3 Adult Education and Competency-based Assessment

This chapter assists examiners with undertaking a standardised approach to all aspects of conducting flight tests and proficiency checks.

3.1 Flight test and proficiency check

The purpose of the flight test and proficiency check is to determine whether an applicant meets the skills, competency and proficiency requirements for the grant of a licence, rating or endorsement, or the ongoing proficiency of a rating or endorsement.

There is often confusion about the difference between skills, competency and proficiency. These terms are sometimes used interchangeably, but they have varying definitions:

Skills

A skill is an ability and capacity acquired through deliberate, systematic, and sustained effort to smoothly and adaptively carry out complex activities or job functions involving ideas or concepts, things (technical skills), and interaction with people (human factors/non-technical skills).

In summary, skills are specific learned activities such as the use of an aircraft navigation system.

Competency

Competency is a combination of skills, knowledge and attributes required to perform a task to the prescribed standard. It is the achievement of specified competency standards and performance criteria outlined in an appropriate syllabus, usually involving an objective yes/no assessment.

A competency-based approach typically includes a long list of items to train and assess. The downside is that it often misses how competencies work together in different combinations to produce a desired result. This desired result is considered ‘proficiency’.

Proficiency

Proficiency is the result or output of demonstrating a series of defined competencies. Rather than demonstrating knowledge about a concept or subject matter, it demonstrates a level of performance. Whereas competency is a baseline, proficiency may vary with currency, revision and experience.

One of the things that differentiates competency from proficiency is the factor of context: A person would be described as being proficient in the operation of an aircraft, if they are able to not only operate the aircraft, but to manage the aircraft effectively in a wide range of variables, situations and contexts, especially when under pressure.

Another factor that differentiates competency from proficiency is agility/flexibility: the ability to change and adapt in the light of new situations.

In summary, competency is a demonstration that you can do something: proficiency is a demonstration of how well and flexibly you can do something.

So, what does this mean for the examiner? The examiner can only determine if an applicant for the grant of a licence or rating is proficient after having observed a representative number of competencies in a range of situations and contexts.

3.1.1 Flight test and proficiency check aims

The aims of a flight test and proficiency check are to:

1. determine, through observation of practical demonstrations, that an applicant has acquired or has maintained the required level of proficiency for the qualification as prescribed in the Part 61 MOS.

2. provide feedback from the examiner to the HOO of information concerning items or components of the test or check that are most frequently below the desired competency, and those that consistently exceed the desired competency to help improve and standardise flight instruction processes and training outcomes provided by training providers.
3. assist in maintaining and, where possible, improving flight safety standards by having examiners display exemplary attitudes and behaviours during flight tests and proficiency checks.

It is essential that a common standard is applied by all examiners.

Because every flight is conducted in different and sometimes widely varying conditions and circumstances, the examiner must consider a range of factors which may affect the assessment process. However, it is not appropriate to make allowance for poor training. Applicants must be assessed according to the required competencies for the licence, rating or endorsement sought. Examiners must exercise sound judgement and impartiality throughout their duties, and clearly understand the principles, methods and factors affecting the assessment process.

3.2 Assessment principles and methods

The principles, methods and factors outlined below apply as much to knowledge assessment in the briefing room as to practical assessment in the aircraft or flight simulation training device (FSTD).

The role of the examiner is to observe the performance of the applicant without influencing their performance.

Applicants and assessment situations vary. An effective examiner must be able to adapt to different personality types and assessment situations.

To accurately assess levels of performance and provide constructive feedback to the applicant the examiner will need to use a number of assessment tools and techniques, including:

- the Flight Examiner Handbook
- CASA flight test forms
- operator test and proficiency check forms
- specific questions and/or activities
- simulation and real-world scenarios.

3.2.1 Principles of effective assessment

The term assessment is generally used to describe the process of gathering measurable information and evidence about the performance of an individual or team and comparing this with a defined set of competency standards. A judgement is then made as to whether the competency standards have been met.

