Report 03 - Description of Tasks & Instructions

For the purposes of this report we adopt the definition of 'task' proposed by the Association of Language Testers in Europe (2001): "*a combination of rubric, item and response.*" By 'rubric', we mean the task instructions – as delivered by the interlocutor; by 'item', we mean individual question or recording, where a group of items can be responded to by referring to a single rubric.

In TEA Version 2010, the introduction (in which the candidate's identity is confirmed and the contents of the test are summarized) is followed by 6 tasks:

Part of Test	Type of Task	Delivery Type	Nature of Interaction	Number of Items	Time Restrictions
Part 1	Interview	Direct	Examiner interviews candidate using verbal questions based on familiar aviation topics	7 maximum	Maximum 5 – 6 minutes
Part 2A	Comprehension Task	Semi Direct Voice only	Candidate asked to describe the details of short recorded items set in non-routine aviation contexts	10	Maximum response time of 20 seconds
Part 2B	Interactive Comprehension	Semi Direct Voice only	Candidate asked to respond (by asking questions) to short recorded items set in general aviation contexts	3	Maximum response time of 20 seconds
Part 2C	Interactive Comprehension	Semi Direct Voice only	Candidate asked to respond (by giving advice) to short recorded items set in general aviation contexts	3	Maximum response time of 20 seconds
Part 3	Picture Description	Direct	Candidate asked to describe and compare sets of 2 pictures	2	Maximum response time of 1 minute

| Part 3 | Discussion | Direct | Examiner asks candidate to participate in discussion of more general aviation nature based on verbal questions thematically linked to the 2nd picture set | 9 scripted questions with interlocutor flexibility | Maximum 4 – 5 minutes |

The test format and task-types are standardised, in order that:

- candidates have equal opportunity to demonstrate language proficiency
- candidates familiar with TEA format perform their best
- the test produces consistent & reliable results.

For large candidatures, unstructured tasks produce unreliable results because of human variables.

In line with test design protocol, the test tasks are designed to develop in complexity as a candidate progresses through the test. Candidates can feel initially comfortable and demonstrate their level as complexity levels increase as the test progresses. In Part 3, interlocutors are able to probe language ability further.

Below, each task is described individually along with:
- a list of which language competences each task is designed to measure*
- a rationale for item design
- a description of interlocutor behaviour
- a description of development (evolution) to the present version
- a presentation of the task rubric.

*TEA development works in line with Messick's view of a 'construct-centered approach' in which he outlines that, once the language constructs (or competences) have been identified, the test developer must ask himself "what behaviours or behaviours should reveal those constructs, and what tasks or situations should elicit those behaviours? Thus, the nature of the construct guides the selection or construction of relevant tasks" (1994:17). Of the productive, receptive and interactive competences listed in **Report 01 – Description of Test Purpose, Specifications & Construction**, some are the main or *primary focus* of certain tasks i.e. the task is designed to specifically assess these competences. Some tasks, however, also allow for the assessment of other competences since the task generates opportunities for demonstration of different abilities (or lack of). These competences have been labelled under each task's *secondary focus*. See **Report 05 – Analysis of Competences Measured in TEA** for further information.

For information about item development and trialling, see *Report 08 – Item Development & Version Content*.

Part 1 Interview

Compared with a discussion or conversation, the interview task is tightly structured and takes the form of Question-Answer. The interlocutor is trained to stick to the scripted *sets* of questions (1 set per candidate) but allow the candidate a degree of freedom in their responses. The questions are chosen for their success in eliciting a representative sample of the candidate's plain English.

Language competencies task is designed to measure

Primary Focus
- Talk about familiar, common, concrete and work-related topics specific to the candidate's role in aviation
- Talk about familiar, common, concrete and work-related topics common to pilot – controller roles in general
- Use a range of basic and complex grammatical structures as appropriate to the function of the task
- Use a range of work-related vocabulary
- Produce connected stretches of language, sometimes at length
- Use a range of phonological features (sound, stress, rhythm, and intonation) to produce speech intelligible to the international aviation community
- Recognise the illocutionary force (the communication purpose) of the speaker
- Process linguistic difficulties such as tense and modality
- Where necessary, demonstrate discourse management strategies to resolve misunderstanding
- Respond immediately, appropriately and informatively

Secondary Focus
- When lacking vocabulary, use circumlocution strategies
- Manage and maintain the speaker-listener relationship

