



**ICAEEA**

International Civil Aviation English Association

# *Selecting ICAO LPR testing systems: A look at the test instrument*

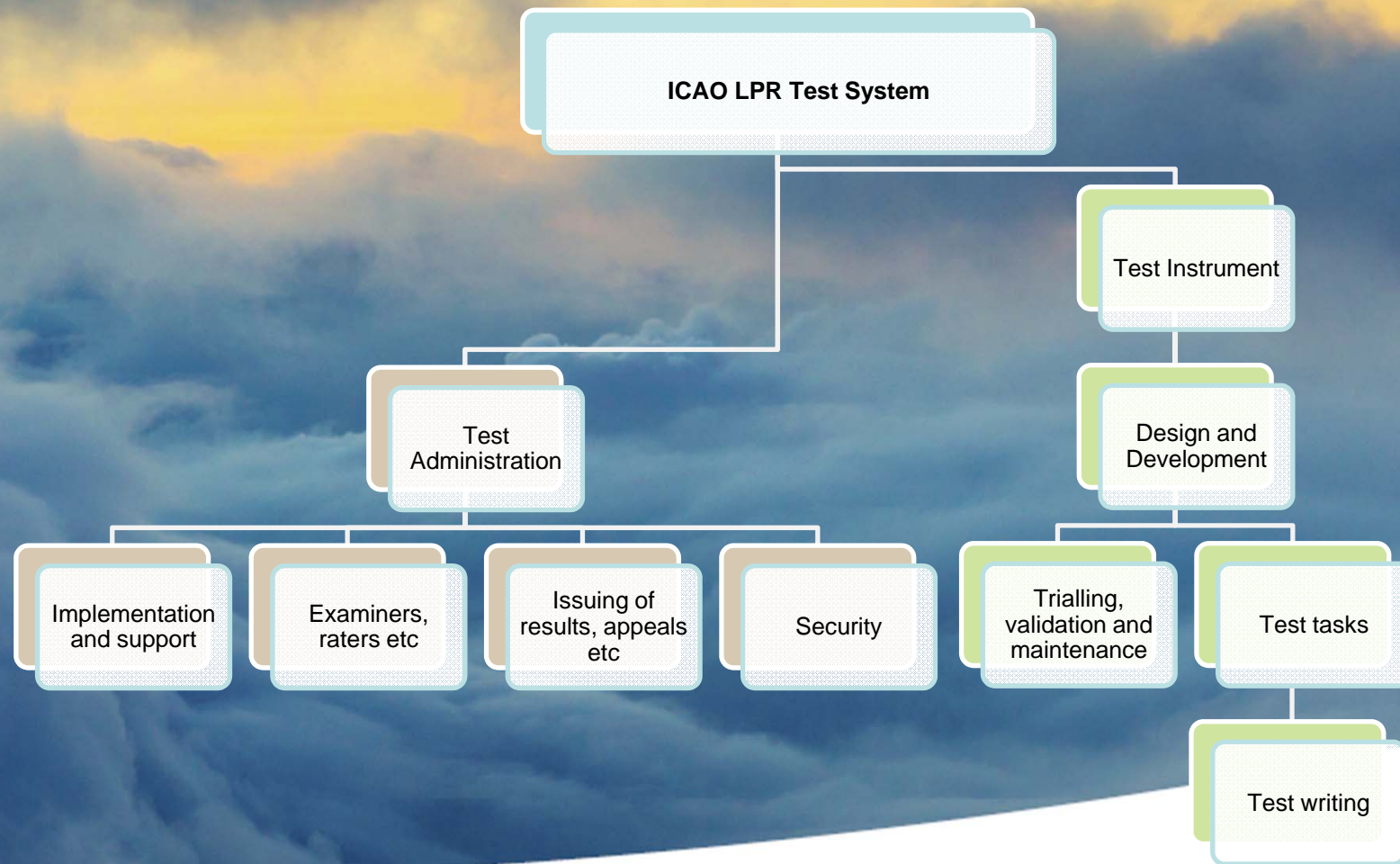


ICAO Inter-regional English Language  
Proficiency Workshop  
(Kuwait, 9-11 November 2015)

INTERNATIONAL  
CIVIL AVIATION ENGLISH  
ASSOCIATION



- Essentials of effective LPR tests
- Overview of valid testing practices
- Best practice in test instrument design
  - Evaluate a range of test task types
- Guide to recognising effective test instruments





# Key criteria: The test instrument

## Validity

- Authenticity
- Reliability
- Maintenance



# LPR testing quiz

1. What would be the overall recommended minimum speaking time required for effective evaluation in a test?
  - a) 10 minutes
  - b) 20 minutes
  - c) 30 minutes
  - d) 40 minutes

2. The LPRs require comprehension to be assessed.  
Which of the following best summarises the ideal approach?
- a) Candidates listen to recordings then describe what they heard
  - b) Candidates listen to recordings then answer prescribed questions
  - c) Candidates do a listening test, which allows them to sit a speaking test if they achieve a minimum level
  - d) Candidates do a speaking test where comprehension can be assessed indirectly



3. If tests include discrete items which are scored as right or wrong, they should contain at least

- a) 10 items
- b) 20 items
- c) 30 items
- d) 40 items



4. The LPRs are applicable to both licensed pilots and ATCOs. Which of the following best reflects how testing should be done for both groups?

