	
Students are well educated as to what constitutes academic integrity as evident in a highly considered Program of Learning. Expectations in regard to plagiarism, assistance by others and referencing are made clear to students.	
Students declare that work is their own.	
Assessment tasks are not reused from year to year.	

Process of Evaluating Quality Assessment Tasks

Process of Evaluating Quality Assessment Tasks

When evaluating the suite of assessment items for a unit, teachers should evaluate the validity of the assessment by applying the six criteria and answering this question: **Is the suite of tasks fit for purpose?**

Coverage of BSSS Accredited Courses

Coverage of the curriculum refers to assessment that measures key knowledge, understandings and skills conveyed in the unit goals, content descriptions and Achievement Standards.

1. **Outstanding Coverage of BSSS Accredited Courses** – Assessment tasks are strategically planned for alignment with Achievement Standards, unit goals and content descriptors. Assessments are not too big: assessing irrelevant content or criteria; nor too small: missing important content or criteria.

2. **High Coverage of BSSS Accredited Courses -** Assessment tasks are thoughtfully planned. Assessments are not too big: assessing irrelevant criteria; nor too small: missing important criteria.

3. **Satisfactory Coverage of BSSS Accredited Courses -** Assessment tasks are appropriately planned. Assessments are not too big: assessing irrelevant criteria; nor too small: missing important criteria.

4. **Minimum Coverage of BSSS Accredited Courses –** Assessment tasks require refinement. Assessments are uneven. Some tasks are either too big: assessing irrelevant criteria; or too small: missing important criteria.

5. **No Coverage of BSSS Accredited Courses –** Assessment tasks are unplanned. Assessments are uneven. Some tasks are either too big: assessing irrelevant criteria; or too small: missing important criteria.

Reliability

Reliability of assessment refers to minimising variance of non-relevant factors in assessment measurements through clarity of instructions, standardisation of assessment conditions, comprehensive and clear marking guides, school-based moderation, and consensus of grade level evidence.

1. **Outstanding Reliability -** Assessment tasks and marking are strategically designed to remove all sources of non-relevant variation in measurements.

2. **High Reliability -** Assessment tasks and marking are thoughtfully designed to remove sources of large variation in measurements except for the key knowledge, skills and understandings of the student in the unit.

3. **Satisfactory Reliability** - Assessment tasks and marking are designed with some consideration of reducing variation in measurements caused by factors other than the key knowledge, skills, and understandings of the student in the unit.

4. **Minimal Reliability -** Assessment tasks and marking are designed with minimal consideration of reducing variation in measurements caused by factors other than the key knowledge, skills, and understandings of the student in the unit.

5. **No Reliability -** Performance in the assessment tasks is largely determined by factors other than the key knowledge, skills, and understandings of the student in the unit.

Bias Awareness

Bias awareness in assessment refers to reducing assessment that marginalises or favours students or student groups on factors such as gender, socio-economic status, disability, ethnicity, or which privileges a view of knowledge.

1. **Outstanding Bias Awareness -** The suite of assessment tasks is strategically designed to be sensitive and empowering for all students, catering for the diverse needs of gender, socio-economic status, disabilities and/or cultures, and that do not marginalise or favour a student or group of students, or advantage or disadvantage certain background knowledge or ways of thinking.

2. **High Bias Awareness -** The suite of assessment tasks is designed that promote the diverse needs of gender, socio-economic status, disabilities and/or cultures, and that do not marginalise or favour a student or group of students, or advantage or disadvantage certain background knowledge or ways of thinking.

3. **Satisfactory Bias Awareness -** The suite of assessment tasks is designed to meet the needs of the dominant culture, socio-economic group, or gender, with evidence of minor alterations for genders, socio-economic status and/or cultures, and doesn't overly advantage or disadvantage certain background knowledge or ways of thinking.

4. **Minimal Bias Awareness -** The suite of assessment tasks is designed to meet the needs of the dominant culture, socio-economic group, or gender.

5. **No Bias Awareness -** Assessment tasks are openly skewed to favour or marginalise a student or group of students.

Levels of Thinking

Levels of thinking refers to assessment that makes provision for a range of cognitive demands and levels of understanding based on theoretical models, such as the Bloom or SOLO taxonomies.

