

Technical Guidance Material for the development and revision of examinations and test items

Subject: TECHNICAL GUIDANCE MATERIAL FOR THE DEVELOPMENT AND REVISION OF EXAMINATIONS AND TEST ITEMS

Date: 2019-01-23

1. APPLICABILITY

The guidelines and *instructions contained in this document apply to:

- a) Internal (SACAA) Examiners involved with or responsible for the development and/or revision of theoretical knowledge examination questions towards the issue of ICAO Annex 1 licences and ratings,
- b) External (Designated) Examiners of all disciplines responsible for the development and/or revision of theoretical knowledge examination questions towards the issue of ICAO Annex 1 licences and ratings,
- c) SACAA Examiners and Designated Examiners of all disciplines responsible for the conduct of practical oral examinations towards the issue of ICAO Annex 1 licences and ratings,
- d) Other parties associated with Aviation Training Organizations and which are responsible for the development or revision of theoretical knowledge and practical oral examinations, specifically towards the the issue of ICAO Annex 1 licences and ratings and
- e) Interested parties involved in aviation training or testing that may benefit from the guidance contained herein.

**Note - The prescriptions in this document refer to summative examinations only. Formative (assessment-for-learning) examinations are consequently excluded, although the fundamental developmental principles may still apply and be useful.*

2. PURPOSE

- a) The purpose of this document is to provide guidance to internal (SACAA) technical personnel (inspectors and testing standards officers) and external industry specialists (Designated Examiners and Aviation Training Organizations), on the
 - i. development,
 - ii. moderation,
 - iii. monitoring and
 - iv. revision of both individual examination questions and the regulatory examinations in which they are presented.
- b) Integrated into the guidelines mentioned above and clearly identified as such, are standard procedures to be used for the development of examination items and papers presented by Designated Examiners.
- c) The guidelines contained in this document should be understood and interpreted in the context of the wider regulations dealing with the requirements for the specific licence or rating. This means that the examiner should be familiar with the

prescriptions contained in the SA-CAR and document SA-CATS, the syllabus, the approved scope of training in the ATO Training and Procedures Manual (TPM), where applicable, and other advisory material, such as Technical Guidance Material and AIC's. Prescribed or suggested study material forms an integral part of any training course and will play a major role in the question development process. Finally, any additional technical and academic reference works can play a huge role in helping a question-writer develop high-quality exam items. A bibliography listing the academic reference books used in the design of this document may be found at the end. The user may find this list useful. A multitude of quality educational literature is available, some of it free of charge on-line.

3. ADDITIONAL GUIDANCE

Users are referred to the following documents for additional guidance:

- a) Technical Guidance Material for Designated Flight Examiners and Oversight and
- b) SACAA departmental PEL procedures (for SACAA (internal) Examiners).

4. EXCLUSIONS

The following examination-related and developmental aspects are excluded from the scope of this document:

- a) General curriculum development,
- b) Examination administration,
- c) Hardware and software set-up and programming and
- d) Language proficiency testing.

5. TERMS AND ABBREVIATIONS:

ABBREVIATION	DESCRIPTION
AO	Authorized Officer
ATO	Aviation Training Organization
CAT	Computerized Adaptive test
CBT	Computer-based test
CFT	Computerized Fixed Test
DE	Designated Examiner
DFE	Designated Flight Examiner
ICAO	International Civil Aviation Organization
MCQ	Multiple choice question
MRQ	Multiple response question
PEL	Personnel Licensing department (SACAA)
POI	Principal Operations Inspector
SACAA	South African Civil Aviation Authority
TGM	Technical Guidance Material
TPM	Training and Procedures Manual
YNTF	Yes-No /or True-False type question

6. *GETTING ONTO THE SAME PAGE - TERMINOLOGY AND DEFINITIONS

***Note** - Users should refer to the document "Technical Guidance Material for Designated Flight Examiners and Oversight" for a more comprehensive discussion of the complexity of defining the term(s) assessment/evaluation .

- a) Compulsory schooling, the application of modern educational principles in the workplace and the accessibility of online information mean that many (if not most) working people have some exposure to modern educational terms, sometimes with an insufficient grasp of the theory behind them.
- b) Worldwide, many educational terms are used loosely in formal and informal environments. Educational terms such as curriculum, learning programme, test, examination (or exam), assessment, evaluation, check, grading scale, bell curve and standardized test mean varying things to different people. Even academic literature contains diverging interpretations of some of these terms. There are considerable differences, for example, in the ways that terms are used in the U.S.A, UK and South Africa. To exacerbate matters, ICAO documentation also contains several ambiguities, contradictions and outdated terms.
- c) For these reasons, it is important to clarify from the start how certain terms will be used in this document. Specifically, the meaning of the terms “assessment” and “evaluation” in this document is intended to be consistent with its widespread use in the educational field in South Africa. Although the use of these and other terms in this document is justified and reasonable, this TGM is neither intended as a purely academic work, nor as a dogmatic argument for or against the specific use of debatable terminology. Rather, this TGM is intended to be practical and helpful to technical aviation personnel. The following descriptions should be viewed as explanatory, not rigid definitions.

TERMINOLOGY	DESCRIPTION
Assessment:	Assessment is done when a person's performance is judged against a standard (or criterion). Here it is not important how well someone performed, but simply whether the set standard was achieved or not. In other words, there is no partial degree of success (like a percentage) involved - the standard is either met or it is not. The person conducting the assessment is called an assessor. This is typically done in a competency-based, performance-based or outcomes-based environment.
Evaluation:	This is done by an evaluator and typically involves making a value judgement about an object or a programme. In other words, evaluation is used to test how well someone did in a test or exam or to what extent a goal was achieved. Here, performance is not measured against a fixed standard, but against a normative benchmark, in other words a value judgement is involved. In this document, this term will be used to indicate assessment in a non-competency-based training system. Evaluation is typically used in content-based tests, such as most SACAA flight crew examinations and practical skills tests.
Assessment/ Evaluation:	<p>The emphasis of this TGM is not supposed to be academic, but rather instructional, directive and informative. Consequently, it is not productive to needlessly debate the use of the terms “Assessment” and “Evaluation” as far as they apply in the South African Personnel Licensing aviation environment and the multitudes of licensing activities. For this reason, except where specifically required, the terms are not used independently, but in combination and separated with a “/” symbol. Examples are “Assessment/Evaluation” and “Assess/Evaluate.” This is compatible with the approach taken by the academic Professor Arend Carl (2012: 99).</p> <p>The terms “assessment”, “evaluation” or “assessment/evaluation” are umbrella terms, including a variety of processes, methods, techniques, instruments and tools to gather evidence and make a judgement about a candidate's competence. This means that an “examination”, a “practical skills test”, a “demonstration” and a “proficiency check” are all different kinds of assessments/evaluations. When the umbrella term is therefore used, it may include any of the variations of “assessment/evaluation.”</p>
Summative examination	In the context of this TGM, a summative examination is one component of the end-of-training programme assessment/evaluation phase of an aviation licensing activity. Such an examination is therefore not to be viewed as informal or formative, but as a final judgement of competence. It should not be confused with a feedback-related testing process. Fein (2012: 35) and Dreyer (2014: 17) make this clear.

Examination (or exam):	<p>In this document, the term is used to refer to any assessment/evaluation of theoretical knowledge. The broad use of the term in this document includes any of the following types:</p> <ul style="list-style-type: none"> ▪ oral examinations, ▪ written (paper-based) and ▪ electronic (computer-based). <p>More details are given in the section dedicated to a discussion of each type later in this document.</p>
Examiner:	<p>ICAO uses the term examiner to refer to an aviation technical specialist that is responsible for aviation training development and testing, specifically with reference to internal (CAA) technical personnel. In the South African context, the SACAA refers to its equivalent internal technical personnel as PEL inspectors or Testing Standards Officers. Similarly then, a designated examiner is a technical specialist from the local aviation industry to whom the CAA has delegated certain functions.</p> <p>However, consistent with South African use, only the term Designated examiner is used in this document. For explanation, the following extract from ICAO Document 9379 is quoted:</p> <p><i>"The tasks related to the examination function may also be delegated externally, under the oversight of the CAA, to an organization or an appropriate person (a "designated examiner" with appropriate qualifications)"</i> (ICAO Document 9379, 2012: 1-2-2).</p>
Test:	<p>In the South African aviation context, the term <i>"test"</i> normally refers to a practical evaluation or assessment. For flight crew, for example, such a practical test is normally done after all theoretical knowledge examinations have been passed. In such a case, the <i>"test"</i> functions as the final and summative assessment/evaluation towards the issue of a licence or rating.</p> <p>In this TGM specifically, however, the term is used more loosely to match its broader use in the educational, non-aviation environment. This use is also more user-friendly, intuitive and corresponds closely to the use of the term in academic literature, especially that of American origin. The term will consequently be used to refer to any kind of theoretical knowledge assessment/evaluation. Used in this way, it is synonymous with the phrase <i>"theoretical knowledge examination."</i></p>
Exam item:	<p>This term is sometimes used in literature and has the same meaning as <i>"exam question."</i> International academic literature sometimes uses the term <i>"test item"</i> instead.</p>
Objective item:	<p>The University of Pretoria (2011: Terminology), defines an objective item as follows:</p> <p><i>"Objective items are defined as structured items with limited responses. These items include a variety of selected response questions such as multiple choice, multiple response, matching, fill-in-the-blank and hot-spot items.</i></p> <p><i>The student is required to select one or more answers from options provided, or supply a single word, number or symbol. The structuring of the item and the limitation of the format in which the answers are provided, leads to the fact that the questions are marked accurately and objectively."</i></p>
Developer:	<p>This is a blanket term, referring to the person developing a test item, irrespective of his or her position, designation, employment or the kind of item being developed.</p>
Computer-based Test (CBT):	<p>Generally, the term "Computer-based Test" refers to a variety of examinations conducted on a computer platform, using appropriate software. For the purposes of this document, a CBT refers to a theoretical knowledge examination conducted either on a fixed-base computer or on another electronic device, such as a laptop or tablet. The definition includes a hard copy examination completed on a marking sheet, but generated and marked by a computer system.</p>

7. EXAMINATION PRINCIPLES

- a) Irrespective of the type or application of a South African examination or who it is conducted by, certain fundamental principles remain important. This section gives an overview of the philosophy that should govern aviation examinations in South Africa.

