



Checkpoint

LANGUAGE ASSESSMENT FOR STUDENT
SELECTION AND ADMISSIONS

Structure, platform and scores

July 2022

Contents

1. Introduction	2
2. Assessment description	2
3. Assessment platform	2
4. Assessment structure.....	3
5. Scores	4
5.1 ICAO and the CEFR	4
5.2 The Checkpoint traffic light system	4
5.3 The traffic light system and the CEFR	5
5.4 Listening, reading and speaking scores	5
5.4.1 Task fulfilment	6
5.4.2 Pronunciation	6
5.4.3 Language-in-use.....	6
5.4.4 Fluency.....	6
5.4.5 Coherence and cohesion	7
5.5 Checkpoint score reports.....	7
5.6 Using the score report	7

1. Introduction

Checkpoint is an English Language Proficiency (ELP) assessment designed to measure the English language skills required for successful English-medium flight training. Checkpoint is owned and operated by Latitude Aviation English Services Limited (UK).

At Latitude, we are committed to providing quality language training and assessment products and services, and to helping our stakeholders to understand and use them. This document provides information on Checkpoint's structure, administration platform and scores. The document is intended primarily to help decision-makers and admissions officers at airlines, flight schools, colleges and universities decide if Checkpoint meets their language assessment requirements, but it may be of interest to other stakeholders in flight training such students, student sponsors, English language instructors, aviation assessors and training managers.

2. Assessment description

Checkpoint is a specific-purpose web-based assessment of ELP designed for student pilot selection and admissions. Checkpoint is designed to be used:

- Before knowledge and aptitude assessments conducted in the medium of English;
- Before or after knowledge and aptitude assessments conducted in the students' mother tongue; and/or
- At the end of an English language training course.

Checkpoint scores are designed to align to the Common European Framework of Reference (CEFR) and can be used to:

- Determine student language proficiency for entry to ab-initio flight training programmes; and
- Identify student language training needs.

3. Assessment platform

Checkpoint is administered via a specific-purpose computer-based language testing platform developed and operated by Owl Testing Software, Pittsburgh, USA (www.owlts.com).

The Owl Test Management System (TMS) is an extremely versatile and flexible platform that allows for the administration of large-scale high stakes computer-based language testing programmes. The Owl TMS centralises test content and data and manages user access to the system according to pre-defined roles allowing:



- Latitude to create, administer and monitor assessments and assess candidate performance; and
- Latitude's customers and candidates to use Checkpoint from any location with a stable internet connection.

Owl's clients include:

- The National Board of Certification for Medical Interpreters
- The College of Staten Island / City University of New York
- The Information and Communications Technology Council
- Colombian Ministry of Education
- Yale, Cornell and Columbia Universities

With thousands of test takers worldwide each year, the Owl TMS has a proven track record in reliable and robust language test delivery and management.

4. Assessment structure

NOTE: Detailed assessment familiarisation videos for candidates are available on the Latitude website.

Part 1: Listening (approximately 40 minutes including introduction, task instructions and example items)

Task	Discourse type	Task time (minutes)	Speakers	Text length		Number and type of scored items
				Words	Minutes	
1	Informal student–training centre staff dialogue	11'	2+	1000 (+/- 100)	5-6	8 x 4-option MCQ (Answer the question / Complete the sentence)
2	Informal student-student dialogue	11'	3+	1000 (+/- 100)	5-6	As above
3	Formal training: instructor monologue with some instructor-student interaction	11'	1+	1000 (+/- 100)	5-6	As above

Part 2: Reading (approximately 40 minutes including task instructions and example items)

Task	Discourse type	Task time (minutes)	Text length	Number and type of scored items
1	Extract from ICAO Doc 4444: Procedures for Air Navigation Services	7	300 (+/- 50)	4 x MCQ (True, false or not given)
2	Extract from FAA Aeronautical Information Manual	7	300 (+/- 50)	4 x 4-option MCQ (Complete the sentence)
3	Extract from UK Air Accident Investigation Branch incident report	12	600 (+/- 100)	8 x 4-option MCQ (Answer the question / Complete the sentence)
4	Extract from industry journal on aviation training, safety and management	12	600 (+/- 100)	8 x single-option MCQ (Paragraph matching)