Item Design

As discussed further in both *Report 01 – Description of Test Purpose, Specifications & Construction* and *Report 02 - Overview of Expert Judgements and Action Taken in Test Development*, the following specifications were set out for task construction:

- Questions should focus on common, concrete and work-related topics in order to
  o be assessable by the ICAO Descriptors
  o give candidates at every level the opportunity to demonstrate their ability to respond on such topics

- o allow the candidate to relax a little.
- Questions should focus on plain English in order to avoid eliciting (non-assessable) standardised phraseology, technical or operational language.
- Question sets should elicit a variety of basic language functions equivalent to those listed in ICAO's *Communicative Language Functions*.
- Questions within sets should be disconnected in order that one question will not disturb the whole task, and to avoid the concept of 'easier' or 'more difficult' topics and, as a result, unreliable results.
- Questions within sets should be graded in order of increasing difficulty to allow candidates to relax into the test and allow raters to assess language ability in a linear fashion.
- The language of the questions should be moderated to give candidates at all levels the opportunity to respond

In terms of delivery, the interlocutor's behaviour is governed by the following specifications:

- There is no allowance for interlocutor flexibility since the language level may be compromised and tests may not be standardised.
- The interlocutor can only ask follow-up questions where a follow-up prompt is scripted in parentheses (for example, *Why?* Or *Why not?* after a closed question).
- The interlocutor cannot 'assist' the candidate unless the candidate has interacted (and sought assistance by, for example, asking for a question to be rephrased or repeated).
- The interlocutor should not correct errors, fill silences and interrupt unnecessarily.
- The interlocutor should try to ask all of the questions in the set while adhering to the maximum time limit of 5 to 6 minutes.
- The interlocutor should adopt an even, 'neutral' tone, avoiding criticism, praise and gesturing. (see ***Report 11 - Examiner Training & Assessment using TEA*** for full details on interlocution training and procedures.)

Development

- The number of scripted questions (and the previous time limit of 3 -4 minutes) was extended in order to:
  - o elicit a greater range of language functions
  - o elicit a larger sample of assessable language related to common, concrete and work-related topics.
- Questions sets were developed not by topic, but by range of target language functions.
- Questions sets specific to private pilots, and student pilots and controllers were trialled and operationalised.

Part 1 Question Set - <u>Example</u> Rubric:

Now, in this first part, I'd like to ask you some questions about yourself:

- *Could you tell me about your job?*

- *Can you describe what you do when you are working?*

- *Why did you choose to be a _____ ?*

- *Is a _____'s job important? (Why / Why not?)*

- *What is the difference between the role of a _____ today compared to the past?*

- *How much training is necessary to become a _____ ? (Could you tell me more?)*

- *In the future, what would you like to change about what you do?*

  [The TEA Administrator informs the examiner of each candidate's aviation *role* before the test starts.]

Part 2

Tasks in this part of the test comply with ICAO's requirements for a voice-only element to the test. Candidates are exposed to a range of international accents via semi-direct delivery (a CD of 16 recordings). Since candidates are asked to respond immediately, the tasks elicit responses that can assess performance across all 6 of the Descriptors, including Comprehension and Interactions.

Rationale

As discussed further in both *Report 01 – Description of Test Purpose, Specifications & Construction* and *Report 02 - Overview of Expert Judgements and Action Taken in Test Development*, the following specifications were set out for task construction within Part 2:

- Items should focus on plain English in order to avoid eliciting (non-assessable) standardised phraseology, technical or operational language.
- Items should be disconnected in order that one will not disturb the whole task.
- Items can be repeated if the candidate has used a successful clarification strategy, but can only be repeated once. The need for repetition is a strong indicator of language processing ability.
- A response time limit (of 20 seconds) allows for assessment of language proficiency and induces a degree of stress to partially-replicate the target language use situations.

In terms of delivery, the interlocutor's behaviour is governed by the following specifications:

- The interlocutor can replay a recording once more if the candidate has used a clarification strategy.
- The interlocutor cannot ask any follow-up questions since this would lead to unreliable assessment.
- The interlocutor cannot answer any candidate's clarification strategy directly since that would compromise the Comprehension score.

- Following each recording, the interlocutor gives the appropriate prompt for the task in order to place the demand on linguistic processing within the short-term memory rather than simple repetition or imitation.
- The interlocutor should adopt an even, 'neutral' tone, avoiding criticism, praise and gesturing.