- b) As a starting point, it is useful to ask what an examination is. What is its purpose? Certainly, it is not intended to be
- a mere box-ticking exercise,
 - a formality to get over with,
 - an abstract, meaningless academic exercise,
 - a money-making scheme or
 - an obstacle to a candidate's success.
- c) What is an examination then? It is probably best to let the International Civil Aviation Organization (ICAO) do the explaining. The organization states it succinctly:
"The issue of a licence is evidence that the State is satisfied that the holder has demonstrated an internationally acceptable degree of competency. However, that competency is the result of the applicant's training and capability, not of the examinations or tests passed. Rather, the examination system should be considered an audit of an applicant's knowledge and skill to confirm that the training system has succeeded. By implication, a well-designed system of theoretical examination and practical testing will drive the training system to produce licence applicants of the appropriate standard."
 (ICAO Doc 9379, 2012: II-5-1)
- d) The purpose of developing examinations is not proverbially to "catch people out" but to test their knowledge. However, as a part of proper examination, there is room for thoroughly assessing/evaluating knowledge. This might seem like "misleading" examinees in the way you compile the questions. However, if this is done fairly and transparently and in to test subject matter comprehension, it is definitely ethical and reasonable. It is not acceptable to try and "trick" people with silly, unnecessary or ambiguous catches, words or phrases that distract from the actual knowledge testing.
- e) In the educational and training world, a multitude of assessment/evaluation methods are available. Depending on the industry and application, examination developers may, for example, use examinations, practical tests, demonstration, role play, oral examinations, Computer-based tests, essays, assignments, case studies, model-building projects and research projects.
- f) In the South African aviation assessment/evaluation environment, the methods used are more limited and mostly entail written, oral and CBT examinations, practical demonstrations and simulation. As stated earlier, this TGM specifically provides information and guidance in respect of theoretical knowledge examinations.
- g) Which type of assessment/evaluation method is best? The University of Pretoria gives useful guidance (2011: Introduction page):
"A good lecturer endeavours in his/her total assessment plan to test all the different levels of knowledge. Oral examinations are eminently useful in testing the highest levels of Bloom's taxonomy, namely synthesis and evaluation. Written examinations may be used for the whole spectrum. Objective items are eminently suited for testing the lower levels. However this does not imply that higher levels cannot be tested in this way, although it is more difficult to formulate such questions."
- h) The principle illustrated in the quote above is that the purpose of an examination should inform everything else that follows. The developer could ask:
"What is the best way to test ...?"
- i) All questions should test for the minimum acceptable knowledge standard. It is a misconception that an examination should cater for a variety of difficulty levels and degrees of examinee competence. Testing is done in respect of a minimum certification level.
- j) All questions should test for the minimum acceptable knowledge standard. It is a misconception that an examination should cater for a variety of difficulty levels and degrees of examinee competence. Testing is done in respect of a minimum certification level.

- k) The pass rate for most SACAA examinations is 75%. This means that a 25% error is already allowed. In the high-stakes aviation testing environment, it is not dangerous to dilute an examination to help poorly prepared candidates. Don't be too soft!
- l) Questions should have practical relevance as far as is compatible with the aim of the examination. In cases where pure theoretical knowledge is assessed/evaluated, it is still important not to test peripheral or trivial aspects. It is a common error that developers feel that an examination should be made more difficult by adding questions on incidental subject matter.
- m) The main focus of an examination should always be core and operational knowledge. Do not waste empty questions that only test rote, overly abstract or irrelevant knowledge. Make each item count.
- n) Item quality is important and should ideally have priority above quantity. However, practically it is inefficient to spend too much time on a question. Do not over-design an item. Often, a minimum number of questions is necessary for an examination to be usable.
- o) Do not be shy to compile authentic, integrated i.e. scenario-based questions. These are excellent testers of real-life understanding. Some topics may not allow this – this is fine.
- p) The SACAA does not presently conduct any Open book examinations. However, this practice may be vastly superior to a Closed book examination in some areas. In subjects assessing/evaluating knowledge of aviation legislation, some topics might warrant an Open book approach. This is particularly applicable to subject matter that does not have to be committed to memory. Emergency and operationally critical subject matter might not be compatible with an Open book approach.
- q) Proper examination instructions to candidates are crucial to ensuring an accurate and fair examination.
- r) In the Questionmark White Paper titled *"Five Steps to Better Tests,"* several helpful tips for test item development are summarized (Peterson, D. 2013: 7-9).

They are:

- i. Insist on creating clear, readability and unambiguous items - Is it clear what the item is asking?
 - ii. Use standardized styling - *"Use a style guide for consistency"*
 - iii. With MCQ-type questions, the distractors should be believable
 - iv. Keep the impact of the verb you use in mind - the cognitive level of the question can be raised or lowered
 - v. Avoid bias and stereotypes
 - vi. Make sure that spelling and grammar are in order
- s) **A few practical instructions follow:**
- i. The maximum mark allocated to a question has to be in proportion to the number of correct choices, the difficulty level of a question or the time required for a proper response;
 - ii. The time allowed to answer an examination should be carefully calculated and confirmed by practical pre-launch testing;
 - iii. When additional visual images, graphs or other attachments are part of an item, additional time and marks should be allocated;
 - iv. The time allocated to calculation questions especially should be carefully confirmed;
 - v. The practice of negative marking has some educational merit, but is not currently applied by the SACAA and
 - vi. YNTF questions should be avoided.

8. PRINCIPLES OF LEARNING

- a) Although a thorough discussion of learning principles is beyond the scope of this document, it is important to understand a few basic, modern learning principles if an accurate assessment or evaluation is to be done.
- b) A few of the fundamentals of learning will be discussed in this section, although the user is encouraged to read much more widely than this document in order to be well-informed about teaching, training and learning. Despite being a specialized field of work, aviation is not unique as far as the application of educational and training principles is concerned. Similar approaches to training and testing are found in other, unrelated industries and many recent patterns / trends in aviation are mirrored elsewhere. In fact, some people even describe many prevalent aviation training and testing practices as archaic.
- c) What is the general trend? In aviation, ICAO is strongly advocating / encouraging a competency-based approach to training and testing. This model is related to the Outcomes-based education system that was implemented in the South African school system years ago. Many modern learning principles are integrated in such a system, which makes knowledge of these principles important for any trainer, assessor or evaluator. The SACAA does not currently use a competency-based model, but even for a content-based system such as ours, many of the principles remain valid and useful.
- d) Fundamentally, there are a few basics / principles that are important to know about. These basics are briefly listed below, each theme followed by a short explanation:
 - i. The use of a constructivist approach to learning,
 - ii. The difference between breadth and depth of learning,
 - iii. Procedural and declarative knowledge,
 - iv. The use of Bloom's taxonomy and the use of verbs corresponding to cognitive levels,
 - v. An emphasis on problem-solving and decision-making in training and testing.

i. The use of a constructivist approach to learning

Learning is explained in the following way (Nieman, M.M., Monyai, R.B., 2006: 72):

"In other words, learning is not seen as a product or result of the individual's genetic predisposition. Rather it entails change in the learner's insight, comprehension, behaviour, perception or motivation and this leads to added knowledge or the ability to do something that the learner could not do before."

Constructivism is an educational theory that emphasizes the primary importance of the student in a learning programme. The theory holds that a trainer or instructor cannot put knowledge inside the mind of a student, but that the student has to "*construct*" meaning in his or her own mind. Teaching is not a one-way process in which the instructor talks and the student listens, but an interactive process with learning as the foremost goal. It is the proverbial case of working smarter, not just harder.

The role of the instructor is to understand how learning happens and create conditions that allow the student to make sense of the material. What is important in constructivism is that the student should participate in the learning. He or she must take responsibility for his or her own learning and play an active part in the learning process. The student has a greater "*hands-on*" role than in a traditional training set-up. If the student does not understand what is taught, the training process is actually pointless. The idea of learner-centredness is important, because the learner is the person that should learn something.

Constructivism does not mean that training standards are lowered (in fact, they might now even be higher), but rather that the instructor has to do his or her job with learning in mind, not just by reciting information to the student. The student, conversely, has to take responsibility for his or her own learning and cannot sit back and passively receive information. This is about moving beyond rote-only learning to meaningful learning.

ii. The difference between breadth and depth of learning

How much should a student know about a topic? Is it better to understand a single topic very well, or know a little bit about a wide variety of topics?

The Alberta Education department states: (Parsons, J. and Beauchamp, L., 2012: 131):

"A person who is a mile wide and an inch deep is not an educated person.

But a person who is a mile deep and an inch wide is not an educated person either"

What is depth and breadth of learning? Moldoveanu and Martin (quoted in Parsons and Beauchamp, 2012: 135) explain that *"Breadth of mind"* is about how many facts and variables someone can comfortably take into account when thinking, but *"Depth of mind"* refers to how advanced and complex the thinking is that one can comfortably do.