- a) As the language and content is similar, one test is sufficient for both groups
- b) One test is sufficient for both groups, but different raters should be used, reflecting the different language needs of each group
- c) Both groups should do the same listening test, but different forms of a speaking test
- d) There should be two different test forms: one for pilots and another for ATCOs

5. Authenticity in language testing is mostly concerned with

- a) the type of language used in real-world situations
- b) the complexity and difficulty level of a range of task types
- c) the type of language and how it is used in real-world situations
- d) simulating on-the-job real-world stresses that affect language usage

6. The overarching concept of reliability in language testing relates to

- a) the precision of the test results
- b) how dependably raters evaluate
- c) the extent to which raters agree with each other
- d) the stability of the test and its consistency in measurement



7. Which of the following affects reliability in language testing?

- a) The consistency in how the raters evaluate
- b) The way the test is designed and constructed
- c) The equivalence of test tasks within a test bank
- d) All of the above

## 8. Validity in LPR language testing is mostly about

- a) fairness of the results
- b) the methods used for determining results
- c) a test's effectiveness in ranking candidates' abilities
- d) how closely test tasks allow candidates to communicate in ways which mirror real communication contexts so that results are meaningful

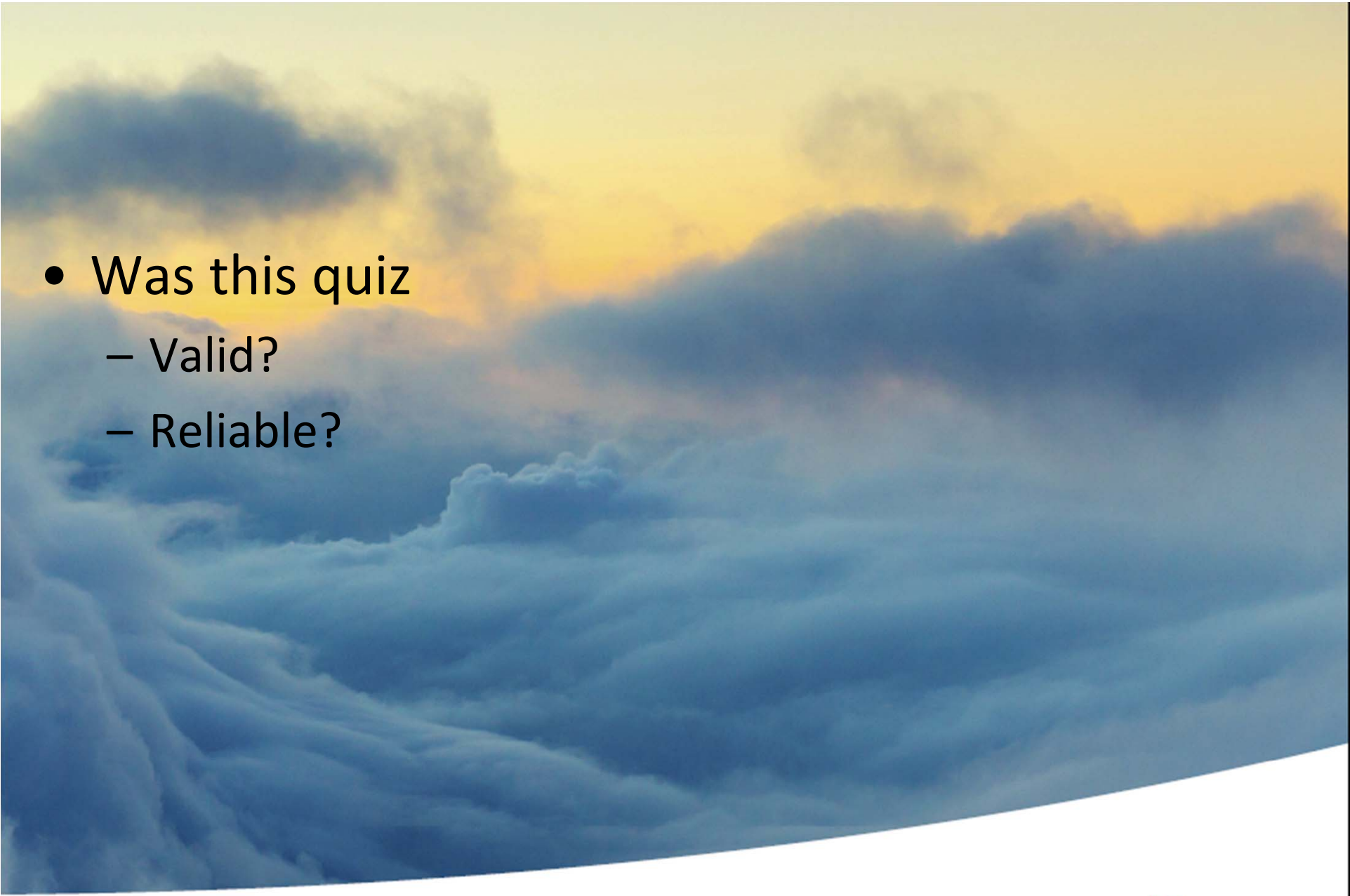
9. Which of the following compromise validity in language testing?