Part 3: Speaking (approximately 25 minutes including task instructions)

Task	Title	Task description	Response preparation?	Response (minutes)
1	Interview	Candidates read, listen to and answer 3 general aviation-related questions	No. Candidates respond to the questions as they are presented	4.5
2a	Storyboard	Candidates describe a scenario related to aviation or flight training captured in an illustrated storyboard	Yes. Candidates have one minute to look at the storyboard before responding	1.5
2b	Storyboard follow-up	Candidates respond to written follow-up prompts related to the topic presented in the storyboard	Yes. Candidates have one minute to look at prompts before responding	2
3	Presentation	Candidates listen to and give a report on a presentation at flight school. Candidates may make notes if they wish	Yes. After the presentation, candidates have one minute to prepare before responding	3

5. Scores

5.1 ICAO and the CEFR

Unlike most aviation English tests, Checkpoint does not measure language proficiency according to the ICAO Rating Scale. This is because the ICAO Rating Scale is an inappropriate measure of language proficiency for ab-initio aviation students for two key reasons:

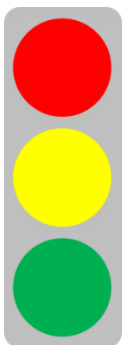
1. *The ICAO Rating Scale addresses only spoken language (speaking and listening); it does not address reading ... skills¹.* As reading is a skill crucial to successful ab-initio flight training, measurement of student language proficiency according to the ICAO Rating Scale is necessarily under-representative of the language skills required for successful ab-initio flight training.
2. *The sole object of ICAO language proficiency requirements is aeronautical radiotelephony communications².* Students on entry to professional flight training are very unlikely to possess working knowledge of flight operations or have experience with standard radiotelephony (RT) communications. Therefore, measurement of student language proficiency using tests designed to meet the ICAO language proficiency requirements is a threat to both test fairness and the validity and reliability of language test scores.

In seeking a more valid scale of measurement for Checkpoint, Latitude conducted research³ into student ELP requirements and the suitability of the Common European Framework of Reference (CEFR) for language assessment in the context of ab-initio flight training. In summary, this research involved linking aviation instructor's perceptions of the minimum levels of ELP required by students to CEFR listening, reading and speaking tasks and associated descriptors. The results showed that:

- The CEFR contains descriptors of ELP that are relevant to the context of ab-initio flight training; and
- CEFR B2 describes a minimum entry-level of ELP for English-medium flight training.

5.2 The Checkpoint traffic light system

Checkpoint scores are reported using a traffic-light system as follows:



Red: CEFR B1 or below. Language is likely to be an obstacle to successful flight training for candidates that score red in any part of the assessment. We recommend that candidates who score red in any part of the assessment undergo 200+ hours of language training before beginning flight training.

Yellow: CEFR B1-B2. Candidates that score yellow in any part of the assessment may encounter language-related difficulties during flight training. We recommend that candidates who score yellow in any part of the assessment undergo 25-200 hours of language training before beginning flight training.

Green: CEFR B2+. Candidates that score green in all parts of the assessment are unlikely to encounter language-related difficulties during flight training.

¹ ICAO document 9835, *Manual on the implementation of Language Proficiency Requirements*, Section 4.5.5.a.

² Ibid, section 3.2.7.