Parsons and Beauchamp (2012: 131) explain it like this:

"Breadth and depth can also be explored as it applies to deep and surface learning.

What is deep learning? It is an approach to learning where learners routinely make use of higher-order cognitive skills, such as the ability to analyze, synthesize, problem solve and think metacognitively in an effort to construct long-term understanding. Unfamiliar ideas are analyzed critically in an effort to link them to known concepts and principles in order that new understandings can be generated and used to problem solve in unfamiliar contexts. Deep learners reflect on the personal significance of their learning, which in turn, promotes understanding and application for life.

What is surface learning? It is the tacit acceptance of information as isolated and unlinked facts, often leading to superficial retention of material. Surface learning, therefore, does not promote understanding or long-term retention of information or knowledge."

iii. Procedural and declarative knowledge

There are many different ways of looking at knowledge, understanding and learning. A particularly useful way is to classify knowledge as having declarative and procedural aspects.

Declarative knowledge can be explained as knowing what to do and why it should be done. This is often knowledge that can be *"declared"* or put into words.

Procedural knowledge is knowing how to do something and when to do it. This is often knowledge that can best be demonstrated with actions.

Training and learning involve both kinds of knowledge. Teaching a student how to do a proper steep turn, for example, requires declarative knowledge in order to understand the sequence of actions involved (e.g. engine set-up, look-out and manoeuvre) and the reasons for actions (you need to raise the aircraft nose or you'll lose altitude). Procedural knowledge entails knowing how to do the manoeuvre and when to do it (e.g. avoiding a bird or doing a pre-stall inspection turn).

When doing training or testing, both of these need to be kept in mind. Declarative knowledge often underpins procedural knowledge.

iv. The use of Bloom's taxonomy and the use of verbs corresponding to cognitive levels*

***Note** - The user of this Technical Guidance Material is encouraged to read more widely about this important topic. A discussion of the use of Bloom's taxonomy should likely best be started by recapping on what the concept of *"Competence"* or *"Competency"* means.

ICAO defines *"Competency"* as follows:

"A combination of skills, knowledge and attitudes required to perform a task to the prescribed standard" (International Civil Aviation Organization, 2011: I-2).

Dr. Melissa Fein explains the idea in a similar way:

"Knowledge, skills and abilities (KSA's) are the job-related qualifications and personal attributes necessary for an individual to perform effectively in a particular job or practice a particular profession" (Fein, M.: 2012: 73).

What does this mean? Very simply, the idea is that human learning can happen in one or more areas - physical, cognitive and psychological (the latter is sometimes referred to as the emotional or attitudinal aspect). Explained in another way, to learn something can have a thinking component (e.g. understanding something) or a physical skill (e.g. co-ordination) or a psychological part to it (e.g. having situational awareness). Logically, most tasks require a combination of these three aspects. When one possesses the right blend of thinking, physical and psychological capabilities to successfully do a job, one can be said to be competent.

A well-known and widely used classification of these abilities was developed in 1956. This was done in the USA by Dr. Benjamin Bloom, a specialist at the University of Chicago. Dr. Bloom's work was titled *"Taxonomy of educational objectives"* and the resulting classification system, quite logically, became known as *"Bloom's taxonomy"* (Booyse and Du Plessis, 2014: 84).

Bloom's taxonomy divides learning into the three areas (or learning domains) mentioned above. Only the cognitive (or thinking domain) is mentioned in this document, due to its relevance to the discussion. Although Bloom's taxonomy was refined slightly in 2001, the original version will be discussed here, mostly because it is also used in some of the other aviation references listed in the bibliography, such as the FAA Aviation Instructor's Handbook.

Bloom's taxonomy shows that thinking happens on six levels of complexity (Dreyer, J.M., 2014: 80). In the following table, these levels are arranged in sequence of simple to complex, from top to bottom (Booyse and Du Plessis, 2014: 84-85; Dreyer, 2014: 80-81; Fein, 2012: 73-74):

The six levels of cognition in Bloom's taxonomy	
Cognitive level	Ability required
Knowledge	Remembering facts
Comprehension	Grasping the meaning of information
Application	Using something that was learnt in a new, but practical scenario
Analysis	Understanding the content and structure of information (or deconstructing something that was learnt into its components)
Synthesis	Combining unrelated pieces of knowledge to generate a new piece of knowledge (or formulating new structures with existing knowledge)
Evaluation	Justifying a theory or position (or making a judgement on the validity of an argument)

The significance of the taxonomy is immense. It helps to make it easier to understand the different levels of thinking complexity. The taxonomy applies to aviation training and testing as much as to other industries.

The taxonomy also helps in the development of oral and written examination questions and scenario's. In a situation that is relevant to the situation of a DFE, for example, the taxonomy can act as a guideline to design an Line-oriented Evaluation (LOE) that requires enough from a candidate during a test. Solving a sudden in-flight emergency, for example, requires at least "Application"-level thinking. In contrast, simply identifying a VOR morse code identifier can most definitely not be seen as a complicated thinking exercise.

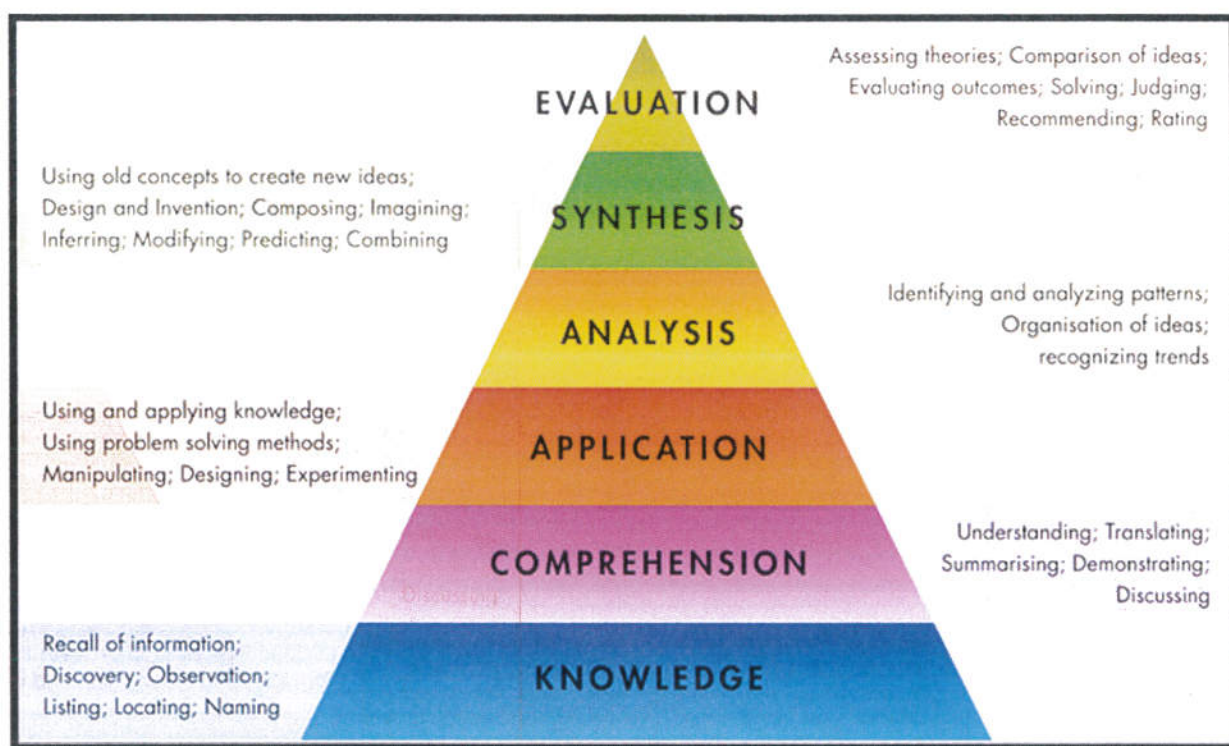


Image 1 - Bloom's taxonomy in pyramid form

Two things are important to keep in mind about cognition complexity, as explained in Bloom's taxonomy – The first is that the cognitive levels build on each other. If one cannot remember information, one is unlikely to be able to understand it or apply it. A useful way in which this is sometimes displayed is graphically, like in Image 1 (Pedro, 2017). In this pyramid image, the lower levels of thinking are shown on the bottom. They are depicted as forming the foundation for the other, higher levels of thinking.

The idea of a hierarchical structure of knowledge is also explained in the following quote (Fein, 2012: 73-74): *"These categories are presumed to be hierarchical in the sense that it is assumed that individuals are thought to need competence in lower-level skills such as knowledge and comprehension before higher-level skills can be mastered."*

The second matter to remember is that higher levels of complexity are not necessarily the same thing as level of difficulty. It is possible to have a "Recall" level question that is very difficult and an "Evaluation" level question that is very easy (Brookhart and Nitko, 2008, cited in Dreyer, 2014: 80). Remembering a long list of emergency memory recall items for a complicated aircraft emergency might be difficult, even though the cognitive complexity of the thinking required is quite low - in this example, the demand is made on one's ability to remember, not on how one mentally manipulates concepts and ideas or reaches an innovative conclusion. On the other hand, substantiating an opinion on why it is unsafe to drive under the influence of alcohol is a simple task, although it demands more complex thinking and the structuring of an argument by combining information and sequencing the different parts of an argument.