- a) When candidates are able to prepare responses
- b) When abilities not intended to be measured affect results
- c) When scoring processes are not based on robust rationale
- d) When insufficient information about candidates' abilities is collected
- e) All of the above



10. Tests banks which are made publicly available or have limited numbers of test versions

- a) are not valid language tests
- b) can be fairer because they allows candidates to prepare and be relaxed
- c) are only effective for assessing some skills such as pronunciation, fluency or interactions
- d) are effective because they encourage candidates to rehearse responses which promotes language learning

- 
- Was this quiz
    - Valid?
    - Reliable?

# Key criteria: The test instruments

- **Validity:** Test targets what we want to measure for accuracy and confidence in results
- **Authenticity:** Relevance and applicability - aligns with the kind of language and how it is used in real world situations
- **Reliability:** Consistency and dependability
- **Maintenance:** Monitoring and renewal of content



# Best practice in LPR testing

- **Reliability** consistency and dependability

## Raters

- Pool of raters' evaluations calibrated
- Two or more raters' evaluations are similar
- Each rater evaluates consistently over time

## Test instrument

- Scores for items consistent
- Quality of test tasks and the effect they have on results

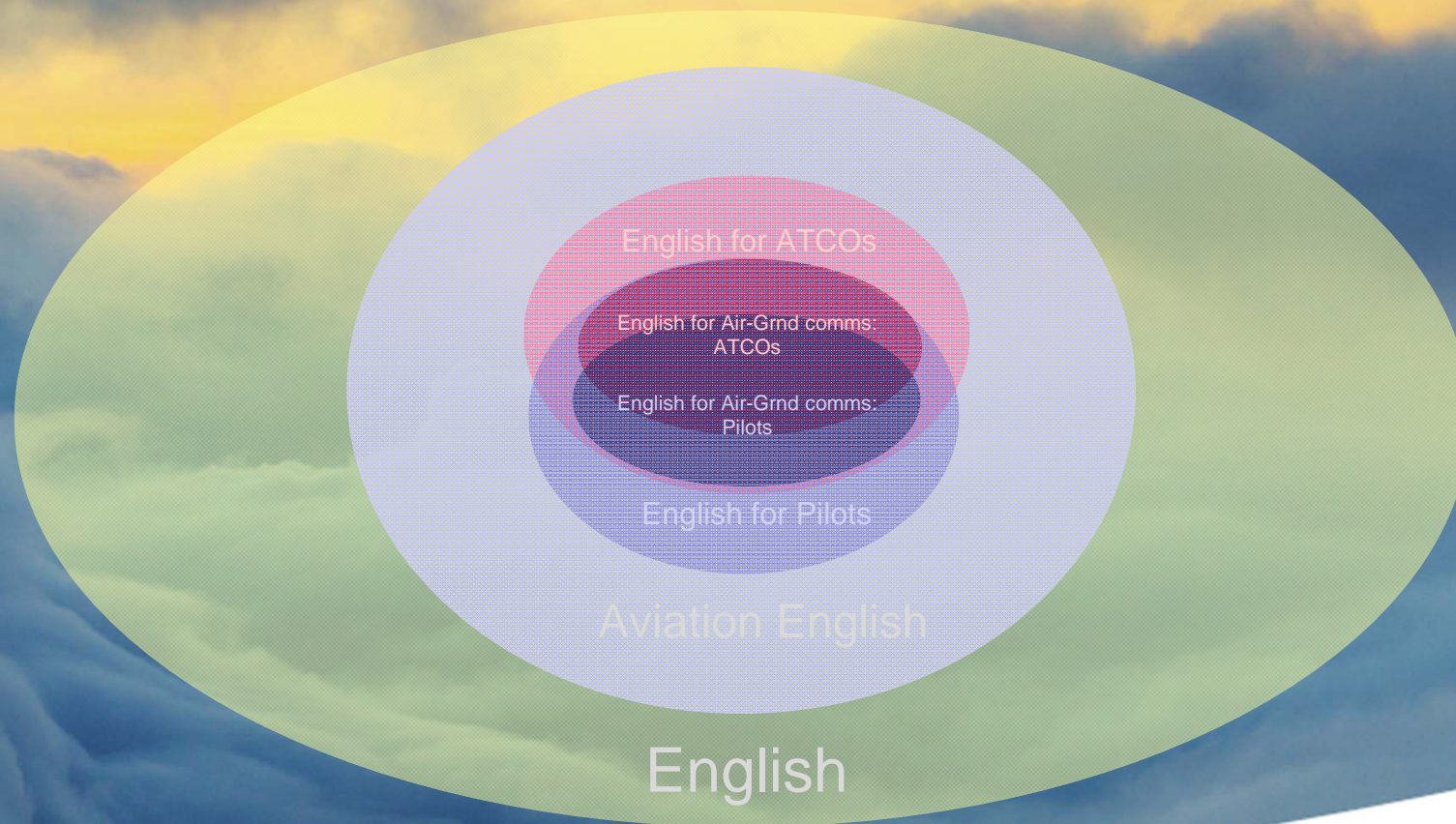
- **Validity** accuracy and confidence in results
  - Content of test (contexts, scenarios and the language they generate relate to ATCOs/pilots)
  - Scores on parts of the test are meaningful
  - The way parts of the test contribute to an overall result is meaningful
  - Full spectrum of language can be evaluated (e.g. ICAO Levels 2, 3, 4 and 5)