Part 2A Comprehension Task

This is a short-message processing task. Candidates hear 10 speakers, both pilots and controllers, speaking in 10 non-routine situations set in a variety of aviation-specific traffic operations, both on the ground and in the air. While short, grammatically-simple items may be easily recalled in short-term memory and imitated, longer or more complex items need to be stored and processed internally before being reprocessed, possibly through paraphrase (Underhill, 1987). In order to allow for the assessment of Levels 1 – 6, a variety and balance of items is required so that even candidates at higher levels of proficiency experience some problems processing the items accurately. Underhill states that "differences of familiarity of vocabulary and complexity of structure can make short sentences difficult and long ones easy" (p.87), thus enabling the construction of items of different levels of difficulty.

Language competencies task is designed to measure

Primary Focus
- Talk about familiar, common, concrete and work-related topics common to pilot – controller roles in general
- Use a range of basic and complex grammatical structures as appropriate to the function of the task
- Use a range of work-related vocabulary
- Use a range of phonological features (sound, stress, rhythm, and intonation) to produce speech intelligible to the international aviation community
- Understand the specific details of short messages delivered by both pilots and controllers in plain English in non-routine situations at different phases of flight (tower, ground, departure, en-route, approach)
- Understand a range of native and non-native speakers in terms of accent and rate of speech
- Process linguistic difficulties such as
  - tense
  - modality
  - lower frequency work-related vocabulary
  - negation
  - contraction
- Recognise the illocutionary force (the communication purpose) of the speaker
- Where necessary, demonstrate discourse management strategies to resolve misunderstanding
- Respond immediately, appropriately and informatively

Secondary Focus
- When lacking vocabulary, use circumlocution strategies
- Manage and maintain the speaker-listener relationship

Rationale

The construction of the items is based on ICAO's categorisation of dominant communicative functions in pilot-controller dialogue where carrying out ATC and piloting tasks. These categories are:
1. Triggering actions
2. Sharing information
3. Managing the pilot-controller relationship
4. Managing the dialogue.

In Part 2A, the primary communicative function is clearly to share information through accurately relaying important information. The situations are constructed of two linked parts and have four possible formats (although their internal order is not fixed):
- Statement of Situation or Action + Elaborated Statement of Situation or Action
- Statement of Situation or Action + Request
- Statement of Situation or Action + Question
- Statement of Situation or Action + Command.

Candidates are asked to explain their understanding of each situation. Aspects of item construction that are intended to affect ease of comprehension include:
- Accent
- Delivery speed & use of contraction
- Speaker context & purpose
- Tense & modality
- Lexical familiarity
- Expression of certainty or doubt

Items are co-written with operational experts to ensure the authenticity of the situations and the ATC and piloting tasks. Items are balanced within each test version so as to characterise a fair variety of sub-topics (within the broader categories of Health, Technical, ATC & Ground, Weather and Interference) and construct features (see *Report 08 – Item Development & Version Content* for full details.)

Development

- Different rubrics and prompts were trialled and operationalised. Candidate performance suggested that candidates needed to be made aware that it was important to relay all the critical information of each message, and that if they needed to hear a recording again they needed to ask. Candidate feedback suggested that "*What was the message?*" was clearer and more appropriate than "*What's happening?*" or "*What did you understand?*".
- Several aspects of the recordings themselves were trialled to make recordings clear yet authentic:
  - Background noise was abandoned as a result of negative feedback from candidates.
  - Different 'tinny' effects were trialled to replicate the sound and clarity of typical, aeronautical radiotelephony.
  - Clicks were added to replicate the sounds that define the beginning and end of authentic aeronautical radiotelephony.

Part 2A Rubric

Now you're going to hear some recordings of international speakers.

*Firstly, you'll hear pilots and controllers talking in **non-routine** situations.*

*After each recording, **tell me the message**.*

***ALL** the information is important.*

*If you need to hear it again, **please ask**.*

You have 10 seconds to start speaking. OK?

Play recordings 1 – 10, pausing after each one.
After each recording, give the following prompt:

What was the message?

You **must** give this instruction each time.

Notes to Examiner:

- If the candidate begins speaking within 10 seconds, allow a maximum of a further 10 seconds before moving to   the next stage. [Say *Thank you*, and continue.]

- The examiner may play a recording a second time only when a candidate has used a clarification   strategy. In this case, the examiner states:

OK, I'll play it once more.

The examiner may not play the recording a third time.