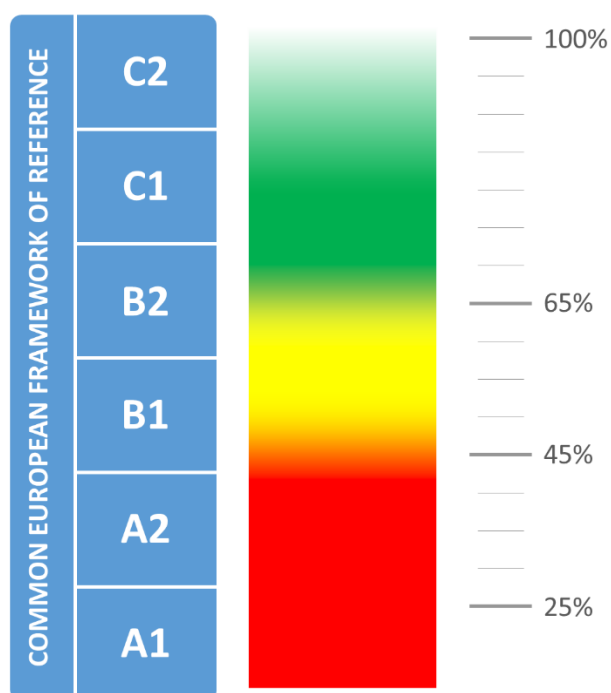
³ A full discussion of the issues associated with using the ICAO Rating Scale for measuring student ELP, along with presentation of the research summarised above, can be found in Emery, H. (2016) *Aviation English For The Next Generation* in Enright, A. and Borowska, A. (eds.) (2016) *Changing Perspectives on Aviation English Training*, Studi Naukowe 29, University of Warsaw.

5.3 The traffic light system and the CEFR

Checkpoint point scores are designed to align with the Common European Framework of Reference. This means that, in each of the Checkpoint listening, reading and speaking assessments:

- Candidates who score red may be able to perform the tasks described at the A1 and A2 levels, but have a poor chance of performing the easier tasks described at the B1 level.
- Candidates who score yellow have a good chance of performing the easier tasks described at B1 and, depending on scores, may have a good chance of performing the harder tasks at B1 and a reasonable chance of performing the easier tasks at B2.
- Candidates who score green have a good chance of performing the easier language tasks described at B2 and, depending on scores, may have a reasonable to good chance of performing language tasks at C1 and C2.

The diagram below shows how Checkpoint scores are designed to align to the CEFR. The percentages correspond to listening and reading assessment scores where 45% and 65% represent cut-scores for the red-yellow and yellow-green levels respectively.



5.4 Listening, reading and speaking scores

Checkpoint listening and reading assessments are scored automatically. Candidates receive a traffic light score and a percentage score for each of the listening and reading assessments. During the speaking assessment, the candidate's voice is recorded by the computer for subsequent rating by Latitude's raters. Spoken performance is rated according to the Checkpoint rating scale for speaking. The rating scale is comprised of three levels – red, yellow and green (arranged vertically) and five criteria – *task fulfilment*, *pronunciation*, *language-in-use*, *fluency* and *coherence and cohesion* (arranged horizontally). Raters award each candidate a level in each criterion. The candidate's overall speaking score is the lowest of any score in the five criteria.

5.4.1 Task fulfilment

Task fulfilment focuses on how well the candidate addresses the requirements of the speaking tasks.

In task one, the candidate responds to three general questions related to aviation. Task fulfilment in task 1 relates to the degree to which the candidate's responses address the topics raised in the questions, and the level of detail, reasons and examples that the candidate provides as evidence to support their ideas.

In task 2, the candidate provides a narrative based on a sequence of pictures and responds to follow-up prompts related to the pictures. Task fulfilment in task 2 relates to:

- The extent to which the candidate formulates a coherent narrative;
- The quality and accuracy of the descriptions of the visual information in the prompts; and
- The degree to which the candidate's responses to the follow-up prompts address the topics raised, and the level of detail, reasons and examples that the candidate provides as evidence to support their ideas.

In task 3, the candidate provides a report based on a presentation at flight school. Task fulfilment in task 3 relates to:

- The candidate's selection and presentation of information central to the presentation;
- The accuracy of the information included in the candidate's report;
- The extent to which the candidate:
 - Highlights or emphasises the importance of information; and
 - Provides personal commentary on the presentation.