Assessment is an essential and continuous (ongoing) component of the flight crew licensing process. An effective assessment provides critical information to the examiner and CASA as the regulator, as well as providing vital feedback to the applicant. Both the examiner and the applicant need to know how well the applicant is performing with reference to a clearly defined and acceptable level of competency.

A good assessment provides practical and specific feedback to applicants, including direction and guidance on how to raise their level of performance. Most importantly, a well-designed and effective assessment process contributes to the development of aeronautical decision-making and judgment skills by helping to develop the applicant’s ability to evaluate his or her own knowledge and performance accurately.

A well-designed and effective assessment also allows the examiner to assess the performance of the applicant over a comprehensive set of competency standards, thus highlighting the areas in which an applicant’s performance is not yet competent. If, however, several applicants have problems at the same point in the flight test, the examiner may recognise the need for clearer or more detailed instructions by the examiner, or special emphasis in the assessment of subsequent performance.

Alternatively, if several applicants demonstrate an unacceptable level of competency in a particular flight manoeuvre, such as basic instrument flying, the examiner might recognise the need for a detailed debrief with the training provider's HOO to address any organisational factors which may be present, such as inadequate instrument scan training, or poor instrument flight instruction.

An assessment may be made about the performance of an individual, or individual members operating as a team. Assessment involves a degree of judgement by the examiner against established principles of effective assessment.

An effective assessment must be rigorous, credible and defensible. A number of key assessment principles are recognised within the aviation and training system. The 4 key assessment principles are summarised below.

Table 2. Assessment principles

Assessment principles	
Validity	Validity is concerned with whether an assessment accurately measures what it is designed to measure and nothing else. That is, the scope of a flight test or proficiency check must be such that when applicants are assessed as proficient, they have met the competency requirements for the grant or revalidation of the licence, rating or endorsement.
Reliability	Reliability is concerned with whether an assessment measures consistently between repeated situations, to ensure comparable (similar) results between applicants over time. In the context of the flight test, assessment reliability ensures that 2 identical performances would result in the same assessment result.
Flexibility	Flexibility is concerned with making appropriate modifications to assessment procedures and methods to better suit the particular needs and personality of the applicant. In the flight test context, the examiner must also allow for variables to suit the context and environment in which the flight test is undertaken. For example, allowance should be made for flying accuracy in turbulent conditions. However, it must be noted that flexibility is intended to apply to the context and process of assessment, and not to the assessment standards.
Objectivity	Objectivity is concerned with ensuring that the examiner's personal opinions will not affect the outcome or assessment of the test. While it is inevitable that all flight test assessments are influenced to some degree by subjective opinions, examiners must ensure, as far as possible, that assessments are made in accordance with the applicable competency standards. Assessments will be less subjective, and therefore more valid, if the examiner has in-depth knowledge of the evaluation process and the expertise to accurately assess applicants without prejudice. To achieve objectivity, the assessment process should be well documented and should not go beyond the requirements for the grant or revalidation of the licence, rating or endorsement.