Parts 2B & 2C Interactive Comprehension

These tasks build on the sentence-processing task in Part 2A but are set in more general aviation situations (Comprehension may be affected by the same type of features listed above, under Part 2A). Candidates are asked to respond appropriately to 6 speakers briefly describing situational complications - by asking questions to find out more about each situation (3 in Part 2B), and giving advice to the speaker (3 in Part 2C).

Language competencies task is designed to measure

Primary Focus
- Talk about aviation-related topics in a broader context
- Use a range of basic and complex grammatical structures as appropriate to the function of the task
- Use a range of work-related vocabulary
- Use a range of phonological features (sound, stress, rhythm, and intonation) to produce speech intelligible to the international aviation community
- Understand a range of native and non-native speakers in terms of accent and rate of speech
- Process linguistic difficulties such as
  - tense
  - modality
  - lower frequency work-related vocabulary
  - negation
  - contraction
- Understand and respond to short messages describing linguistic or situational complications or an unexpected turn of events (in an aviation context).
- Where necessary, demonstrate discourse management strategies to resolve misunderstanding

- Respond immediately, appropriately and informatively

Secondary Focus
- When lacking vocabulary, use circumlocution strategies
- Manage and maintain the speaker-listener relationship

Rationale

- Problem-solving tasks are appropriate to the target language context.
- Situations are de-contextualised in order to elicit two target language functions in plain English "on various topics that are related to radiotelephony communications or aviation operations, without replicating radiotelephony communications specifically" (ICAO 9835: 6.2.8.9).
- Items are balanced within each test version so as to characterise a fair variety of sub-topics (within the broader categories of Health, Technical, Human and Environment) and construct features.

In terms of delivery in these tasks, the interlocutor's behaviour is governed by the following specifications:

- The interlocutor's prompt for Part 2B is "*Ask Questions*". If the candidate does not ask questions, the interlocutor says once only, politely but firmly, "*No, ask the speaker questions*" to reinforce understanding of the task.
- The interlocutor's prompt for Part 2C is "*Give Advice*". If the candidate does not ask questions, the interlocutor says once only, politely but firmly, "*No, give the speaker advice*" to reinforce understanding of the task.

Development

- Parts 2B and 2C were split into separate tasks after candidate feedback suggested that the original two-part format sometimes sounded inauthentic. Although it did not affect the nature of the task, it was agreed that any item that could cause confusion could impair the assessment of comprehension since the ability to anticipate, based on background knowledge of the way things work naturally, can play a part in listening comprehension ability. In trialling and operationalising the two tasks independently of each other, this potential disturbance was overcome.
- The splitting of the tasks rendered the need for an example situation redundant. However, it was noticed in trialling that candidates at lower-levels occasionally failed to understand the demands of the task. To combat the following improvements were made:
  o clearer pre-test instructions in the *Notes for TEA Candidates* to indicate that there are no correct answers, but that candidates should show they understand the situations by giving relevant responses
  o interlocutors were permitted to re-prompt candidates who were not following the task instructions, where necessary.

Part 2B Rubric:

*Now you'll hear 3 recordings in more **general** situations.*

*Listen to the speaker and **ask some questions** to find out more about the situation.*

You have 10 seconds to respond.

*If you need to hear it again, **please ask**. OK?*

Play recordings 11 – 13, pausing after each one.
After each recording, give the following prompt:

Ask questions.

You **must** give this instruction each time.

Notes to Examiner:

- If the candidate does not ask questions, politely but firmly say:

  No, ask the speaker questions.

- If the candidate does not understand the recording, say *Thank you*, and continue.

- If the candidate expresses a lack of understanding of the instructions, briefly paraphrase the instructions for them.

- If the candidate begins speaking within 10 seconds, allow a maximum of a further 10 seconds before moving to the next stage. [Say *Thank you*, and continue.]

- The examiner may play a recording a second time only when a candidate has used a clarification strategy. In this case, the examiner states:

OK, I'll play it once more.

The examiner may not play the recording a third time.

Part 3 Pictures Description

Candidates are given a set of two thematically-linked photographs and asked to describe and compare them with as much information as possible. They are told they have 1 minute to speak and when to start (and stop). Although candidates are asked to describe and compare the picture sets, assessment does not focus on task achievement, so the task allows for personal interpretation and style. The task also includes preparation time since studies by Wigglessorth (1997) and Skehan and Foster (1997) indicate that preparation time and forethought can enhance performance.

