Applied Linguistics Papers 27/1: 2023, 63–78

# Best Practices for Test Construction based on Test of English for Aviation Personnel (TEAP)

**Olena PETRASHCHUK**
University of Warsaw, Aviation Communication Research Centre
E-mail: o.petrashchuk@uw.edu.pl,

**Anna P. BOROWSKA**
University of Warsaw, Aviation Communication Research Centre
E-mail: a.borowska@uw.edu.pl,

**Abstract:** The paper presents best practices necessary for constructing mandatory tests for aviation/aeronautical purposes. Based on the TEAP test, we will show an example of the test form and content in accordance with the ICAO recommendations. We will analyse all the components and talk about their functions. The TEAP test has been used in Ukraine since November 2007. The Test specifications were piloted on 1114 Ukrainian ATCOs during the years of 2007–2009. The TEAP has been authorised by Ukrainian Civil Aviation Authority as a test to be used in Ukraine for licensure purposes of both controllers and pilots. In this paper, the language construct is considered as a benchmark for all aspects of ICAO language proficiency requirements implementation through aviation personnel teaching and testing. According to the *Dictionary of Language Testing* (1999), a 'construct' is defined as an ability or sets of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores. The Test practices will be presented on a large-scale format including examiners and raters' functions which should be viewed as crucial factors to provide the best test qualities – validity, reliability and practicality. Therefore, the TEAP specifications will be presented with the reference to the ICAO Chapter 6 provisions and the roles of examiners/raters as part of the testing system.

**Keywords**: Test of English for Aviation Personnel, Aeronautical English, test design, specialist language assessment, proficiency test

## Introduction

In aviation industry professional communication plays a crucial role, particularly during air-to-ground interactions. According to the studies on the role of language in aviation accidents, nearly two thirds of the accidents showed contributing role of language proficiency in radiotelephony communication (ICAO 2010: 1.2.8–1.2.7). It became a serious ground to initiate new language requirements for pilots and controllers introduced in March 2008. Radiotelephony communication is mainly realised by a deliberately invented code called a standardised phraseology which is characterised by specific features and the restricted register. In cases when phraseology is not sufficient, mostly in non-routine or emergency situations, plain language is required to be used in order to minimise misunderstandings between a flight crew and a control station.

Due to strong safety issues, plain English should be learned appropriately and assessed accurately as recommended by ICAO Doc 9835, Circulars No. 323 and No. 189. For this

purpose, language proficiency tests, also defined as high-stake tests, are used (ICAO 2010: 5.2.10, Ch. 6). Accordingly, English is internationally recognised common language for communication in aviation. Though local languages may be used within the territory of an ICAO individual member state, English standardised phraseology and plain English are required to be used for international flights.

Any test construct might be a challenge for test designers. Moreover, a language high-stake proficiency test for aviation licensure purposes seems to be a complex task due to the fact that the particular tasks should be accurately designed and carefully piloted in order to meet ICAO requirements. There are several factors that make language proficiency testing for licensing requirements a case of exceptionally high-stakes testing. Inadequate aviation language testing can result in either serious safety gaps or have highly negative social and economic consequences (ICAO 2010: 6.2.2.1). The paper aims at sharing good practices regarding aviation English test development in compliance with ICAO recommendations.

## 1. Statement of Purpose

TEAP design results from a large-scale and long-term work based on a theory and practice of language testing in general. In order to define a conceptual approach to the test design ICAO fundamental principles were considered (ICAO 2010: 6.2). The main aim of good proficiency language testing is to provide objective and reliable measurement of the language proficiency level related to the beyond test language use. In case of aviation personnel testing, it means that test-takers should demonstrate the performance appropriate to the language requirements at their workplaces.

Language communication via radio is of two-fold nature. On the one hand, the radiotelephony exchange participants should use standardised phraseology which is a code based on English in order to avoid misunderstandings. On the other hand, they should be able to use English in a natural conversation including all the technical aviation terms necessary, but at the same time, often under stress due to emergencies, they should try to avoid misunderstandings by applying more efforts, i.e. communication strategies, to communicate appropriately (A. Borowska 2017). Therefore, the best approach to test design seems to be a communicative one: "In an aviation context, proficiency testing should establish the ability of test-takers to effectively use appropriate language in operational conditions" (ICAO 2010: 6.2.5.4).