At this point, it is important to briefly link Bloom's taxonomy to the idea of higher-order thinking skills. In Image 1, the bottom half of the pyramid represents lower-order thinking and the top half higher-order thinking levels (Nieman and Monyai, 2006: 80-82). While the lower levels are obviously important to a pilot, on their own they are not enough for the complex functions of the pilot role, such as managing work load, maintaining situational awareness, handling emergencies, interpreting information and so forth. Obviously, the same principle applies to a test or check. For

example, merely determining during an Instrument Rating skills test whether a pilot can remember pieces of information does not tell anyone whether that pilot can actually apply what he or she learnt during training.

The following extracts from the FAA Aviation Instructor's Handbook explains the idea of higher-order thinking skills:

"Higher Order Thinking Skills (HOTS)

The constructivist theory of learning explains and supports the learning of HOTS, which is commonly called aeronautical decision-making (ADM) in aviation. HOTS lie in the last three categories on Bloom's Taxonomy of Learning: analysis, synthesis, and evaluation skills. Teaching the higher level thinking skills which are essential to judgment, decision-making, and critical thinking is important to aviation because a common thread in aviation accidents is the absence of higher order thinking skills ... HOTS are taught like other cognitive skills, from simple to complex and from concrete to abstract."

"It must be remembered that critical thinking skills should be taught in the context of subject matter. Learners progress from simple to complex; therefore, they need some information before they can think about a subject beyond rote learning. For example, knowing that compliance with the weight-and-balance limits of any aircraft is critical to flight safety will not help an aviation student interpret weight-and-balance charts unless he or she knows something about how center of gravity interacts with weight and balance."
(Federal Aviation Administration, Flight Standards Service, 2008: 2-5, 2-6)

A very important tool when using the taxonomy is to carefully select action verbs when phrasing questions or designing training or testing activities. These action words generally correlate with certain cognitive levels and explain clearly what the Examiner wishes to hear from a candidate.

Competence		Skills Demonstrated	Example
I	Knowledge: remembering information	Define, identify, label, state, list, match, select	1. State the standard temperature at sea level. 2. Define a logbook entry.
II	Comprehension: explaining the meaning of information	Describe, generalize, paraphrase, summarize, estimate, discuss	1. In one sentence explain why aviation uses a standard temperature. 2. Describe why a log entry is required by the FAA.
III	Application: using abstractions in concrete situation	Determine, chart, implement, prepare, solve, use, develop, explain, apply, relate, instruct, show, teaches	1. Using a standard lapse rate, determine what the temperature would be at a pressure altitude of 4000'. 2. Determine when a logbook entry is required.
IV	Analysis: breaking down a whole into component parts	Points out, differentiate distinguish, examine discriminate, compare, outline, prioritize, recognize, subdivide	1. Compare what the different temperatures would be at certain pressure altitudes based on the standard lapse rate. 2. Determine information required for logbook entry.
V	Synthesis: putting parts together to form a new and integrated whole	Create, design, plan, organize, generate, write, adapt, compare, formulate, devise, model, revise, incorporate	1. Generate a chart depicting temperatures for altitudes up to 12,000'. 2. Write a logbook entry for an oil change.
VI	Evaluation: making judgments about the merits of ideas, materials, or phenomena	Appraise, critique, judge, weigh, evaluate, select, compare and contrast, defend, interpret, support	1. Evaluate the importance of this information for a pilot. 2. Evaluate the necessity of keeping logbook entries.

Image 2 - The six major levels of Bloom's Taxonomy of the Cognitive Domain with types of behaviour and examples of objectives

The table above (Image 2) has additional information on Bloom's taxonomy with action verbs and gives some practical examples (Federal Aviation Administration, Flight Standards Service, 2008: 2-13)

Booyse and Du Plessis (2014: 89-93) explain how various actions can be required during demands can be made on candidates to relate to different cognitive levels in *Bloom's taxonomy. A few relevant examples are given in the *table below.

**Note - The 2001 revised taxonomy is used by the authors and reflected here.*

Level	Cognitive level	Demand	Action words
1.	Remembering	<ul style="list-style-type: none"> a. Mention a simple law or equation b. Retrieve information c. Explain terminology d. Recall complex content 	Recite / List / Define / Name / State / Recall
2.	Understanding	<ul style="list-style-type: none"> a. Explain simple relationships b. Identify content in a table c. Explain particular terminology d. Give 1-step answers e. Explain processes f. Identify principles 	Give example Explain Describe Summarize Discuss Identify
3.	Application	<ul style="list-style-type: none"> a. Perform well-known procedures in familiar contexts b. Know what procedure is required to solve a problem (all information given to candidate) c. Identify and use procedures d. Simplify procedures 	Demonstrate Interpret Decide and apply Choose an action
4.	Analysis	<ul style="list-style-type: none"> a. Solve non-routine, unseen problems b. Explain more complex relationships between concepts c. Solve problems with 2 or more steps d. Read and interpret scenarios e. Interpret and make adjustments f. Extrapolate from solutions already obtained to solve problems in unfamiliar contexts g. Give answers to "what if" questions 	Investigate Classify Compare Interpret
5.	Evaluation	<ul style="list-style-type: none"> a. Substantiate an opinion b. Critique solutions to problems c. Provide a solution and explain why 	Judge Evaluate Decide on the best route
6.	Synthesis	Use complex reasoning linked to abstract ideas Generalize patterns observed and make predictions Redesign Work with complex problems that involve insight and logic leaps Create new solutions to problems Modify procedures or methods	Design Create Construct Forecast Rearrange Modify Formulate Generate

v. **An emphasis on problem-solving and decision-making in training and testing**

The world is becoming more complex. Dealing with this effectively requires people to “*think on their feet*.” It is no longer enough to recite facts verbatim with no insight or ability to apply this information (Maybe it has actually never been enough). This awareness is starting to feature in education and training worldwide. Compare the following learning goals from different disciplines below and observe how similar some of the training objectives are.

A few critical outcomes of the South African High school English Home language curriculum:

Identify and solve problems and make decisions using critical and creative thinking

Work effectively as individuals and with others as members of a team

Communicate effectively using visual, symbolic and/or language skills in various modes

Demonstrate an understanding of the world as a set of related systems by recognising that problem solving contexts do not exist in isolation

Pilot core competency descriptions (ICAO Doc 9995)

Demonstrates effective oral, non-verbal and written communications, in normal and non-normal situations

Accurately identifies risks and resolves problems. Uses the appropriate decision-making processes

Perceives and comprehends all of the relevant information available and anticipates what could happen that may affect the operation.

Air traffic controller performance criteria (ICAO Doc 9868)

Identifies potentially hazardous situations (e.g. amount of separation with other aircraft, objects, airspace and ground, consequences of adverse weather, navigational deviations and capacity overload)

Selects communication mode that takes into account the requirements of the situation, including speed, accuracy and level of detail of the communication

Demonstrates active listening by asking relevant questions and providing feedback

The demand on experts in fields outside of aviation to apply higher-order thinking skills and solve new kinds of problems is present. In the same way, future flight crew members will increasingly be required to think on their feet - for example to analyse problems, create new solutions to non-routine problems and substantiate these convincingly.

9. THE STANDARD OF ENGLISH TO BE USED IN QUESTION DESIGN

- a) Official SACAA examination questions are written in English. Which standard of English is required though? This is a relevant question and the following is intended to provide some guidance.
- b) If the purpose of an SACAA aviation theoretical knowledge examination is fundamentally to test knowledge, then that is exactly what it should do - nothing more and nothing less.
- c) Do not test someone's command of English, nor the candidate's ability to guess, interpret intentions, reading speed or stress-handling ability. There are other tests available for this, if required.
- d) Notwithstanding this, it is reasonable and appropriate to expect that aviation professional should have a reasonable level of proficiency in English. This applies legally to operational personnel, specifically flight crew and Air traffic controllers. These licence-holders are legally required to have a command of English consistent with the minimum ICAO operational level, namely Level 4. The operational language proficiency levels are presented in rubric form in Image 3 below.

ICAO LANGUAGE PROFICIENCY RATING SCALE

1.1 Expert, extended and operational levels

LEVEL	PRONUNCIATION <i>Assumes a dialect and/or accent intelligible to the aeronautical community.</i>	STRUCTURE <i>Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.</i>	VOCABULARY	FLUENCY	COMPREHENSION	INTERACTIONS
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work-related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work-related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work-related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work-related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.

Levels 1, 2 and 3 are on subsequent page

Image 3 - ICAO Operational language proficiency rubric

- e) There might not necessarily be ICAO requirements for other licence-holders, but in some cases South African aviation law has prescriptions in excess of ICAO Annex 1. In the case of Cabin crew members, for example, SA-CAR 64 states the following:

"Language

SA-CAR 64.01.8 (1) Cabin crew members shall have sufficient ability in reading, speaking and understanding the English language to enable them to adequately carry out their responsibilities as cabin crew members."

- f) It is reasonable to assume that aviation professionals will have overall reading and language skills proportional to the level described in the table.