3.2.2 Types and methods of assessment

Table 3. Types of assessment

Types of assessment	
Recognition of prior learning	Recognition of prior learning (RPL) is the acknowledgement of knowledge, skills or competencies gained as a result of previous experience or formal training or study undertaken. In the aviation context, this is especially beneficial for applicants transferring formal qualifications gained through the Defence Forces or other National Aviation Authorities.
Diagnostic assessment	Diagnostic assessment is used to assist learners, and trainers to determine the training needs of the learner. A diagnostic assessment indicates the gap between the learner's current knowledge, skills and competencies, and the desired level of knowledge, skill or competency for a particular task or role.
Traditional assessment	Traditional assessment involves the kind of written or oral testing (e.g. multiple choice, matching) and grading that is most familiar to assessors and applicants. To achieve a passing score on a traditional assessment, the applicant generally has a set amount of time to recognise or reproduce memorised terms, definitions and data. There is usually only one correct answer.
Formative assessment	Formative assessment occurs progressively throughout a training program and is used to provide feedback to trainees on their progress during training. As the term suggests, formative assessment helps trainees to form the desired knowledge, skills and understanding that will eventually be needed to demonstrate the required competencies. The emphasis of formative assessment is on providing feedback to learners during the learning process. This is beneficial to learners in developing their own self-assessment skills for the improvement of their own performance and progress.
Summative assessment	Summative assessment is the most common form of assessment. It usually occurs at the end of a learning segment or formal training program. This is the formal process of collecting sufficient evidence to assess whether a person is competent in relation to a particular set of performance criteria. In the aviation context, summative assessments are used to determine competency for the grant of a licence, rating or endorsement. The flight test is a demonstration of the applied knowledge, skills and competencies, and comprises the assessment upon which granting of the formal qualification is based. The main approach to summative assessment is competency-based assessment, by which defined standards or specific performance criteria are used to assess the applicant's acceptable performance on a task. This type of assessment is preferable in the aviation context, since competence or the ability to meet performance criteria is desirable when assessing skilled performance.
Authentic or holistic assessment	Authentic assessment requires the applicant to perform real-world tasks and actively demonstrate a meaningful application of knowledge, skills and competencies. Rather than develop separate assessment tasks to assess each unit of competency, a holistic approach seeks to integrate the assessment of knowledge, skills and performance in one assessment task or activity. That is, the applicant must demonstrate in-

Types of assessment	
	depth knowledge by formulating a solution to demonstrate application of skills, or competency, rather than merely choosing an outcome or response.

Generally, there are several assessment methods that examiners may use to make judgements regarding an applicant's overall proficiency. These are not mutually exclusive, and in practice, are often used together to formulate a more accurate 'overall' assessment.

Examples of assessment methods:

- observation of actual performance such as observing a flight sequence
- using a range of different question types in order to assesses the applicant's ability to listen, interpret and communicate ideas about information
- simulating a situation such as role-play interaction between the applicant and examiner
- presentation by the applicant of a variety of evidence of previous experience or training which addresses current standards.

3.2.3 Dimensions of competency

The above-described types and methods of assessment may be used independently or in a more holistic way by conducting the assessment against all 'dimensions of competency'. This means that the assessment is not narrowly based on a specific task or skill, but embraces all aspects of proficiency, utilises some, or all, of the methods described above, and represents an integrated and holistic approach to the assessment.

The assessment may take into account the dimensions of competency outlined below.

Table 4. Dimensions of competency

Dimensions of competency	
Task skills	Performing at an acceptable level of competency. This may include carrying out individual tasks such as conducting a level turn.
Task management skills	Managing a number of different tasks at any one time. This involves being able to integrate several tasks to complete an outcome. For example, operating aeronautical radio whilst taxiing an aircraft.
Contingency management skills	Responding and reacting appropriately to unexpected problems, changes in routine and breakdown. For example, if the weather unexpectedly deteriorates, alternative strategies are employed to ensure a safe outcome.
Job or role environment skills	Fulfilling the responsibilities and expectations of the workplace. Each workplace is unique and requires the individual to be able to adjust to the environment in which they are working. This may include, for example, working with different flight crews, following workplace procedures, or complying with organisational policies.
Transferability skills	Transferring skills and knowledge to new situations and contexts. This requires the ability to adapt to different work situations and demands. For example, a pilot should be able to transfer baseline skills and knowledge from one situation to another. For example, instead of just assessing the applicant's performance while executing a level turn against the specified competency standard (task skills), it may be more realistic to observe the applicant performing the manoeuvre to avoid a simulated cloud bank

Dimensions of competency	
	(contingency skills) where the turn is required to position the aircraft to avoid the un-forecast weather and conduct a diversion procedure to an alternate aerodrome in accordance with the procedures contained within the company manuals (role and transferability skills).

By assessing all dimensions of competency, the skill is being applied to a new circumstance (transfer of skill), while managing a somewhat complex undertaking. This approach combines knowledge, understanding, problem solving, technical skills and application into the assessment.