1. **Outstanding Levels of Thinking -** Comprehensive assessment tasks are designed that allow students to engage at progressively higher cognitive demands. The suite of assessments demonstrates that there are high expectations for all learners at all levels of learning and opportunities for extending all learners are strategically planned for. Assessment tasks are flexible and varied, promoting a range of assessment modes.

2. **High Levels of Thinking -** Clear assessment tasks are designed that allow students to engage at progressively higher cognitive demands. The suite of assessments demonstrates that there are expectations for all learners at all levels of learning and opportunities for extending all learners are planned for. Assessment tasks are flexible and varied, covering a range of assessment modes.

3. **Satisfactory Levels of Thinking -** Assessment tasks are designed around the thinking progression of the Achievement Standard. The suite of assessments demonstrates that there are some expectations for most learners to extend their thinking at all levels of learning. Assessment demonstrates some assessment modes.

4. **Minimal Levels of Thinking -** Assessment tasks are limited for the top students who are unable to show the extent of their thinking. The suite of assessments demonstrates that there are minimal expectations for learners to extend their thinking. Assessment is generally in one mode with some small changes to make each task different.

5. **No Levels of Thinking -** Assessment tasks are one dimensional and do not encourage a range of thinking levels. The suite of assessments demonstrates that there are no expectations for learners to extend their thinking. Assessment is in one mode only.

Student Engagement

Student engagement refers to assessment that promotes student involvement and ownership.

1. **Outstanding Student Engagement -** Assessment tasks are strategically planned to engage students. Assessment tasks are explicitly and purposefully connected to contemporary issues, student lived experiences, interests, or prior knowledge. The suite of assessment tasks clearly supports student ownership.

2. **High Student Engagement -** Assessment tasks are thoughtfully planned to engage students. Assessment tasks are explicitly connected to contemporary issues, student lived experiences, interests, or prior knowledge. The suite of assessment tasks supports student ownership.

3. **Satisfactory Student Engagement -** Assessment tasks are appropriately planned to engage students. Assessment tasks are implicitly connected to contemporary issues, student lived experiences, interests, or prior knowledge. The suite of assessment tasks supports student ownership.

4. **Minimal Student Engagement -** Assessment tasks require refinement. Assessment tasks are connected minimally to contemporary issues, student lived experiences, interests, or prior knowledge. The suite of assessment tasks supports student ownership but are not aligned to unit goals, content descriptions and Achievement Standards.

5. **No Student Engagement -** Assessment tasks are unplanned. Assessment tasks lack connection to contemporary issues, student lived experiences, interests, or prior knowledge. The suite of assessment tasks does not support student ownership through choice, decision making opportunities and procedural choices and are not aligned to unit goals, content descriptions and Achievement Standards.

Academic Integrity

Academic Integrity refers to assessment that promotes genuine and original work from students.

1. **Outstanding Academic Integrity -** Students are required to engage in genuine deep learning at a level of challenge appropriate to the student and tasks make provision for sense making or knowledge construction. Assessment is designed to ensure authenticity from students and requires individualised responses.

2. **High Academic Integrity -** Academic integrity is discussed with students with expectations with respect to academic integrity and the consequences of cheating or plagiarising made clear. Assessment is designed to encourage original thinking from students and require individualised responses that will be different.

3. **Satisfactory Academic Integrity -** Assessment is designed so that a majority of the assessment encourages original thinking from students or requires individualised responses. Expectations in regard to plagiarism, assistance by others and referencing are referred to in the assessment task description.

4. **Minimal Academic Integrity -** Minimal evidence of academic integrity processes are in place. Assessment allows for the possibility of identical responses from students. Expectations in regard to plagiarism and referencing are inconsistent or applied inconsistently.

5. **No Academic Integrity -** Academic integrity is not mentioned in any documentation. Assessment requires identical responses from students. Expectations in regard to plagiarism and referencing are not addressed.