10. ORAL EXAMINATIONS

- a) ICAO Doc 9379 states the following in regards to oral examinations:
(International Civil Aviation Organization, 2012: II-5-2)
"Oral examinations require an experienced examiner with an in-depth knowledge of the specialist subject matter to conduct a one-on-one examination of a candidate, and sound processes, including statistical analyses, should be put in place to ensure consistency and objectivity. Most commonly, an oral theory examination is a component of a practical test"
- b) Essentially, an oral examination refers to the practice of conducting an assessment/evaluation verbally. Although this can be done informally, this TGM addresses formal examination only. Formal oral examination can be done either through a structured or an unstructured interview (Dreyer, 2014: 114-115).
- c) Dreyer explains the difference between these two types of interviews as follows (2014: 115):
- A structured interview is comparable to a written examination with a fixed set of questions, except that the candidate responds verbally;

- ii. An unstructured interview allows room for diverting the line of questioning and probing, follow-up questions. This type of interview is characterized by questions such as "Please explain that in more detail."
- d) Oral examinations are widely used in aviation, especially in conjunction with a practical assessment/evaluation.
- e) Nieman and Monyai (2006: 132) refer to several types of oral questions.

The following three are relevant to this document:

- Closed questions -** These questions are closed because they do not allow further discussion. They can only be answered with a short response, such as "Yes" or "No."
Examples of such a question are:
"Is the international emergency frequency 121.5 MHZ?" or
"What is the international emergency frequency?"
- Open questions -** These questions stimulate thought and allow excellent testing of knowledge. These questions often have several possible answers. They require candidate participation.
An example of such a question is:
"Explain to me how you would determine the most suitable take-off alternate aerodrome?"
- Follow-up questions -** These are penetrating types of questions and excellent for going deeper into a topic. They are used to follow-up on a candidate's response or get additional information after a previous question.
An example of such a question is:
"Can you explain to me how you came to that conclusion?"

- f) Guidelines for effective questioning are given below. Although these originate from educational literature, they remain useful in aviation oral examinations (Booyse, C. and Du Plessis, E., 2014: 88-89):
 - i. DE's should concentrate on giving "wait time" after asking a question. Research shows that a longer wait time causes responses to be longer and more in-depth, made with more confidence and include alternative explanations.
 - ii. DE's may change the formulation of a question to increase the level of demand, especially by moving from factual or basic conceptual questioning to comprehension and the application of thoughts
 - iii. Questions should require candidates to give thoughtful answers, not just agree to leading questions.
 - iv. Effective questioning is a skill and can be practiced. This might even be done with colleagues. With some thought and practice, a question can be made a powerful probing tool. Compare the following two questions and see how a different ending turns a feeble question into a powerful tool.

*"Some people describe friction as the opposite of slipperiness.
Do you disagree or disagree?"*
*"Some people describe friction as the opposite of slipperiness.
What do you think?"*

- g) Chan (2009) explains that the following are necessary to ensure good oral assessment:
 - i. Be sure that a student knows what the assessment objectives are.
 - ii. Prepare marking sheet or rubric beforehand to make grading consistent and fair.
 - iii. Document responses or progress to prevent a student contesting a decision later.
- h) It is important that an Examiner or a Designated Examiner phrases questions clearly and unambiguously. An oral examination must be delivered clearly and professionally. The following questioning habits or deficiencies are not considered acceptable for a Designated Examiner:

- i. Mumbling and incomprehensible speech,
 - ii. Poor grammar and sentence construction, slang and swearing,
 - iii. Poor technical language and a display of clear-cut ignorance in his or her field of work,
 - iv. Poor verbalization and evaluation techniques, including the inappropriate use of closed-ended, confusing, overly simple and ineffective questions,
 - v. Unsettling peculiarities and offensive behaviour and
 - vi. A disinterest and apathy about the assessment/evaluation.
- i) Oral examinations are powerful assessment/evaluation tools, having the following capabilities:
- i. they allow interactive, probing examination with follow-up questions, especially of suspected areas of deficiency,
 - ii. they allow candidates that are not proficient at expressing themselves the opportunity to explain their answers,
 - iii. examinees are able to ask for clarification,
 - iv. higher-order thinking can be powerfully examined,
 - v. they are very useful when elaborate answers are required and
 - vi. they allow a systematic exploration of a candidate's knowledge.
- j) Dreyer gives the following guidelines for oral assessment/evaluation (2014:115):
- i. The examiner should listen more and speak less,
 - ii. Use silence to give the candidate a chance to think and respond,
 - iii. Do not debate a topic with a candidate,
 - iv. Do not give your own point of view, the purpose of the examination is to collect evidence from the candidate,
 - v. Do not interrupt unnecessarily. Do this only when the examinee has misunderstood the question or has lost track or has become confused.
- k) Qualities of verbal questions are compared in the following table:

GOOD question technique	POOR questioning technique
Easily understood	Bewildering
<i>What is the difference between procedural and non-procedural air traffic control?</i>	<i>If an aircraft is entering your airspace, you know, when you're the controller on duty, and the radar is not working, what is that called?</i>
Composed of common words	Oversize
<i>What is a pilot's first priority in case of engine failure?</i>	<i>List all the steps you would take if you had an engine failure.</i>

Promotes thinking	Toss - up
<i>Why is it so important to maintain the ideal glide speed for an aircraft?</i>	<i>Is the glide speed for your aircraft important during a forced landing?</i>
Practical - operational	Irrelevant
<i>What should a NITS briefing contain?</i>	<i>What fee does the SACAA charge for your annual licence renewal?</i>
Applicable/appropriate	Leading
<i>Explain the legal requirements for the extension of an MPI validity period?</i>	<i>If a pilot asks you to extend the MPI validity period by falsifying the maintenance records, are you allowed to do this?</i>
Only one correct answer	Trick
<i>What is the normal climb speed for this aircraft?</i>	<i>What types of climb speeds are there for this aircraft?</i>

11. WRITTEN EXAMINATIONS

- a) In this TGM, the term “written examination” is used to refer to any theoretical knowledge assessment/evaluation that is answered by a physical handwritten response.
- b) This includes any form of question, including MCQ's and MRQ's, as long these are not associated with a CBT system, as explained in the next section.
- c) Traditionally, this type of examination is often associated with essay-type questions. However, a multitude of questions may be presented to a candidate, virtually identical to a CBT examination.
- d) Although written examinations are widely used in secondary and tertiary educational environments, the SACAA does not presently conduct any written examinations.
- e) In South African aviation, the practice of conducting written examinations is mostly prevalent at operators and during formative (in-training) assessment/evaluation.
- f) Practically, there is a strong case for written examinations, although much effort is required to conduct them adequately, for the following reasons:
 - i. Several papers must be available for circulation to avoid exposure of content,
 - ii. A marking memorandum must accompany every paper,
 - iii. Moderation of both the papers and memoranda is required,
 - iv. Papers must be stored securely and
 - v. Marking takes time and is likely to have some degree of subjectivity.

- g) An alternative to fixed papers is to have a set of questions available electronically and use these to compile a fresh paper for each session. However, this places an additional burden on the assessor/evaluator to monitor the frequency of use and success rate of each question.
- h) However, as explained earlier, the advantage of written examinations is that the cognitive level of testing that is feasible far exceeds that of CBT examinations.
- i) Other advantages include:
 - vi. reasonably flexible testing practices,
 - vii. lower cost and overheads than with CBT examinations,
 - viii. the hands-on involvement of the Examiner or Designated Examiner and
 - ix. For smaller-scale examinations, less effort than with CBT examinations.
- j) Certain prerequisites must be met in order to conduct written examinations:
 - i. Clear, hard copy instructions must be provided to each candidate,
 - ii. Criteria must be communicated to candidates, preferably by marking rubric,
 - iii. Mark allocation and time limits must be provided,
 - iv. If possible, an expected time to be spent on each question must be communicated and
 - v. Question layout and presentation should be logical and understandable (this means that numbering and bullets, spacing, font and type and page numbering and referencing should be given attention).

12. COMPUTERIZED FIXED TESTS

- a) The term "*Computerized Fixed test*" refers to a "*fixed-length, fixed-form computerized exam without any adaptive item selection*" (Parshall, C.G., Spray, J.A., Kalohm, J.C. and Davey, T., 2002: 92).
- b) A CFT is the CBT type examination most similar to a hard copy paper-based examination (Parshall et al, 2002: 92) and provides several of the basic, simple benefits of CBT testing. All ICAO Annex 1 licence examinations conducted by and under the control of the SACAA are presently of the CFT type with the randomization selected and set at maximum level.
- c) Adaptive functionality is found in a Computerized Adaptive Test (CAT). This means that the CBT system selects a highly individualized question set to each examinee, based on his or her response to previous questions in the same examination.
- d) CAT functionality is not to be confused with the randomization function of modern CBT systems. Such randomization functions select questions randomly, if activated. This functionality should always be active to minimize cheating and item exposure.
- e) For the purpose of this document, the term "*Computerized Fixed test*" includes
 - i. all electronic examinations towards the issue of a licence (including tablet- and laptop-delivered examinations) and
 - ii. all examinations compiled using the standard electronic examination system but delivered in another manner (like hard copy papers generated by special software functionality).
- f) In other words, CFT is used as a variety of CBT.

- g) In Doc 9379, the Manual of Procedures for Establishment and Management of a State's Personnel Licensing System, ICAO summarizes the context of CBT examinations in the following way (The International Civil Aviation Organization, 2012: II-5-2):

"Written examinations are provided in paper form by many Licensing Authorities although computer-based examinations are increasingly popular.

Computer-based examinations have the benefit of potentially providing unique but equivalent examinations for every candidate, drawing on an extensive database of questions (to combat cheating); they can deal with large numbers of candidates at once, who do not necessarily have to start and finish at the same time; they can be marked and results provided almost immediately; and they can potentially explore knowledge more effectively than paper-based examinations. On the other hand, computer-based examinations are expensive to set up, require built-in security systems to avoid compromising the questions, and require ready access to trained computer technicians to maintain hardware and software."