3.2.4 Evidence as an assessment method

Evidence is the information gathered which, when matched against current standards, provides proof of competence or proficiency. Evidence can take many forms and be gathered from a number of sources. Evidence can be direct, indirect or supplementary.

Direct evidence

Examples of direct evidence:

- direct observation
- oral questioning
- demonstration of specific skills.

Indirect evidence

Examples of indirect evidence:

- assessment of qualities of a final product
- review of previous work undertaken
- written tests of underpinning knowledge.

Supplementary

Examples of supplementary evidence:

- testimonials from colleagues
- reports from supervisors
- work diaries or logbooks
- examples of reports or work documents.

No single form of evidence is better than another. Quality evidence should be provided by the applicant and should meet the 4 ‘rules’ of evidence below.

Table 5. Rules of evidence

Rules of evidence	
Valid	Relates to the competency standards for the licence, rating or endorsement being assessed.
Sufficient	Provides enough evidence to make a judgement about the competence of the individual in relation to the current standards for the licence, rating or endorsement being sought.
Current	Is recent enough to show that the evidence produced is still able to be applied to the current standards for the applicable licence, rating or endorsement.
Authentic	Can be verified that the evidence is the applicant's own work.

There are many potential sources of evidence which may be used to make judgements regarding an individual's competency. These are not mutually exclusive and, in practice, are often used together to formulate a more accurate 'overall' assessment.

3.2.5 Effective questioning techniques as an assessment method

Most assessment types described above will require some form of questioning. While factual recall questions provide an indication of an applicant's knowledge, it is important for the examiner to create a climate of enquiry and engagement in high quality, high order questioning if formative assessment is to be effective.

Effective questioning is vital because it makes the applicant's thinking visible. It identifies prior knowledge, reasoning ability and the specified degree of applicant understanding. Questioning techniques are below.

Table 6. Questioning techniques

Questioning techniques	
Key questions	A good way to foster a culture of inquiry is to open with a big question that gets the applicant to think critically about what they have learnt. By asking a big question you can start thinking that immediately engages the applicant about their learning and it can raise motivation. For example, you might ask the applicant how important they think effective non-technical skills are in their assessment.
Open and closed questions	Open questions require learners to think and formulate a response. If an examiner asks an applicant to explain why they performed that manoeuvre, then the applicant has to provide the explanation in their own words. This gives the examiner feedback especially if the examiner observes the applicant's body language. Closed questions usually only require the applicant to answer 'yes' or 'no' and as such are not particularly valuable.
The strategic pause	The thinking time at the 'pause' point is crucial, the quality of the response and the confidence level of applicant are raised by even a short amount of thinking time.

Questioning techniques	
Socratic questioning and Socratic circles	The 6 steps of Socratic questioning (named after the classical Greek philosopher Socrates) create a critical atmosphere that probes thinking: 1. Clarification: • Why do you say that? • Could you explain that further? 2. Challenge assumptions: • Is this always the case? • Why do you think that assumption holds here? 3. Evidence as a basis for argument: • Why do you say that? • Is there reason to doubt this evidence? 4. Viewpoints and perspectives, which challenge the applicant to investigate other ways of looking at the same issue: • What is the counter argument for…? • Can you look at this in another way? 5. Implications and consequences, given that actions have consequences, this is an area ripe for questioning: • But if that happened, what else would result? • How does… affect …? 6. Question the question, just when applicants think they have a valid answer, you can further challenge misunderstandings: • Why do you think I asked that question? • Why was the question important?

3.2.6 Types of questions to avoid

Asking, closed questions ‘Do you understand?’ or ‘Do you have any questions?’ has limited effect in an assessment environment. Assurance by the applicant that they do understand or that they have no questions provides no evidence of their comprehension, or that they even know the subject under discussion.

Table 7. Type of questions to avoid

Types of questions to avoid	
Puzzle	‘What is the first action you should take if a conventional gear airplane with a weak right brake is swerving left in a right crosswind during a full flap, power-on wheel landing?’
Oversize	‘What do you do before beginning an engine overhaul?’
Toss-up	‘In an emergency, should you squawk 7700 or pick a landing spot?’
Bewilderment	‘In reading the altimeter, you know you set the QNH for the nearest station pressure. If you take temperature into account, as when flying from a cold air mass through a warm front, what precaution should you take when in a mountainous area?’