# Validity in LPR tests

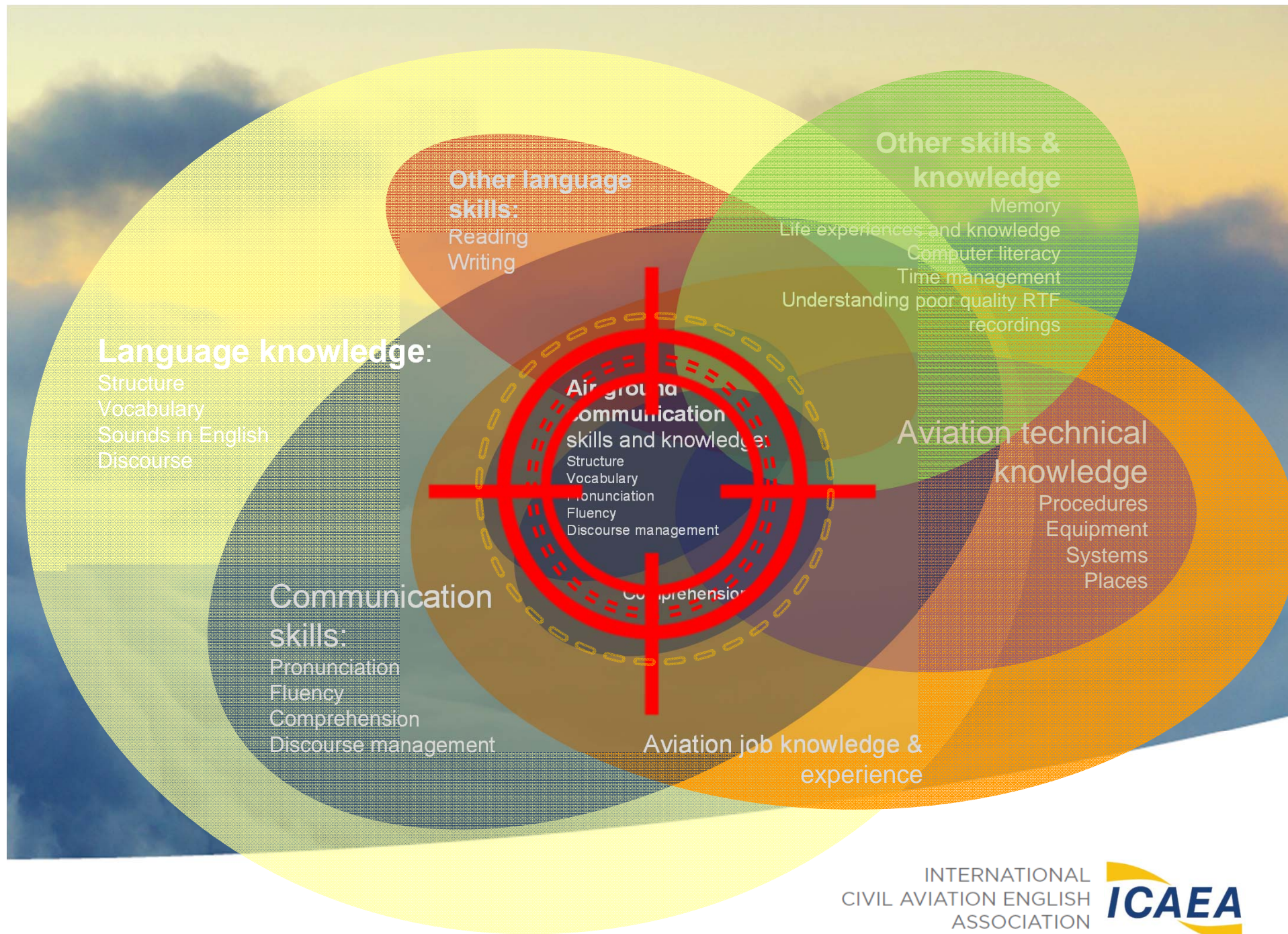
Inferences drawn based on test scores derived from tests developed in response to the ICAO LPRs should be reflective of candidates' ability to communicate in non-routine aviation situations