- h) Although it is possible to use a variety of question types in a CBT system, the vast majority of questions are objective types (MCQ's and MRQ's). For this reason, the following objective type-terminology is used (University of Pretoria, 2011: Terminology page):

Stem -	The wording of the question
Options -	The possible answers to an item
Distractor(s) -	The option(s) containing incorrect answers
Key -	The correct answer option.

- i) A large amount of information is available for users interested in understanding objective items better. Certain electronic sources listed in the references of this TGM provide a good starting point. The SACAA PEL department may be contacted if more sources are required.
- j) Both multiple-choice and multiple response questions may have graphs, diagrams, charts or other visual media attached. These will be designated as "Advanced" variations and extra marks and time will have to be allocated to them.
- k) Be careful of YNTF questions. The danger with YNTF is that an examinee has a 50% chance of guessing the answer correctly. Extra care needs to be taken with YNTF to avoid this. However, YNTF can be valuable, because they have the potential to be powerful tools to indicate what a candidate (or examinees in general) does not know.
- l) There are 2 slightly different varieties of Multiple-response questions – one requires that all the correct choices have to be selected to get the question right and the other allows marks to be accumulated cumulatively. The cumulative type carries risk if not designed correctly, namely that an examinee could simply select all the options and get full marks for the question. Multiple-response questions (MRQ's) are powerful testing tools, but should be carefully designed. When MRQ's are dichotomously scored, a combined correct response should be awarded the same marks as a standard MCQ.
- m) Multiple choice questions may have anything from 3 to 5 choices. 4 choices is normal and should be the default number of choices.
- n) In a multiple-choice question, only one option should be completely correct, although it is fine to have this choice correct because it is more appropriate than another. The exam instructions do say that examinees should select the most correct option.
- o) Advanced multiple-choice and multiple-response questions have charts, graphs, diagrams, pictures and such attached. Do not use videos or sound clips.

- p) To minimize the risk of premature exposure of repeated test items, a minimum database size is required. As a minimum, the developer should develop 3 questions per topic for each examination question presented from that topic (called the coverage here). However, experience has shown that a safer minimum for high-stakes examinations is a coverage factor of **10**.
- q) When developing large-scale examination items, it is good advice for the developer to keep a note pad close for when those out-of-the office times when he or she suddenly thinks of a nice question!

13. *DESIGNING AND MODERATING TEST ITEMS

***Note 1-** *The procedures described in this section apply to SACAA Examiners and all categories of Designated Examiners.*

It is assumed that the following conditions are met:

- i. *relevant regulatory prescriptions are in place and have been complied with (including specific ATO and operator training and procedural approvals),*
- ii. *a syllabus or training programme specifying the subject content has been developed and published,*
- iii. *for large-scale examination development, a coordinator has been appointed and each member of the development team has been allocated a specific subject area for which he or she has met the required Subject Matter Expert (SME) criteria,*
- iv. *the question developer has been both mandated and trained to develop examination questions,*
- v. *all Non-disclosure agreement legalities have been complied with,*
- vi. *the technical guidance provided in this TGM document has been incorporated into the design of items and*
- vii. *required guidance material has been developed for the use of the trainer/trainee.*

a) Before starting

- i. The progressive phases of the test item design process are given sequentially below.
- ii. Procedures applicable to all three categories of examinations are combined without specifying the examination type. In a case where a different method or procedure applies to each specific type of examination, this procedure is specified in the appropriate column. Where an activity is not applicable to a specific type of examination, this is indicated by a black-shaded field.
- iii. The regulations governing the licensing activity must be thoroughly reviewed and complied with. This includes that component of SACAA procedures documented in ISO procedures.
- iv. In most cases, regulatory prescriptions pre-determine the assessment/evaluation method to be used. A practical demonstration of operational competence by a Cabin Crew member, for example, may not be done by CFT. The purpose of the assessment/evaluation often determines the allowable method(s). Users of this TGM should refer to the relevant Part and Sub-parts of the Regulations.
- v. Although there is more regulatory flexibility with the design of a theoretical examination, logistical considerations often prevail. The only practical way for the SACAA to host the tens of thousands of flight crew theoretical knowledge examinations which it conducts annually, for example, is by CFT.
- vi. Despite some restrictions, there is still often significant testing leeway available to the developer. This should be used to decide on the most effective and meaningful testing method.
- vii. Subject matter not specified in the syllabus may be not examined.
- viii. It is recommended that items not be written during meetings or agreed on purely by consensus, unless mandated by special circumstances. ICAO gives the following guidance in this regard (International Civil Aviation Organization, 2012: II-5-3):
"Sometimes, questions are written by a committee but care should be taken with this technique to avoid "group-thinking", where a group seeking consensus tends to override realistic appraisal of alternative courses of action (e.g. introducing questions about new technology or new procedures)."

b) ***General procedure for the design and moderation of test items**

**Note - Only once the basic pre-design requirements listed above have been met, should development proceed.*

Phase 1 - Course overview

Oral exam	Written exam	CBT exam
Review the training programme objectives. The aim of the training course should be understood before starting with the design process. This should inform everything else that follows. Questions to be asked may include: <ul style="list-style-type: none"> – “Where does this exam fit into the training course?” – “What training preceded this examination?” – “What is the purpose of this examination?” – “What is the best available method to test candidates’ knowledge of this subject area?” 		

Phase 2 - Syllabus review

Oral exam	Written exam	CBT exam
Study the applicable syllabus topic requirements in the official syllabus. Include the syllabus guidelines, this TGM and supporting curricular documents, such as Aeronautical Information Circulars.		

Phase 3 - Review subject matter and topics in the course material

Oral exam	Written exam	CBT exam
Locate the content relevant to the topic in the suggested study material. The training course goals or outcomes, learning objectives or assessment criteria and study material should be integrated and coherent. The developer should confirm which study material is suggested or prescribed for the training course and confirm that industry or operator guidance material corresponds to the suggestion or prescription. The developer should be familiar with the study material. If necessary, the developer should recap on the subject content. This should be done prior to starting with development.		

Phase 4 - Refer to the style guide

Oral exam	Written exam	CBT exam
If one is available, the style guide should be consulted before starting. It is important to understand any question layout and typesetting requirements. In a modern environment, especially with CBT examinations, standardized question design, phraseology and emphasis techniques are necessary to avoid confusing candidates. Any specifications must be adhered to from the very beginning. Once an item is stylistically incorrect, it becomes very inefficient (or even impossible) to fix.		

Phase 5 - Consider fundamental examination sheet instructions

Oral exam	Written exam	CBT exam
<p>Basic instructions must preferably be determined at the beginning of the process.</p> <p>This is because examination instructions are often interwoven with responses. Certain types of questions must also be responded to in a specific way. Some examples of such scenarios are:</p> <ul style="list-style-type: none"> • Certain types of examinations may require a scenario to be studied and responded to in increments, until the candidate's knowledge has been exhaustively assessed/evaluated. In such a case, this must be clarified to the candidate. • Multiple-response type questions are marked differently, depending on whether they are scored dichotomously or cumulatively. This is an extremely important consideration and affects both the question development and the instructions. <p>Such requirements must at least be thought about intelligently prior to starting with the design process. This is especially important if the developer is not the person that will personally be conducting the examination.</p> <p>It makes for simpler and efficient item development if the process is started systematically and methodically. Compiling an instruction list afterwards increases work load exponentially and increases the chances for errors and omissions.</p>		
For oral examinations, the developer must make notes of any specific requirements that are crucial to the way in which a candidate responds. This relates closely to the marking guidelines or memorandum. For example, a candidate may be required to respond in detail to a limited number of questions. On the opposite extreme there may be a requirement to cover much subject matter in which case shorter answers may be preferred.	For written and CBT examinations, the developer must make notes of any specific requirements that are crucial to the way in which a candidate responds. In most cases, no clarifying questions are allowed and the importance of instructions becomes greater than with oral examinations.	
	In written examinations, action verbs are very important. An instruction like <i>"List"</i> means something very different from one like <i>"Justify."</i> A candidate must be sensitized to this in the instructions.	In CBT forced-response examinations, instructions assume a strong efficiency role. There are many different CBT applications and the individual characteristics of each must not confuse first-time examinees. The potential for error is increased in a forced-response environment and this places a responsibility in the developer to ensure consistency between questions, responses required and instructions.