Glossary

Term	Definition
Ability	having skills or proficiency in an area.
Academic Integrity	the assurance that student work is the genuine product of the student being assessed.
Appropriate	sound, suitable, relevant, some connection between variables, fulfilling foundation requirements, major issues addressed and planned.
Assessment Plan	comes from the Program of Learning and has 3 steps: Step 1: What do we want students to show evidence of? (what are the big ideas in the unit?) Step 2: What evidence do we need to collect to support these claims? Step 3: What task will allow this evidence to be collected?
Bias	favouring a student or students over others based on factors other than the key knowledge, skills, and understandings of the student in the unit.
Capability	the capacity of a student to perform a specific task.
Conceptual understanding	refers to students who grasp ideas in a transferrable way and apply ideas across domains.
Cognitive understanding	refers to the mental processes involved in gaining knowledge and comprehension.
Curriculum Coverage	assessment covers the curriculum and its intent without being 'too small' (construct under-representation) and failing to assess what it should or 'too big' (construct irrelevant variance) assessing things which it should not.
Embedded practice	seamless and frequent high-quality practice, which is evidence based, practice aligns with curriculum.
Fair and equitable	fair consideration means students receive non-discriminatory practices regardless of their personal situations. Equitable consideration means students receive the same opportunity

	to reach a specific objective while specific considerations are in place. See BSSS website for more information.
Knowledge, understanding and skills	knowledge is discrete information, ideas and facts, understanding is building connections between inter-related ideas of a concept, skills relates to the ability to apply knowledge and understanding in familiar and unfamiliar contexts.
Learning progression	a continuum that maps key stages in the development of a learning domain (e.g. reading and mathematics) from simple beginnings through to complex interpretations and applications.
Levels of Thinking	using theoretical frameworks such as Bloom's Taxonomy (1956), Anderson and Krathwohl's Taxonomy (Bloom's revised taxonomy) (2001) or SOLO Taxonomy (Biggs & Collis, 1982) to describe phases of understanding and application, and the interconnectedness with other concepts or ideas.
Marking Bias	factors which influence marking and cause marking to be skewed. See here for an overview.
Non- relevant variance in measurements	factors other than knowledge, understanding and skills in the unit.
Program of Learning	A plan that a college develops to implement a course for a subject to ensure that the content descriptions are taught and learned.
Reliability	valid inferences of student knowledge, understanding and skills in the domain, by minimising the influence of non-relevant factors in the measurement.
Sensitive and empowering	having or displaying an appreciation and consideration of others' points of view.
Strategic	deliberate, evidence based, future focused, acutely focused on the needs of students, having breadth and depth, alignment of all variables.
Student Engagement	students who are unmotivated to complete an assessment will not produce reliable or valid assessment results as they will not demonstrate what they truly know or understand.
Thoughtful	reflective, attentive to the present context, responds to the needs of students, has breadth, purposeful.

Types of marking/ in school moderation	see here for an overview of in-school marking and moderation.
Validity	how fit for purpose the assessment is for the domain being assessed.

References

Anderson, L. W. and Krathwohl, D. R., et al (Eds.) (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon.

Bae, S., & Kokka, K. (2016). *Student Engagement in Assessments: What Students and Teachers Find Engaging*. Stanford, CA. Stanford Centre for Opportunity and Policy in Education and Stanford Centre for Assessment, Learning and Equity. Retrieved from: https://www.newcastle.edu.au/__data/assets/pdf_file/0004/318550/student-engagement-assessments.pdf

Bennett, R. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18: 1, 5 – 25

Biggs, J. B., & Collis, K.F. (1982). *Evaluating the quality of learning. The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Black, P. & Wiliam, D. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5).

Bloom, B. S. (1956). Taxonomies of educational objectives. *Handbook 1. Cognitive Domain*. NY: McKay.

Bramley, T. & Dhawan, V. (2010). *Estimates of Reliability of Qualifications.* Cambridge Assessment.

Carr, M., McGee, C., Jones, A., McKinley, E., Bell, B., Barr, H., and Simpson, T. (2005). *Strategic Research Initiative Literature Review; The Effects of Curricula and Assessment on Pedagogical Approaches and on Educational Outcomes; Report to the Ministry of Education.* Retrieved from: https://www.educationcounts.govt.nz/__data/assets/pdf_file/0003/9273/The-Effects-of-Curricula-and-Assessment.pdf

Centre for Education Statistics and Evaluation (CESE). (2017). Cognitive load theory: Research that teachers really need to understand, NSW Department of Education

Charles Sturt University. (2020). *Designing For Academic Integrity*. Retrieved from: https://www.csu.edu.au/division/learning-and-teaching/home/assessment-and-moderation/assessment-resources-and-information/designing-for-academic-integrity

Chen, O., Kalyuga, S. & Sweller, J. (2017). The Expertise Reversal Effect is a Variant of the More General Element Interactivity Effect. *Educ Psychol Rev,* 29. 393–405.