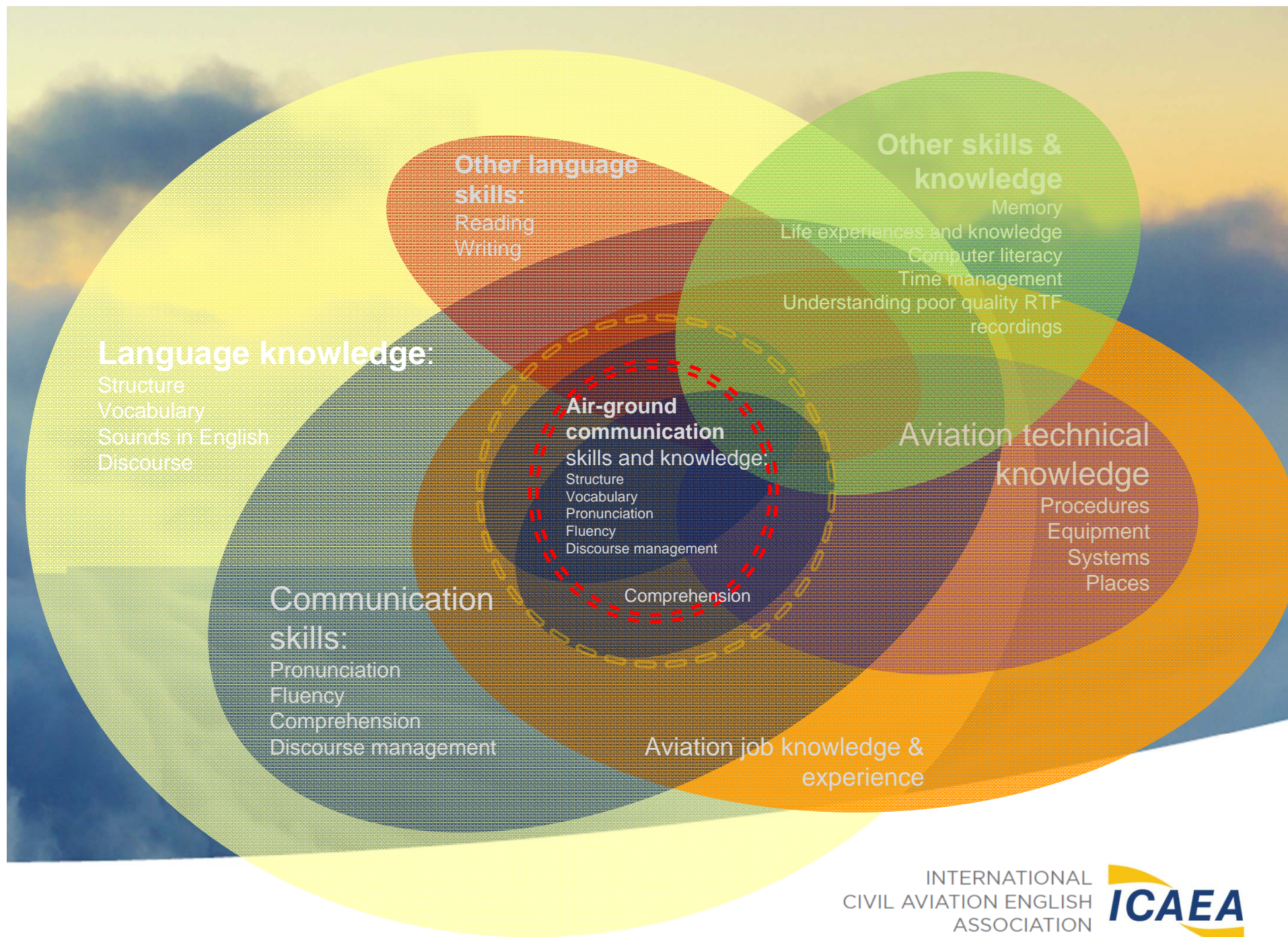
# Language assessed













# Task type Example

Candidate computer screen display

## Listening comprehension

You will hear 20 statements and conversations, each 10-20 seconds long. After you hear each one, a question will be asked about what was said. After you hear the question choose the best answer.

Audio	Item
1. The first jet aircraft to move from the use of Nickel Cadmium batteries to extensively Lithium Ion batteries was the Dreamliner.	According to the statement, which type of aircraft first used Lithium Ion batteries?  a) <input type="radio"/> The 777 b) <input checked="" type="radio"/> The B787 c) <input type="radio"/> The ATR72 d) <input type="radio"/> Most jet aircraft

# The test instrument task types

## Consider the following for each task type

1. Is the task designed for ATCOs (**A**), pilots (**P**), both ATCOs and pilots (**B**)?
2. Does the task allow candidates to produce a sufficient range and complexity of language (e.g. up to Level 5)? **Y/N**
3. Which language skills from the ICAO rating scale does the task allow for assessment (**P, S, V, I, C, F**)?
4. Does the task allow candidates to produce interactive, cohesive and co-constructed communication (authenticity)? **Y/N** or **NA**
5. How well does the task mirror real-world communication in the way pilots or controllers communicate (authenticity) – *High (H), Medium (M) or Low (L)*?
6. Would the task allow candidates of the same language level to produce perform to the same standard (is the task likely to have reliability)? **Y/N**
7. Referring to the diagram, consider what skills the task assesses (**A-G**)? Where on the diagram does this task sit for assessment purposes? (Write the *Test Task number* on the diagram)
8. Is the task useful and would it be appropriate in an ICAO LPR test? **Y/N**

# Task type 1

Candidate computer screen display

Listen to a short recording about news report about an aviation incident. After the recording finishes you will hear a beep. You will have 30 seconds to summarise what you heard. Your response will be recorded.





# Task type 2

Candidate paper-based instructions

## INSTRUCTIONS

You will be given a task card describing a non-routine situation.  
You will have 30 seconds to study the situation and prepare your response.  
You will have 3 minutes to respond.

### ***Situation***

You are en-route, at FL360, flying west to Dubai 40 miles east of the Bangkok FIR.  
The cabin crew call you to advise a passenger is unconscious and is not responding.