Phase 5 - Review learning objectives and/or assessment criteria

Oral exam	Written exam	CBT exam
<p>Determine which cognitive level of knowledge is required by studying the learning objectives and assessment criteria of the programme.</p> <p>Proper assessment objectives state precisely what an examinee should be able to demonstrate and therefore what an item should contain. In the absence of published objectives and criteria, the developer has no option but to try and reconcile the overall goals of the training programme and subject syllabus with a justifiable, reasonable and broad (i.e. less-detailed) perspective of the required breadth and depth of knowledge. This is clearly not ideal, but is (unfortunately) very likely the way in which most training and testing is done in many industries and countries.</p>		

Phase 6 - Write the test item(s)		
Oral exam	Written exam	CBT exam
<p>Keep the type and purpose of the item in mind when selecting the language style and register.</p> <p>If necessary, the relevant section of this TGM should be consulted to better understand the characteristics and strengths of each type of question.</p> <p>The drafting process may be unique to a specific area of expertise, ATO or operator. For example, some organizations may require CBT questions to be captured in Word first, before being transferred to a software application. In other cases, it might be considered more efficient to capture CBT questions directly on the computer system. This part of the procedure falls outside of the scope of this TGM and is not prescribed in this document. Developers should refer to their respective organizational procedures for this information.</p> <p>Each item must be provided with</p> <ol style="list-style-type: none"> a sample solution or correct response, the official objective or criterion, the study material reference and the mark allocated to each correct or partially correct response. 		
Oral items should also be written down, despite the fact that they will not be presented on paper.		<p>CBT examinations most often use MCQ's and variations in bulk (large-scale) examinations. Quantity is often critical and this means that planning for enough time is important. A reasonable estimate for the compilation of questions is:</p> <p>Allow 10 minutes for each simple question plus 25% for moderation (i.e. 12 – 13 minutes per question).</p> <p>To be safe, plan on 3 to 4 questions per hour. Be very disciplined with your time.</p>

Phase 7 - Initial (1 st) moderation		
Oral exam	Written exam	CBT exam
<p>The initial review is the first and basic appraisal of the item.</p> <p>The term review is used synonymous with the term moderation. The initial review is analogous to the first proofreading of a manuscript and may be done by the developer or another SME.</p> <p>If the developer does the review, he or she should try and be as objective as possible. For this reason, it is not realistic to try and do this immediately after writing the item. Allow some time to pass, even if other tasks have to be prioritized. It helps to develop several questions for the same examination topic reviews them in bulk to be more efficient.</p>		
A DE that is not associated with an ATO full-time is not required to employ additional staff members to satisfy the moderation requirements specified in this document. This applies to a DE that does oral examination of a candidate as part of a practical assessment/ evaluation. However, such a DFE is encouraged to contact the PEL department for moderation input.		

Phase 8 - Technical (2 nd) review		
Oral exam	Written exam	CBT exam
<p>The technical review is the second level of moderation.</p> <p>This review <i>must</i> be done by a second SME.</p> <p>The purpose of moderation is to identify errors (often well hidden), assumptions and latent ambiguities. It should also balance extremes in an item. The reason for the latter review is that an experienced item developer should be able to remain objective, disciplined and clinical when crafting a question, but it is simply human nature to have an inherent bias towards or against certain topics. This can occur because of personal operational experience or simply preference. This easily leads to an over-emphasis of interesting topics or neglect of important ones.</p> <p>Technical moderation should focus on the technical correctness of the question and the terminology. This review should also check for fairness and consistency of the question with the syllabus and objectives/criteria, as applicable. In the absence of the latter, the technical review plays an even more important role, namely that of moderating the breadth and depth of knowledge assessment/evaluation.</p> <p>The technical moderator should view the item with proverbial distrust and not assume anything.</p>		
	For written examinations, the technical review also involves checking that the solution presented in the memorandum is correct and justifiable.	

Phase 9 - Language (3 rd) review		
Oral exam	Written exam	CBT exam
<p>Language review is the third level of moderation.</p> <p>Ideally, all three reviews should ideally be done by different developers, although the language moderation may be done by an experienced administrative staff member. This review may, however, be done by the initial developer.</p> <p>The language review checks unnecessary or overly technical terminology, grammar and sentence construction and overall spelling and readability. This review should also specifically check for the compatibility between the stimulus and options.</p> <p>During this level of moderation, the reviewer also checks for adherence to the style guide.</p>		
	For a written examination, the 3 rd moderation level may not be done electronically, but a hard copy draft should be reviewed.	

14. *ANALYTICS (ITEM ANALYSIS)

**Note - The discussion of analytics only refers to CFT items. No data analysis of oral and written examination responses is done.*

- a) For our purposes, analytics refers to the data analysis of candidate performance in individual examination items. This serves several useful purposes, including
 - i. ensuring exam database quality,
 - ii. timely correction of incorrect or inappropriate items and
 - iii. training of exam item developers.
- b) A multitude of analytic parameters are often available as part of computer systems. These may include, for example:
 - i. Item difficulty (p -value analysis),
 - ii. Variations of Item discrimination (or Item-Total correlation),
 - iii. Distractor analysis,
 - iv. Coefficient- α (alpha) or Cronbach's Alpha reliability coefficient and
 - v. Gap analysis.
- c) For examinations conducted by the SACAA, only the Item difficulty and Item discrimination (Item-total correlation) analyses are currently practical.
- d) The following item performance parameters are regarded as warranting a review of the affected test item(s), as indicated in the paragraph above:
 - i. A p -value of greater than 0.7 or lower than 0.2
 - ii. An Item-total correlation value lower than +0.2
- e) An exceedance of the borderline values, as specified above, does not necessarily indicate a dysfunctional item, but justifies a closer inspection of the item(s).
- f) An important consideration for the statistical analysis of CBT examination items is the number of times that the item has been attempted. A simple, extreme example of the inaccuracy of statistical analysis is that if an item has only been attempted once, it will show either a 100% success rate or 100% failure rate. Clearly, an item needs to have been presented in examinations a number of times for any statistics to be usable. For the large-scale examinations conducted by the SACAA, experience has shown that a figure of **50 attempts per item** is the minimum for a meaningful analysis to be possible.
- g) A useful example of what item analysis can typically reveal is given in an online article by Questionmark, the examination software provider, partially quoted below (Peterson, 2011):

"The Item Analysis report shows you how a question itself (or even the class or course materials) is performing.

 - i. *If pretty much everyone is answering correctly, the question may be too easy.*
 - ii. *If almost no one is getting the question correct, it may be too hard or poorly written.*
 - iii. *If the majority of participants are selecting the same wrong answer:*
 - The question may be poorly written.*
 - The wrong answer may be flagged as correct.*
 - There may be a problem with what is being taught in the class or course materials.*
 - iv. *If the answer selection is evenly distributed among the choices:*
 - The question may be poorly written and confusing.*
 - The topic may not be covered well in the class or course materials.*
 - If the participants who answer the question correctly also tend to pass the assessment, it shows that the question is a "good" question – it's testing what it's supposed to test."*

15. *PROCEDURE FOR THE REVISION OF ITEMS

**Note - The procedure for the revision of a test item is dependent on the type of examination which it is applied. The procedure for each type of examination is described independently below.*

- a) Generally, the following principles apply to the routine revision of examination questions, in other words revision for the purpose of content maintenance. This does not refer to revision initiated by investigation:
- i. A test item should preferably not be presented to the same candidate more than once every 3 months;
 - ii. CBT questions be analytically monitored every year and a report drawn for items exceeding the values listed in the section on Analytics (Item analysis);
 - iii. CBT items with negative Item-total correlation values should immediately be temporarily removed from the active database, reviewed and corrected. This is the most critical of the analytical exceedances.
 - iv. CBT items with low item-total correlation values (lower than 0.2) should be reviewed;
 - v. CBT items with high p -value exceedances (greater than 0.8) should be withdrawn;
 - vi. CBT items with high p -value exceedances (higher than 0.7) should be reviewed;
 - vii. CBT items with low p -value exceedances (lower than 0.2) should be reviewed;
 - viii. CBT items with high Item-total correlation and lower p -values (from 0.2 to 0.7) do not warrant revision.

b) Revision of oral examination item(s)

- i. Problems with oral examination questions are unlikely because of the interactive and adaptive nature of the assessment/evaluation method. It is more likely that a deficient interview style is the culprit. It is also possible that an uncorrected deficiency in a DE's knowledge is a source of error.
- ii. Oral examinations are most often used in one-on-one assessment/evaluation situations. This makes the identification of interview, knowledge or content deficiencies difficult.
- iii. It is recommended that a DE deliberately submits to constructive input from another DE or SME, particularly if this can be done during a real assessment/evaluation situation.
- iv. A second counter-measure is for a DE pays close attention to Examiner (AO) feedback regarding the oral examination aspect during the post-oversight debriefing. Tendencies and poor habits may be revealed here.
- v. However, thorough initial moderation (quality control) of oral examination questions prior to use will prevent or minimize problems later.
- vi. To identify possible problem items, a DE should make notes of candidate performance on each item. This records candidate response patterns and may reveal potential problem items. Even if this does not reveal a problem, it nevertheless has the potential to indicate training-related problems. This should be communicated to the PEL department technical inspectorate.
- vii. If a DE identifies or suspects a problem item, it should not be used until fixed or replaced.
- viii. A DE not employed by an operator or ATO should liaise with the PEL department for technical assistance.

c) Revision of written examination item(s)

- i. The SACAA does not conduct any written examinations for the issue of Annex 1 licences.
- ii. For all practical purposes, written aviation examinations are therefore the exclusive domain of ATO's and operators.
- iii. The CFI (in case of an ATO) or Head of Training (in case of an operator) should monitor candidate performance by regular sampling of responses. Sampling 30% of attempted examinations is regarded as sufficient to identify recurring problem areas.
- iv. Problems may be indicated by one or more of the following:
 - v. A pattern of incorrect response(s), even by above average candidates,
 - vi. A recurring pattern of candidates not answering a specific question,
 - vii. Several attempts to respond to a question, indicated by scratching and corrections,
 - viii. A recurring pattern of candidates not completing a specific paper in time or
 - ix. Recurring requests for clarification of specific items.
- x. In case a potential problem with a fixed paper is identified, it should either be withdrawn from use for completed frequency ap formative Problems with oral examination questions are unlikely because of the interactive and adaptive nature of the assessment/evaluation method. It is more likely that a deficient interview style is the culprit. It is also possible that an uncorrected deficiency in a DE's knowledge is a source of error.
- xi. The paper or item should be reviewed by at least two in-house SME's in conjunction with the marking memorandum.
- xii. The PEL department inspectorate may be consulted for guidance.
- xiii. Where the revision of a paper or item necessitates a revision to an approval, the amendment should be submitted to the POI.

d) *Revision of CBT examination item(s)

***Note -** The procedures listed here do not address examination problems resulting from hardware, software, database configuration /or programming malfunctions or errors. Content revision under contractual arrangement with external service providers is also excluded from the procedures listed below.