Language competencies task is designed to measure

Primary Focus
- Talk about familiar, common, concrete and work-related topics common to pilot – controller roles in general
- Talk about aviation-related topics in a broader context
- Use a range of basic and complex grammatical structures as appropriate to the function of the task
- Use a range of work-related vocabulary
- When lacking vocabulary, use circumlocution strategies
- Produce connected stretches of language, sometimes at length
- Use a range of phonological features (sound, stress, rhythm, and intonation) to produce speech intelligible to the international aviation community

Rationale
- The selected sets of 2 photographs should
  - be of good quality and clearly set in aviation contexts
  - be linked thematically (e.g. linked to the topic of 'aviation security')
  - illustrate broad scenes of activity and detail in order to potentially elicit a range of sentence structures and vocabulary
  - depict typical and atypical or unexpected scenes in order to potentially elicit a variety of language functions beyond description and comparison (e.g. speculation, opinion-giving, prediction)
  - contain features that are clearly comparable (in terms of similar or different activities, views, numbers, problems, etc.)

  In terms of delivery, the interlocutor's behaviour is governed by the following specifications:
- If a candidate ends their turn early, or it is clear they are not sticking to task, the interlocutor can prompt or re-direct them by saying "*You have more time*" or "*Please describe and compare the pictures*" in order to elicit the most appropriate sample possible.

- The interlocutor cannot prompt the candidate to describe specific aspects of the pictures or ask follow-up questions since that would compromise test reliability.
- The interlocutor takes the pictures back at the end of the candidate's turn so they may not be used as a reference point later in the test.

Development
- Trials indicated that a greater variety of language, and language functions, was elicited from a set of 2 pictures than 1 picture. Candidates are asked to describe and compare and, although many do not directly compare, a greater variety of appropriate, assessable elicited language makes the format advantageous.
- A variety of rubrics were trialled for the new 2-picture task. The following rubric – "*Now I'm going to show you two pictures related to Health. Please look at them for a few moments….. Please describe both pictures with as much information as possible. Tell me what is the same and what is different. You have one minute to speak. Please start now.*" – evolved into the final rubric (see below) after the following respondent feedback and discussion:
  o It was unnecessary to state the thematic link of the 2 photographs.
  o The wording of the rubric could be greatly reduced once it was established that 'compare' was understood by the majority of respondents and that the improved *Notes for TEA Candidates* could help to alleviate any doubts.
  o The task instructions should be given before the thinking time commenced.
  o One minute was adequate time for respondents of all levels to talk about both pictures at some length.

Part 3 Picture Task Rubric:

Each candidate should be presented with two sets of pictures in part three.
The first set is for description, the second set is for description and extended discussion.

*Now I'm going to show you **two pictures** which I'd like you to describe and compare with as much information as possible. Please think about them for a few moments.*

The examiner places the first set on the table in front of the candidate.

After 15 seconds, the examiner states:

*OK, you have **one minute** to speak. Please start now.*

After a maximum of 60 seconds, the examiner states:

Thank you.

*Now I'm going to show you **two more pictures** which I'd like you to describe and compare with as much information as possible. Again, please think about them for a few moments.*

The examiner places the second set on the table in front of the candidate.

After 15 seconds, the examiner states:

*OK, you have **one minute** to speak. Please start now.*

After a maximum of 60 seconds, the examiner states:

Thank you.

Part 3 Discussion

In Part 3, the two-way discussion, the 'frame' is looser and the examiner can accommodate their language level to the level of the candidate by fashioning appropriate questions from graded question prompts. The scripted sub-sets provide support for lower level candidates while still allowing higher level candidates the opportunity to demonstrate their proficiency. This approach is supported by Underhill, who writes:

"the tightly-controlled interview that is more like a question and answer test will not easily elicit the candidate's best language performance, and it is therefore more important at higher levels that the interviewer stand back a little and allow the candidate to display his conversational fluency skills. Filling conversational pauses naturally and correcting one's own errors are two of the features that distinguish higher-level from intermediate-level candidates. Ultimately, the interviewer must be prepared to yield the initiative to the candidate, as is the aim in discussion/conversation." (p.56)

Language competencies task is designed to measure

Primary Focus
- Talk about familiar, common, concrete and work-related topics common to pilot – controller roles in general
- Talk about aviation-related topics in a broader context
- Use a range of basic and complex grammatical structures as appropriate to the function of the task
- Use a range of work-related vocabulary
- When lacking vocabulary, use circumlocution strategies
- Produce connected stretches of language, sometimes at length

- Use a range of phonological features (sound, stress, rhythm, and intonation) to produce speech intelligible to the international aviation community
- Process linguistic difficulties such as
  - tense
  - modality
  - lower frequency work-related vocabulary
- Recognise the illocutionary force (the communication purpose) of the speaker
- Where necessary, demonstrate discourse management strategies to resolve misunderstanding
- Manage and maintain the speaker-listener relationship
- Respond immediately, appropriately and informatively

Secondary Focus
- Talk about familiar, common, concrete and work-related topics specific to the candidate's role in aviation.