### ***Task***

**Describe your reaction and what you would do in this situation.**

# Task type 3

Candidate computer instructions

## Part E: Readbacks

*You will hear several spoken radiotelephony messages. Say an appropriate readback after each one. For each question, you will see a call sign for your information. Use this call sign to say an appropriate readback.*

### **Example: Cessna 29**

When you hear: "Cessna 29, hold at next intersection."  
one possible answer is: "Hold at next intersection, Cessna 29." OR  
another possible answer is: "Holding at next intersection, Cessna 29."



1. Sky Bright 315
2. Rainbow 395
3. Noble 405
4. Orange 295

# Task type 4

Examiner script

## Part 2: **Work and operations**

Ask 4 or 5 of the following questions. Rephrase and prompt as necessary. Ask additional related questions if needed (4 minutes).

1. Describe an unusual situation you have experienced in your job and what you had to do to manage it?
2. How has the job of a controller changed over the last 20 years?
3. What do you need to do if there is a pilot reports an unruly passenger onboard?
4. Tell me about why you decided to become a controller.
5. Describe the most recent job-related training session you had. What did you practise and how useful was it for your job?

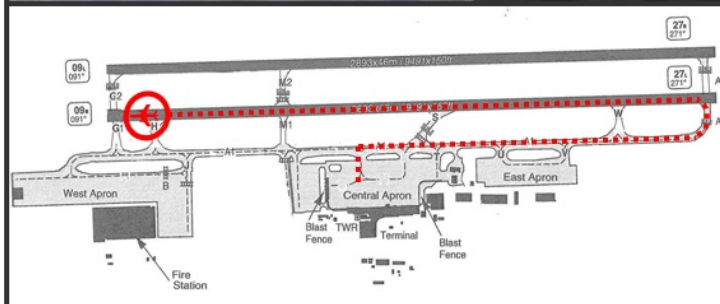


# Task type 5

**Part 2: P**  
You will p  
in the pic



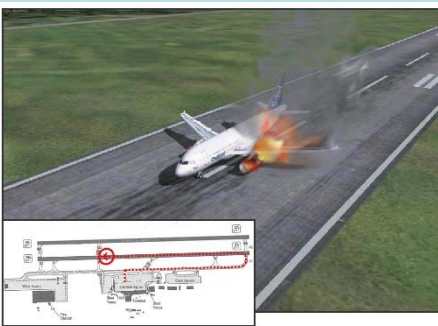
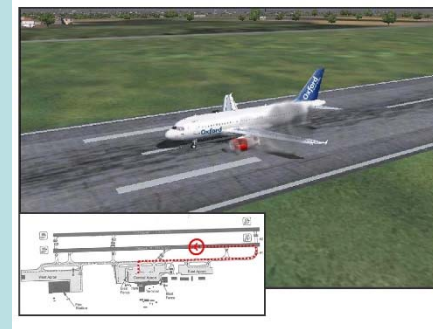
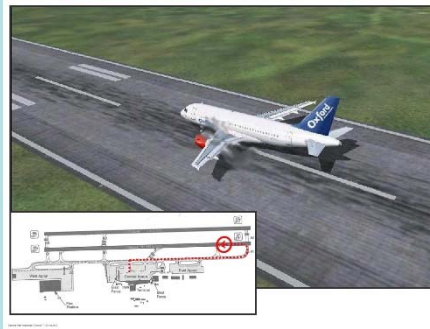
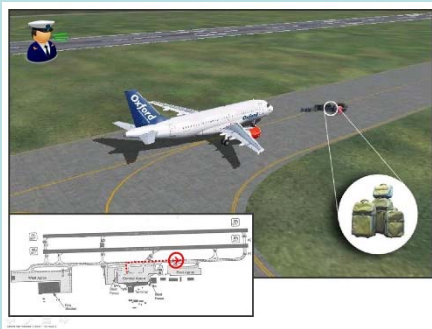
ompts



# Task type 5

## Part 2: Pilot and ATC role play

You will play the role of a pilot. Talk to and respond to ATC using the prompts in the pictures. Your responses will be recorded.





# Task type 6

Examiner script

## Section 2: Part A - Incident debrief and description

Tell the candidate that you are going to show them a series of pictures related to the ATC-PILOT roleplay they just completed in Section 1. Remind them that they can refer to this picture to describe what happened during the incident. You are their supervisor. Interview the candidate using the prompts below:

- a) What happened this morning?
- b) What do you think caused this to happen?
- c) What did you have to do?
- d) Do you think you made the right decisions?
- e) Do you have any suggestions for the company to help manage possible future events like this?

Picture A



Picture B



Picture C





# Task type 7

## Candidate instructions

You will hear 3 short pilot-ATC exchanges in non-routine situations. After each recording explain what you heard. You may take notes. You will then be asked questions about the incident.



### Exchange 1

## Examiner instructions

### Prompt as required:

- What happened?
- Where was the problem?
- Who first noticed the problem?
- What did ATC advise the pilots would happen next?

### Speculating: Ask each of the following.

1. Why did you think this problem happened?
2. Do you think the aircraft was inbound or outbound? Why?
3. What might have happened if this problem hadn't been identified?

## Audio

P1: Boston Tower, Japan 7 heavy, we request hold short for ten minutes due to a system trouble.

T: Japan Air 7 heavy, Boston Tower, you need ten minutes?