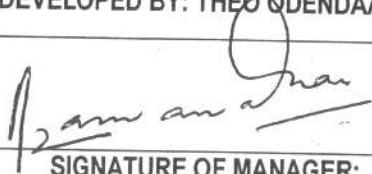
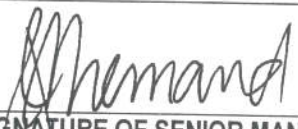

- i. CBT examinations comprise the bulk of the theoretical knowledge assessment/evaluation conducted by the SACAA. These can rightfully be called large-scale.
- ii. In consequence, errors are likely to affect larger numbers of candidates than with other examinations.
- iii. It is therefore reiterated that preventative measures should be in place to minimize the impact of database content errors.
- iv. PEL departmental guidelines give administrative procedures for the remark process and should be consulted.
- v. Suspected errors in a CBT database are identified by five potential mechanisms:
 - vi. The periodic revision or updating of database sections,
 - vii. Reporting of suspected errors by examinees or interested parties,
 - viii. The monitoring of item performance by system analytics capability,
 - ix. Incidental identification in consequence of remark applications and
 - x. Coincidental, non-deliberate detection in consequence of unconnected database-related work.
- xi. A party identifying or reporting a suspected database error should report this to the Manager: Examinations, the SME or inspector responsible for database content or the SM: PEL.
- xii. Once a potential error has been identified, it is allocated to an SME by the Manager: Examinations for investigation.
- xiii. If investigations reveals a legitimate problem, this is discussed with a second SME and reasonable consensus reached. If consensus is not possible, a second SME is involved.
- xiv. If consensus is still not reached, the Manager: Examinations is advised and a decision on retaining or scrapping the item is made.
- xv. Minor styling or grammatical errors are not subject to the steps mentioned above.
- xvi. Any amendments to an item are logged on the examination software.
- xvii. It is crucial to test the item for correct functioning, display and scanning prior to release.
- xviii. Prior to re-release of an item after modification, a second reviewer (SME) is required to inspect the item. The names of the first and second reviewers are also logged on the system.

16. BIBLIOGRAPHY

- a) Ambrose, S.A., Bridges, M.W., Di Pietro, M., Lovett, M.C. and Norman, M.K. 2010. *How Learning Works. Seven Research-based Principles for Smart Teaching*. San Francisco: John Wiley & Sons, Inc.
- b) Booyse, C. and Du Plessis, E. 2014. *Curriculum Studies. Development, interpretation and practice*. (2nd Ed.). Pretoria: Van Schaik Publishers.
- c) Carl, A.E. 2012. *Teacher Empowerment through Curriculum Development: Theory into Practice*. (4th ed.). Cape Town: Juta & Company Ltd.
- d) Chan C. 2009. *Assessment: Oral Assessment*. [Online]. University of Hong Kong. [Accessed on 2019-01-30]. (Available at https://ar.cetl.hku.hk/am_orals.htm).
- e) Department of Basic Education. 2011. Curriculum and Assessment Policy Statement. Grades 10-12. English Home Language. [Online]. (Available at https://www.education.gov.za/Portals/0/CD/National%20Curriculum%20Statements%20and%20Vocational/CAPS%20FET%20_%20HOME%20_%20ENGLISH%20GR%2010-12%20_%20WEB_5478.pdf?ver=2015-01-27-155402-907). [Accessed on 2019-02-26]. Pretoria: Department of Basic Education.
- f) Department for Education Innovation. *Writing MCQs and other objective items*. 2011. (Version 1.1). [CD-ROM]. Pretoria: University of Pretoria.
- g) Dickinson, M. 2011. *Writing Multiple-Choice Questions for Higher-level Thinking*. [Online]. Learning Solutions. [Accessed on 2019-03-07]. (Available at http://www.learningsolutionsmag.com/articles/804/writing-multiple-choice-questions-for-higher-level-thinking?utm_campaign=lsmag&utm_medium=email&utm_source=lsm-news).
- h) Dreyer, J.M. (Ed.). 2014. *The Educator as Assessor*. (2nd Ed.). Pretoria: Van Schaik Publishers.
- i) Du Toit, E.R., Louw, L.P. and Jacobs, L. (Eds.). 2016. *Help, I'm a student teacher*. (2nd ed.). Pretoria: Van Schaik Publishers.
- j) Federal Aviation Administration, Flight Standards Service. 2008. *Aviation Instructor's Handbook FAA-H-8083-9A*. (Available at http://www.faa.gov/about/office_org/headquarters_offices/avs/offices/afs/afs600. 2015-02-20). Washington, DC: U.S. Department of Transportation.
- k) Fein, M. 2012. *Test Development. Fundamentals for Certification and Evaluation*. Alexandria: The American Society for Training and Development.
- l) Fossey, A. 2015. *Questionmark White Paper. Managing Item Development for Large-Scale Assessment* [Online]. Questionmark Corporation. [Accessed on 2019-01-24]. (Available at http://pages.questionmark.com/rs/312-PUW-641/images/WP_ENUS_Managing_Item_Development_for_Large_Scale_Assessment.pdf).
- m) International Civil Aviation Organization. 2006. *Doc 9734. Safety Oversight Manual. Part A. The Establishment and Management of a State's Safety Oversight System*. 2nd Ed. Montréal: International Civil Aviation Organization.
- n) International Civil Aviation Organization. 2011. *Annex 1 to the Convention on International Civil Aviation, Personnel Licensing*. (11th Ed.). Montréal: International Civil Aviation Organization.
- o) International Civil Aviation Organization. 2012. *Doc 9379. Manual of Procedures for Establishment and Management of a State's Personnel Licensing System*. 2nd Ed. Montréal: International Civil Aviation Organization.
- p) International Civil Aviation Organization. 2013. *Doc 9995, Manual of Evidence-based Training*. International Civil Aviation Organization: Montréal.
- q) International Civil Aviation Organization. 2016. *Doc 9868, Procedures for Air Navigation Services - Training*. 2nd Ed. Montréal: International Civil Aviation Organization.
- r) Lambert, N.M. and McCombs, B.L. (Eds.). 1998. *How Students Learn. Reforming Schools Through Learner-Centered Education*. Washington: American Psychological Association.
- s) Lyman, H.B. 1986. *Test Scores and What They mean*. (4th Ed.). Englewood Cliffs: Prentice-Hall.
- t) Mager, R.F. 1984. *Measuring Instructional Results*. 2nd Ed. California: Pitman Learning, Inc.
- u) Morrison, G.R., Ross, S.M. and Kemp, J.E. 2004. *Designing Effective Instruction*. 4th Ed. Hoboken, New Jersey: John Wiley & Sons, inc.
- v) Nieman, M.M. and Monyai, R.B. 2006. *The educator as mediator of learning*. Pretoria: Van Schaik Publishers.
- w) Parshall, C.G., Spray, J.A., Kalohn, J.C. and Davey, T. 2002. *Practical Considerations in Computer-Based Testing*. New York: Springer-Verlag.
- x) Parsons, J. and Beauchamp, L. 2012. *From Knowledge to Action. Shaping the Future of Curriculum Development in Alberta*. [Online]. [Accessed 2019-02-26] (Available at <https://open.alberta.ca/dataset/bc0bd7df-2bfe-4b8b-8af0->

- db19b17a7721/ resource/5f11d83e-3074-408b-bff7-bcc27987864a/download/5976960-2012-from-knowledge-action-curriculum-development-alberta.pdf. Edmonton: Minister of Education.
- y) Pedro. 2017. [Blog] A longer piece on the taxonomy of Bloom. (Available at: <https://theeconomyofmeaning.com/2017/08/24/a-longer-piece-on-the-taxonomy-of-bloom/>). [Accessed on 2019-02-26].
- z) Peterson, D. 2011. *Improving Multiple-Choice Questions*. blog.questionmark.com. December 19. [Accessed on 2019-03-07]. (Available at <https://blog.questionmark.com/improving-multiple-choice-questions>).
- aa) Peterson, D. 2013. *Questionmark White Paper. Five Steps to Better Tests* [Online]. Questionmark Corporation. [Accessed on 2019-01-24]. (Available at http://pages.questionmark.com/rs/312-PUW-641/images/WP_ENUS_5_Steps_to_Better_Tests_White_Paper.pdf).
- bb) Pope, G. 2009. *Questionmark White Paper. Item Analysis Analytics* [Online]. Questionmark Corporation. [Accessed on 2019-01-24]. (Available at http://pages.questionmark.com/rs/312-PUW-641/images/WP_ENUS_Item_analysis_analytics.pdf).
- cc) Stobart, G. 2008. *Testing Times. The uses and abuses of assessment*. New York: Routledge.

17. DOCUMENT AUTHORISATION

DEVELOPED BY: THEO ODENDAAL		
	RAMA IYER	25 MARCH 2019
SIGNATURE OF MANAGER: EXAMINATIONS	NAME IN BLOCK LETTERS	DATE
REVIEWED & VALIDATED BY:		
	JOHAN NIEMAND	25 MARCH 2019
SIGNATURE OF SENIOR MANAGER: PEL	NAME IN BLOCK LETTERS	DATE
APPROVED BY:		
	SIMON SEGWABE	25 MARCH 2019
SIGNATURE OF EXECUTIVE: ASO	NAME IN BLOCK LETTERS	DATE