Rationale

As discussed further in both *Report 01 – Description of Test Purpose, Specifications & Construction* and *Report 02 - Overview of Expert Judgements and Action Taken in Test Development*, the following specifications were set out for task construction:

- Discussion topics should be thematically linked from the second picture set so as to bridge to the discussion naturally.
- 3 sub-topics of 3 questions should be developed for each discussion page to be used randomly and as necessary by the interlocutor
- Questions should focus on broader, less familiar aviation topics in order to
  - be assessable by the ICAO Descriptors
  - give candidates at higher levels the opportunity to demonstrate their ability to respond on such topics (one of the guiding principles in developing test items for Parts 3 was for item writers to ask themselves whether the candidate would be likely to be able to answer the question easily, sensibly and fully in his own first language).
- Questions should focus on plain English in order to avoid eliciting (non-assessable) standardised phraseology, technical or operational language.
- Question sets should elicit a variety of basic and complex language functions equivalent to those listed in ICAO's *Communicative Language Functions*.
- Questions within sets should be linked in order that a natural discussion may develop
- Questions within sets should be graded in order of increasing difficulty to allow candidates at lower levels the opportunity to respond to the first questions and to allow for assessment of language ability in a linear fashion.
- The language of the questions should be clear but less moderated than Part 1 in order to extend candidates at the higher-levels.

In terms of delivery, the interlocutor's behaviour is governed by the following specifications:

- There is allowance for interlocutor flexibility since a natural discussion is to be encouraged – interlocutors are trained to allow candidates   to show their ability to spontaneously interact and elaborate on answers by asking suitable follow-up prompts
- The interlocutor can 'assist' the candidate if the candidate has interacted (and sought assistance by, for example, asking for a question to   be rephrased or repeated)
- The interlocutor should try to ask all of the questions in at least one set while adhering to the maximum time limit of 4 to 5 minutes.

Part 3 Discussion Task Rubric (<u>partially-hidden</u>)**:**

*Now I'd like to discuss some more general questions related to **Emergencies and Safety**.*

*Begin with **any** question set.*

Let's consider first of all…

Emergency Situations

- *When is a situation an **emergency** situation?*

- *Which emergency situations should passengers be most worried about?*

- *People hear a lot more about emergency situations on television these days. What would you say to people who are worried about flying?*

Depressurisation

- *What can cause depressurisation?*

- *How is depressurisation different to other emergency situations?*

- *What should passengers do in the event of depressurisation?*

<u>Safety in Aviation</u>

- *What recent developments have improved safety in aviation?*

- *Is flying safer now than in the past? (Why? / Why not?)*

- *How do you think aviation safety will improve in the future?*

Thank you very much, that's the end of the test.

Alternative Tasks & Approaches Trialled

Video

Candidates were shown a 30-40 second video of an emergency situation. They were asked to describe the video as it was played, and then to summarise what they had seen afterwards. This task was designed to test candidates' ability to convey information and paraphrase spontaneously as well as use an appropriate range of structures. Candidates were not given any preparation time, in order to simulate a stressful work situation. Trials showed that it discriminated poorly between candidates of different proficiency levels since many candidates struggled to describe the situation quickly. Lower-level candidates tended to lapse into silence, falsely indicating elementary-level proficiency. Many candidates at higher levels reported that they found the task too difficult to complete due to the time pressure.

It was agreed that further adaptations and trialling should end since other limitations to the task type also became apparent:

- the lack of range of structures elicited during the task
- the shortage of multiple videos of suitable quality
- difficulties in assessing equivalence
- copyright issues.

Map Task – Describing a route

A task was developed and trialled in which candidates were given a map and asked to describe the route between two places with situational complications and hazards identified along the way. The task reflected the relevant lexical domains of weather and geography, and the critical language functions of conveying information, giving advice and issuing warnings. It was designed to assess extended fluency and vocabulary since candidates were asked to describe the travel conditions in as much detail as possible. Candidates were given preparation time and a time limit.