Types of questions to avoid	
Trick questions	These questions will cause the applicant to develop the feeling that they are engaged in a battle of wits with the examiner, and the whole significance of the subject involved will be lost. An example of a trick question would be where the alternatives are 1, 2, 3, and 4, but they are placed in the reverse order to trick the applicant to inadvertently answer incorrectly. If attention to detail is an objective, detailed construction of alternatives is preferable to trick.
Irrelevant questions	Diversions, which introduce unrelated facts and thoughts, will only obscure the orderly assessment process. Answers to unrelated questions are not helpful in evaluating the applicant's knowledge of the subject at hand.

3.2.7 Questioning in the context of multi-crew operations

When considering an assessment of the knowledge requirements of 2 applicants, examiners should consider their question distribution strategy to ensure the knowledge requirements for both applicants are assessed in all required competencies. The examiner should ensure the questioning technique is robust to satisfy the testing of the knowledge requirements of both applicants. This is best achieved through questioning each applicant individually.

3.3 Factors affecting assessment

Two key considerations for the reliability and validity of flight test and proficiency check assessments are the accuracy and consistency of the examiner. The accuracy of assessments depends largely on the examiner's knowledge of the assessment criteria, the assessment scenarios, and the assessment methods to be used.

3.3.1 Inter-rater reliability

Inter-rater reliability refers to consistency (or agreement) between examiners regarding both their behavioural observations and their performance ratings. Agreement is important for consistent application of assessment criteria to maintain flight crew licensing standards and improve flight safety. Appropriately trained examiners should be interchangeable; the assessment should not be dependent on any particular examiner.

3.3.2 Accuracy

Two forms of accuracy are important: observation accuracy and rating accuracy.

Observation accuracy is the extent to which observers can correctly identify and record behavioural information. This is critical, as the assessment of non-technical skills (NTSs) typically requires examiners to observe and assess an applicant's NTS while the applicant is performing a simulated or actual task.

Rating accuracy is the extent to which the examiner assigns the correct rating (i.e. competent or not competent) to the particular level of performance observed. This is critical in order to provide a valid assessment of the applicant's skills.

An applicant's ability to perform a task is assessed by:

- determining that the performance achieves the required competencies in Schedule 5 or 6 of the Part 61 MOS
- referencing the performance criteria in Schedule 2 of the Part 61 MOS
- determining that the performance is within the prescribed tolerances in Schedule 8 of the Part 61 MOS.

One of the greatest inconsistencies in achieving inter-rater reliability is the examiner's perception of what is an acceptable degree of error and an acceptable deviation from tolerance (i.e. how long can

that deviation be sustained before the deviation is recognised and that immediate and appropriate corrective action is taken).

Where an observed error has not become safety critical, the examiner should look for evidence that the applicant has the skills to recognise and correct the error. When an applicant is demonstrating sound technique, but minor deviations occur outside the flight tolerances in Schedule 8 of Part 61 MOS, the examiner must permit an acceptable opportunity for the applicant to demonstrate NTS competencies in order to apply corrective action.

The examiner must not accept errors where tolerances are critical, such as descent below a Minimum Descent Altitude or a Decision Altitude.

3.3.3 Judgement errors and biases

In order to conduct effective and objective flight tests, the examiner requires not only a sound knowledge of the characteristics of assessment, but also a good understanding of possible personal bias and judgement errors that can occur throughout the assessment process.

Assessment errors stem from 2 categories of bias:

1. bias that the examiner may bring to the process
2. bias that the applicant being assessed may bring to the process.

Examiner biases

Examiner biases can cause judgment errors and thus influence the assessment of applicants or a particular group.