P1: Affirmative Japan 7 heavy.

T: Japan 7 heavy, roger. Hold short 22 Left. Advise ready for departure.

P2: Hey, Tower, American 5881.

G: American 5881.

P2: Hey, that Japan Air, they may know it, but they've got fuel or something spilling out of the outboard left wing there. Quite a bit.

G: Very good. Thank you very much.

T: Japan Air 7 heavy, Tower.

P1: Japan Air 7, go ahead.

T: Japan Air 7 heavy, the last departure, actually the aircraft that is sitting on the approach end of 22 R just reported a large amount of fuel that appears to be coming from your left wing. Japan Air 7 heavy, we're going to send a fire engine truck out to your aircraft to check if everything is ok.



# Task type 8

Candidate test paper

## Listening comprehension

You will hear a news report about an aviation incident.

Answer the questions by writing short sentences while you are listening. You will hear the recording twice.

### Questions



1. What had the pilot requested?
2. What kind of weather is often problematic in this part of the world?
3. What type of aircraft was it?
4. What is the most likely reason for the cause of the accident?
5. Why was the accident unlikely to have been because of a technical problem?
6. Why do they think the incident happened very quickly?

# Task type 9

Candidate computer screen display

You will hear 15 short pilot transmissions. Explain what each transmission is about.



Audio script

Audio
1. We are reporting loss of pressurisation. I need lower immediately.
2. We advise that we are experiencing some severe turbulence at 22,000 feet.
3. There is a build-up ahead of me. Request deviate 2 degrees left of course.
4. We're meeting severe icing at 230. I'd like to drop down lower.
5. The forecast says we'll have freezing rain. We can expect hazardous run conditions.
6. We're cancelling our mission. My co-pilot is incapacitated and is unable to continue.



# Task type 10

Candidate test paper

## Part 4: Picture description

Describe what you can see in the picture the examiner will show you. You will have 1 minute to prepare then 5 minutes to talk.

Examiner script

Provide one of the pictures for the candidate to describe. Allow 1 minute preparation and 5 minutes to present.