The initial trial proved successful with candidates reporting that they felt comfortable attempting this task as it was work-related, and appropriate, assessable language was elicited. However, the following concerns about the task evolved during repeated trialling:

- Candidates did not always understand the symbols on the map and item writers considered using written labels. However, this would have required a degree of reading comprehension, which would discriminate against some candidates, and was not in keeping with the requirements of 9835.
- Although it initially appeared easy to produce multiple versions of this task, it became clear that that the variety of responses to this task type was too limited to be a reliable indicator of spontaneous language production and range of vocabulary and structures.
- It was noted that many candidates were simply giving directions without reference to conditions. Candidate feedback forms had indicated that candidates felt comfortable with this part of the test because it was strongly work-related and, although this had initially been viewed as a positive factor, it was important that the task forced candidates to face typical and atypical or unexpected complications.

Variations in 2A

Oral Comprehension Questions

Trials were conducted using recorded, short, question prompts that followed the text and challenged candidates to firstly process the details of the text, then the question, before responding. The objective was to measure whether comprehension prompts, specific to each individual item, could be successfully operationalised in integrative testing.

Some example recordings were:

Item Number (Voice 1 – Standard) *Situation 1*
Item (Voice 2 – International accent) *There's something on the runway. We can't take off.*
Prompt (Voice 1 again) *What caused the problem?*

Item Number (Voice 1 – Standard) *Situation 2*
Item (Voice 2 – International accent) *It's possible we won't have brakes on touchdown. Can we have the longest runway?*

Prompt (Voice 1 again)   *What do they need?*

Item Number (Voice 1 – Standard) *Situation 3*
Item (Voice 2 – International accent) *We have sunk in snow. Can you send a tractor to pull us out?*
Prompt (Voice 1 again)   *What does the speaker want?*

The trials proved unsuccessful since the candidates were either over-loaded with information to process, or confused by the format. It was discussed whether interlocutors could deliver the question prompts orally following each recording but it was agreed that, by having to have the whole task scripted, test security and, therefore, reliability would be compromised. A further problem with the task format was that candidates heard the prompt after the recording - not good practice in language testing since candidates should know what they are specifically listening for prior to the text.

20-item trials

It was agreed that trials to measure the effect of extending Part 2A to 20 items were necessary to assess whether extra items would differentiate between levels more effectively than the current system. Stakeholder feedback suggested an extended 2A would be welcome, though feedback from candidates and examiners suggested otherwise.

Trials were conducted in which 20-item tasks were constructed along the formula of the 10-item 2A task. This allowed for a greater variety of accents, linguistic features and lexical domains to form part of the test content, potentially supporting an argument for improved content validity.

The respondent performance (the sample group of 30 candidates consisted of an even range of nationalities/first languages and a range of TEA Comprehension abilities (assessed through pre-testing using TEA)) suggested there was little if any statistical advantage to the 20-item format:

Candidate	Score from 10	Score from 20
Bulgarian	10	19
French	10	18
Bulgarian	9	19
Bulgarian	9	17
Polish	9	17
Bulgarian	8	17
Colombian	8	16
Bulgarian	8	16

Italian	8	16
Italian	8	15
Polish	8	14
French	7	15
Italian	7	13
French	6	15
Russian	6	13
Polish	6	13
Brazilian	6	12
Italian	6	12
Colombian	5	12
Brazilian	5	11
Italian	5	9
Colombian	4	7
Russian	4	6
Brazilian	2	5
Colombian	2	4
Russian	1	3
Russian	1	3
Russian	1	2
Russian	0	1
Russian	0	0

The candidates would not have been awarded a different overall comprehension score and, although content validity could be improved through operationalising the 20-item task, the negative practical implications outweighed the potential positives.

References

Messick, S. (1994). *The interplay of evidence and consequences in the validation of performance assessments*. Educational Researcher 10, 9-20.

Skehan, P and Foster, P (1997) The *influence of planning and post-task activities on accuracy and complexity in task-based learning,* Language Teaching Research, Volume 1/3, pp 185-211

Underhill, N (1987) *Testing Spoken Language* Cambridge: CUP

Wigglesworth, G (1997): *An investigation of planning time and proficiency level on oral test discourse*, Language Testing, Volume 14/1, pp 101-122