Table 8. Examples of examiner judgement errors and biases

Examiner judgment errors and biases	
Confirmation	The tendency to seek out information that supports a pre-conceived belief about the applicant, as opposed to remaining open to the applicant's abilities throughout the assessment process.
Central tendency	The tendency for all, or most applicants to be assessed as 'average'. The examiner really feels that the performance of most applicants is not as good as it should be and therefore 'underscores' an applicant's good performance.
Generosity	The tendency for most applicants to be assessed as competent. This could be caused by an examiner's desire to be known as a nice person, or their reluctance to cope with the possible emotional response in the event of a fail assessment.
Severity	The tendency for most applicants to be assessed as not competent. In this situation, the examiner may feel that the published standards are too low and they should make assessments against their own standards. Or the examiner may be overly critical of an applicant's performance, requiring too high a standard.
Halo effect	When the examiner's impression of the applicant influences the assessment of performance. For example, when testing a friend, acquaintance, high profile individual, or their own trainees, an examiner may assess the individual as competent in error.
Leniency	This is a form of halo effect. It has its source in the examiner's likes, dislikes, opinions, prejudices, moods and political or community influence of people.
Stereotyping	The tendency for an examiner to assume that a member of a group has certain characteristics (e.g. national culture, gender) without having actual information about that individual.

Examiner judgment errors and biases	
Logical error	This error occurs when an examiner assumes that a high degree of ability in one area means a similar degree of competence in another. A competent assessment of one or 2 items does not mean the applicant is also competent on all items to be tested. The full test must be completed and assessed.
Error of narrow criterion	This error occurs when there is a group of applicants to test. The examiner may, in these circumstances, rate each applicant against the others within the group instead of against the standards.
Error of delayed grading	By delaying a 'not competent' assessment which will terminate the test, examiners may award a final 'competent' assessment based upon the overall impression of the flight test. This results in an erroneous assessment and a flight test report that is of little value to the training system.
Blind spot	Examiners must be aware of their own personal biases (other than those listed above). For example, an examiner may believe an applicant is too young or their accent suggests a poor English proficiency. Both age and accents are irrelevant; rather it is the performance of the applicant on all aspects of the flight test that should be considered.
Fundamental attribution	The tendency for people to over emphasise personality-based explanations for behaviours observed in others while under emphasising the role and power of situational influences on the same behaviour. The examiner should always consider the context (environmental influences or task demands) in which the behaviour is being displayed.

Applicant biases

Applicant biases typically affect the applicant's view of their own abilities, and the post-assessment process.

Table 9. Applicant biases

Applicant biases	
Just-world hypothesis	The tendency for people to want to believe that the world is fundamentally just. This can be a potential problem when the applicant receives negative feedback, perhaps believing it to be unjust in this situation. How the examiner delivers negative feedback or a 'fail' result is therefore crucial.
Over confidence effect	This is when an applicant's subjective confidence in his or her judgments is greater than the objective accuracy of those judgments. Over-confidence causes people to overestimate their knowledge, underestimate risks, and exaggerate their ability to control events. Examiners may find that applicants challenge their post-test debrief, particularly if the applicant has an over inflated view about their performance.

Whilst all these biases and resulting errors may appear obvious on paper, they may not be obvious under flight-test conditions, especially as the judgement of the examiner may be obscured by a combination of 2 or more biases. Examiners must therefore be aware of these biases to consciously prevent them from influencing the validity of the flight tests and proficiency checks they conduct.

To avoid the above errors and biases, the examiner should consider adopting the following actions:

- Allow plenty of time for the flight test or proficiency check to observe the applicant over a sufficient time frame.
- Maintain the planned structured flight examination process. A structured and organised process for recording observations increases accuracy and reduces bias. Make sure that this process is repeatable.

- When making conclusions about an applicant's performance, do not rely completely on memory; always refer to notes.
- Consider only information observed and do not make any assumptions.
- Consider only information relevant to each technical and non-technical skill that is being assessed.
- Consider if any rating errors are occurring. Ask yourself: 'Are my ratings affected by halo, leniency, severity, or any other types of bias?'
- Ensure a structured approach is taken when providing a debrief or giving feedback.