Picture A



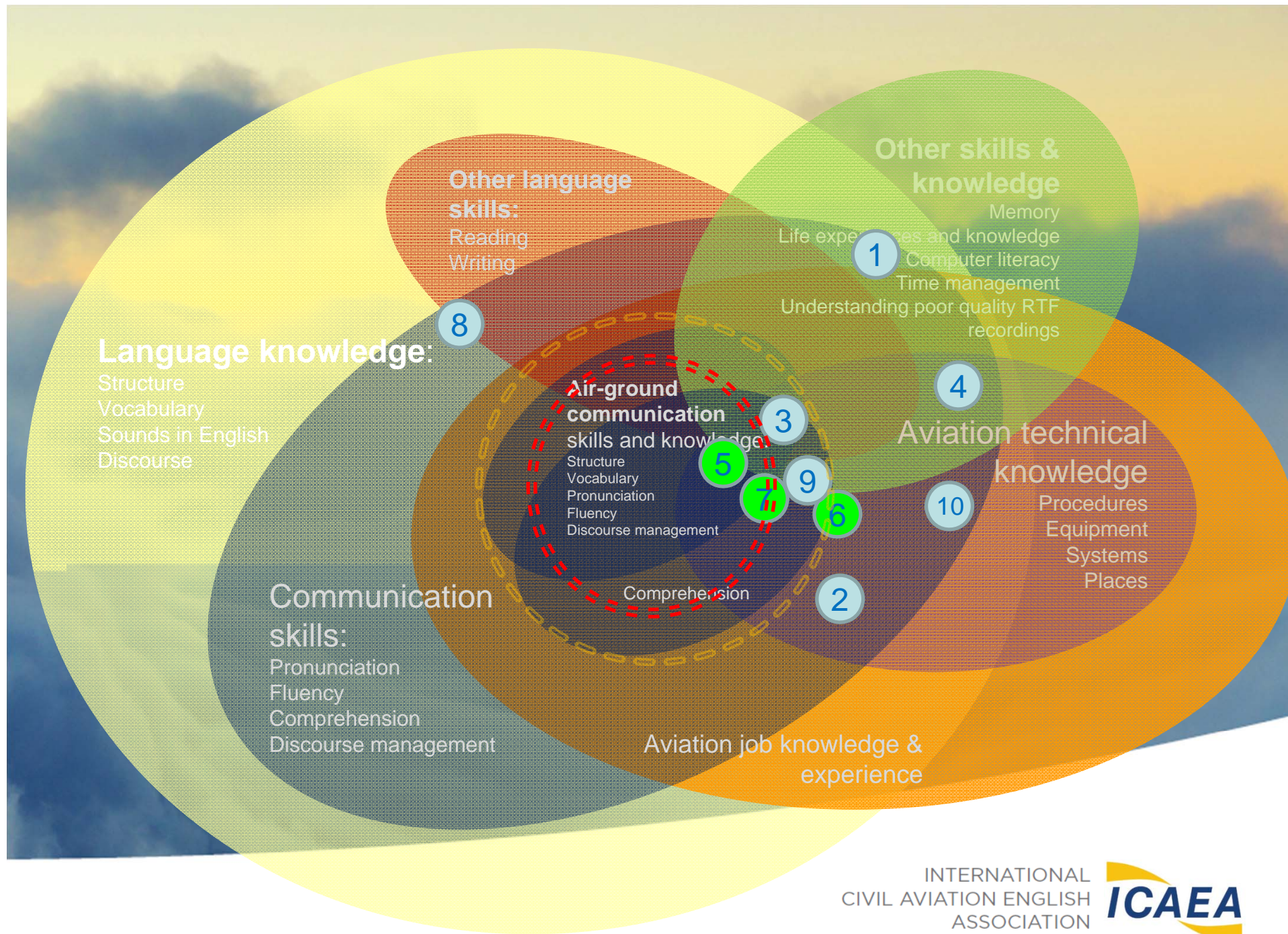
Picture B



Picture C

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10
1. Is the task designed for ATCOs (A), pilots (P), or both pilots and ATCOs (B)?	B	P	P	A	P	B	B	B	B	B
2. Does the task allow candidates to produce/understand a sufficient range and complexity of language (e.g. up to Level 5)? Y/N	Y	Y	N	Y	Y	Y	Y	Y	N	Y
3. Which language skills from the ICAO rating scale does the task allow for assessment (P, S, V, I, C, F)?	P, S, V, C, F	P, S, V, F	P, F	P, S, V, F	P, S, V, I, C, F	P, S, V, I, F	P, S, V, I, C, F	C	P, S, V, C, F	P, S, V, F
4. Does the task allow candidates to produce interactive cohesive and co-constructed communication (authenticity)? Y/N or NA	N	N	N	N	Y	Y	Y	NA	N	N
5. How well does the task mirror real-world communication in the way pilots or controllers communicate (authenticity) High (H), Medium (M) or Low (L)?	L	L	M	L	H	H	M	L	L	L
6. Would the task allow candidates of the same language level to produce perform to the same standard (is the task likely to have reliability)? Y/N	N	N	Y	N	Y	Y	M	Y	Y	N
7. Referring to the diagram, consider what skills the task assesses (A-G)? Where on the diagram does this task sit for assessment purposes? (Write the Test Task number on the diagram)										
8. Is the task useful and would it be appropriate in an ICAO LPR test? Y/N	N	N	N	N	Y	Y	Y	N	N	N







# Check list: what to look for in test tasks

- ✓ Contains sufficient breadth of tasks to allow a spectrum of language levels to be assessed
- ✓ Contain tasks that produce language so that all the skills can be fully evaluated (P, S, V, F, C, I)
- ✓ Include a range of task types to provide multiple opportunities for language skills and use
- ✓ Produce sufficient amount of spoken language for rating (20 + mins)
- ✓ Test tasks do not assess technical knowledge
- ✓ There is a rational and distinct purpose for each part of the test
- ✓ Assesses pilot/ATCO language for on-the-job needs (e.g. air-ground communications)
- ✓ The results the tasks/questions produce align with the ICAO rating bands



# Check list: what to look for in test instruments

- ✓ Assesses pilot or ATCOs (separate test instruments)
- ✓ Contains parts specifically for assessing listening and speaking skills separately
- ✓ Contains content **directly** related to language needs of pilots or ATCOs for use of English in R/T communications
- ✓ Allows for both voice-only and face-to-face communication
- ✓ Mirrors real-world communication contexts (authenticity)
- ✓ Sufficient bank of equivalent test versions/content is available (20+)
- ✓ Content is reviewed and versions are regularly replaced with new versions

# Trademarks of inferior test instruments

- ✗ Content or tasks are not linked to job language use
- ✗ Structure of the test is overly simplistic (not capable of allowing evaluation of skills or ICAO levels adequately)
- ✗ Test tasks are too easy (unable to effectively evaluate higher levels)
- ✗ Contains inadequate variety in tasks and content (limits effectiveness in drawing out different range and use of language) –e.g. just Q & A scenarios
- ✗ Parts of test are not coherent (do not serve a function for assessment)
- ✗ Test tasks rely on technical knowledge, memory or other skills which affect results
- ✗ Test bank contains limited versions or is publicly available (promotes rehearsing for test rather than constructive language development)



# Evidence of reliability

TSPs should provide evidence of:

- Methods of evaluating consistency of different test versions
- Listening tests with items:
  - Reports of internal consistency: statistical analysis e.g. showing correlations among all the variables (a high reliability estimate should be as close to 1 as possible).
  - Reports on item difficulty: statistical analysis showing that the behaviour of
- Methods of determining equivalence between raters
- System checks to monitor/identify rater differences

# Evidence of validity

TSPs should provide evidence of:

- Test trials where data is collected and analysed – comparisons with other known tests (e.g. correlations)
- Data/statistical evidence of justification for parts of tests - scores and effect on results (e.g. analysis of comparisons of results across test parts)
- Statistical evidence of justification for results based on listening scores and ICAO rating scale bands
- Extensive qualitative feedback from examiners and test takers on content, delivery, timing, usability, access, appropriateness, task requirements etc
- Reports from expert review panels (content, task types, standard-setting, authenticity etc)



[www.icaea-aero.org](http://www.icaea-aero.org)

[michael.kay.icaea@gmail.com](mailto:michael.kay.icaea@gmail.com)

Next event: **Training for Operational Personnel**

29-30 April 2016, Portugal