5.4.2 Pronunciation

Pronunciation focuses on how well the candidate can produce the features of the English sound system and the extent to which control of these features assists or impedes raters' understanding of the candidate. These features include:

- Production of individual vowel, diphthong and consonant sounds;
- Pronunciation of words with the correct syllable stress;
- Rise and fall of voice pitch (intonation) to show meaning, for example, certainty, emphasis, query, digression and conclusion; and
- Control of word stress, cadence and pausing to organise speech into meaningful chunks and to indicate the beginning, middle and end of units of speech.

5.4.3 Language-in-use

Language-in-use focuses on the range, control and precision of the candidate's lexical and grammatical resource. Rating language-in-use involves identifying:

- How well the candidate's lexical and grammatical resource allows them to express themselves both spontaneously and with planning in relation to the task prompts;
- The flexibility and variety of the grammatical forms the candidate uses, and how repetitive the structures are;
- The range and precision of lexical resource and how repetitive the candidate's vocabulary is; and
- The extent to which errors with lexis and grammar impede understanding.

5.4.4 Fluency

Fluency focuses on how much language the candidate produces and how smooth and consistent the candidate's speech flow is. Rating fluency involves identifying:

- The speed of the candidate's speech flow (tempo);
- The length of turn the candidate is able to produce, or the ability of the candidate to 'keep going'; and
- The extent of pausing, hesitation, repetition and self-correction.

5.4.5 Coherence and cohesion

Coherence and cohesion focuses on the ability of the candidate to organise and arrange their discourse so that ideas are presented as logical, coherent text. Rating coherence and cohesion involves identifying:

- The extent to which the candidate organises their language both at the sentence level and across longer turns;
- The extent to which candidates employ linguistic devices such as referencing, substitution, discourse markers and other forms of textual cohesion; and
- The variety, appropriacy and accuracy of cohesive devices.

5.5 Checkpoint score reports

Test users receive a score report for each assessment session. Below is a sample assessment report.

#	FIRST NAME	LAST NAME	LISTENING	READING	SPEAKING	SPEAKING				
						TF	P	LIU	F	CC
1	Wang Qiang	Wong	20.83	33.3	R	R	R	R	R	R
2	Tomas	Schmit	62.50	66.67	Y	G	Y	Y	G	G
3	Rashed	Ali Aish	54.17	47.75	Y	Y	Y	Y	Y	Y
4	Wang Ping	Lui	62.50	70.33	Y	G	Y	G	G	Y
5	Fedhel	Talahi	37.50	41.67	R	R	R	Y	Y	R
6	Claude	Corichon	72.92	82.80	G	G	G	G	G	G
7	Miguel	Serra	45.83	54.17	Y	Y	Y	Y	Y	Y
8	Chanchoio	Chaiprasit	67.50	77.80	G	G	G	G	G	G
9	Nguyen	Ahn Dung	58.33	62.50	Y	Y	Y	Y	Y	Y
10	Maxim	Vakorin	41.67	54.17	R	R	Y	R	R	R

5.6 Using the score report

The assessment report comprises:

1. Traffic light scores in each of the listening, reading and speaking assessments. This allows decision makers to view candidate performance 'at-a-glance'. This is particularly useful when selecting or admitting a small number of candidates from a large population. For example, if the purpose were to select two candidates from the ten candidates above based on ELP, a decision maker could quickly identify candidates 6 and 8 as the most proficient.

2. Percentage scores for the listening and reading assessments and scores by criteria for the speaking assessment. These more granular-level scores indicate performance within the red, yellow and green levels. These scores are particularly useful when selecting or admitting candidates from a smaller population where finer distinction between candidate ability needs to be made. For example, if the purpose were to select four candidates from the ten candidates above based on ELP, one might:

- Reject candidates 1, 5 and 10 (as they scored red in some or all parts of the assessment)
- Select (in order of preference):
 - Candidates 6 and 8 (as they both scored green in all parts of the assessment)
 - Candidates 2, 4 (as they scored a mix of green and yellow)
 - Candidate 9 in preference to candidates 3 and 7. Although all three candidates scored yellow in all parts of the assessment, candidate 9's scores in listening and reading were significantly higher in the level than candidates 3 and 7.