Christodoulou, D. (2016). *Making good Progress? The Future of Assessment for Learning*. Oxford University Press

Cluskey Jr., G.R., Ehlen, C., & Raiborn, M. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics*, 4. Retrieved from: http://www.aabri.com/manuscripts/11775.pdf

Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Darr, C. (2005a). A hitchhiker's guide to validity, *Set: Research Information for Teachers*, *2*. 55-56. NZCER.

Darr, C. (2005b). A hitchhiker's guide to reliability. *Set: Research Information for Teachers, 2*. 59-60. NZCER.

Dirksen, D. (2013). Formative assessment: Reliability & validity. In D. Dirksen, *Student assessment: Fast, frequent, and formative* (Ch. 3, pp.17-28). Lanham, MD: Rowman & Littlefield Education.

Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre). (2004). *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes*. Retrieved from EPPI-Centre website: http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass_rv3.pdf?ver=2006-03-02-124720-170

Furtak, E. M., Morrison, D. E. B., and Kroog, H. (2014). Investigating the Link Between Learning Progressions and Classroom Assessment. *Science Education*, 98(4)

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York: Little, Brown and Co.

Griffin, P. (editor.) (2018). *Assessment for teaching* (Second edition). Cambridge University Press, United Kingdom [London]; New York, USA; Port Melbourne, VIC.

Hanover Research (2011). Effective Grading Practices in the Middle School and High School Environments.

Illinois Education Department (2015). *Guiding Principles For Classroom Assessment*. Illinois: Illinois State Board of Education.

Kirschner, P., Sweller, J. & Clark, R. (2006). 'Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching', *Educational Psychologist*, vol. 41, no. 2, pp. 75-86

Masters, G., & Forster, M. (1996). Common 'errors'. In E. Recht (Ed.), *Developmental assessment: Assessment* resource kit (pp. 36-39). Camberwell, Australia: Australian Council for Educational Research.

Masters, G. (2013). Reforming Educational Assessment, Imperatives, Principles and Challenges. *Australian Education Review* No 57. ACER.

Masters, G. N. (2014). *Assessment: Getting to the essence*. Retrieved from Australian Council for Educational Research website: https://www.acer.edu.au/cari/articles/assessment-getting-to-the-essence

Moon, T. R. (2005). The role of assessment in differentiation. *Theory into Practice,* 44(3), 226-233.

Murillo, F.J., & Hidalgo, N. (2017). Students' Conceptions about a Fair Assessment of their Learning. *Studies in Educational Evaluation*, 53.

NSW Centre for Education Statistics & Evaluation (2015). *Re-assessing assessment.* Retrieved from: https://www.cese.nsw.gov.au//images/stories/PDF/Re-assessing_Assessment_v6.pdf

Nuthall, G. (2007). *The Hidden Lives of Learners*. NZCER Press: Wellington.

OECD. (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*: OECD.

Popham, W. J. (2018). *Classroom Assessment – What teachers need to know (8th Edition).* Allyn & Bacon.

Stevens, R., Ructinger, L., Liyanage, S., and Crawford, C. (2017). *Review of Contemporary Research in Assessment and Reporting*. NSW Department of Education

University of Tasmania (2018). *Minimising Plagiarism and Cheating*. Retrieved from: https://www.teaching-learning.utas.edu.au/assessment/choosing-and-designing-assessment-tasks/minimising-plagiarism-and-cheating

University of Waterloo (n.d.). *Encouraging Academic Integrity Online*. Retrieved from: https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/planning-courses/course-design/encouraging-academic-integrity-online

Wiggins, G. P., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units.* Alexandria, Va: ASCD.

Wiliam, D. (2013) Example of a really big mistake: Calling formative assessment formative assessment and not something like "responsive teaching". Twitter available at: https://twitter.com/dylanwiliam/status/393045049337847808 (accessed 6 July 2020)

Wiliam, D. (2014). Principled Assessment Design. *SSAT: Redesigning Schooling – 8*. Retrieved from: http://www.tauntonteachingalliance.co.uk/wp-content/uploads/2016/09/Dylan-Wiliam-Principled-assessment-design.pdf

Willingham, D. (2009). *Why Don't Students Like School*. Jossey-Bass, John Wiley & Sons.

Willingham, D. (2019). How to Teach Critical Thinking. *Education: Future Frontiers*. Occasional Paper Series