It should be noted that the communicative approach to language test design became popular in the mid-1980s. L. Bachman and A. Palmer (1996: 4) include, among fundamental principles of language testing, the need for a correspondence between language test performance and language use. This test quality was named 'usefulness': "in order for a particular language test to be useful for its intended purposes, test performance must correspond in demonstrable ways to language use in non-test situations." Another advantage of the communicative approach and usefulness of language test is the emphasis on strategic competence which is defined as "the ability to employ communicative strategies to compensate for breakdowns as well as to enhance the rhetorical effect of utterances" in the process of communication (H. D. Brown 2004: 10). Thus, the usefulness of language test in aviation context should include ratability of a speech sample. The test task types and test administration are expected to provide rateable speech samples of unrehearsed spontaneous language performance to be assessed against the ICAO Rating Scale for licensure purposes (ICAO 2010: 6.2.8.4).

Furthermore, communicative language testing presents challenges for test designers firstly because of search of real-world tasks, and secondly, due to the difficulty to validate the tasks to be included into the test. In aviation the contexts for such tasks are varied and the sampling of tasks for any assessment procedure needed to be validated by what language users actually do with the language. Another specific feature of language test for aviation purposes is the fact that real-world tasks should agree with radiotelephony format and content. It means that exchanges between controllers and pilots should be made in a standard language developed by ICAO based on aviation procedures, and provide the norms for worldwide communication in commonly occurring situations of air navigation. The procedures are subject to specific conditions, including speech tempo. Therefore, it is required that radiotelephony language should be concise in order to assure precise and unambiguous communication. The same rules apply for emergencies when plain English use is recommended and expected. Following the communicative approach to test design, the test tasks should reflect the working environment of the test-takers (e.g. ground-to-air communications) and, on the other hand, elicit more of natural (plain) language to meet the descriptors of language proficiency proposed by ICAO for licensing of aviation personnel (ICAO 2010: 4.6). It is obvious that the language testing for aviation purposes, with the reference to professional context and ICAO recommendations, should be concerned with the authenticity of test tasks to provide performance-based assessment (C.J. Weir, 1990).

Generally, performance-based assessment of language involves oral production, written production, open-ended responses, integrated performance, group performance and other interactive tasks (H. D. Brown, 2004: 10–11). According to ICAO (2010), the aviation language test should focus on two major skills – speaking and listening, since these are skills crucial for radiotelephony communication characterised by non-visual voice only message exchanges via radio. So, oral production and integrated performance can be appropriate test task types for aviation language testing. It should be also noted that interactive tasks are necessary for aviation context because the assessment is supposed to measure test-takers language samples, including their ability to react, i.e. to respond, to request or to clarify.

A valid test is to be designed in order to match the construct as well as the content being taught. The recommended and appropriate type of test seems to be a proficiency language test, which is not linked to any training course and, therefore, cannot be fully prepared on the basis of learning/teaching materials. Moreover, proficiency tests are suitable for licensing purposes in the aviation community because licensing plays a critical role in the safety of aviation operations (ICAO 2010: 6.2.5.5). An interview is considered to be one of the best test tasks to assess speaking skills. Communicative approach defines performance-based assessment which provides interactive assessment procedure. The test-taker is required to listen to another interlocutor and to respond appropriately. If care is taken in the test design process, tasks can approach the authenticity of real-life language use (H.D. Brown 2004: 11).

Based on theoretical grounds mentioned above, a model of aviation language test system has been suggested. The model of useful aviation English testing (UAET) below (Figure 1) presents test related components so that to provide relevant test system development in order to meet all major criteria of best quality language assessment for aviation purposes. All the components are necessary to make the testing system reliable and in accordance with the ICAO recommendations.

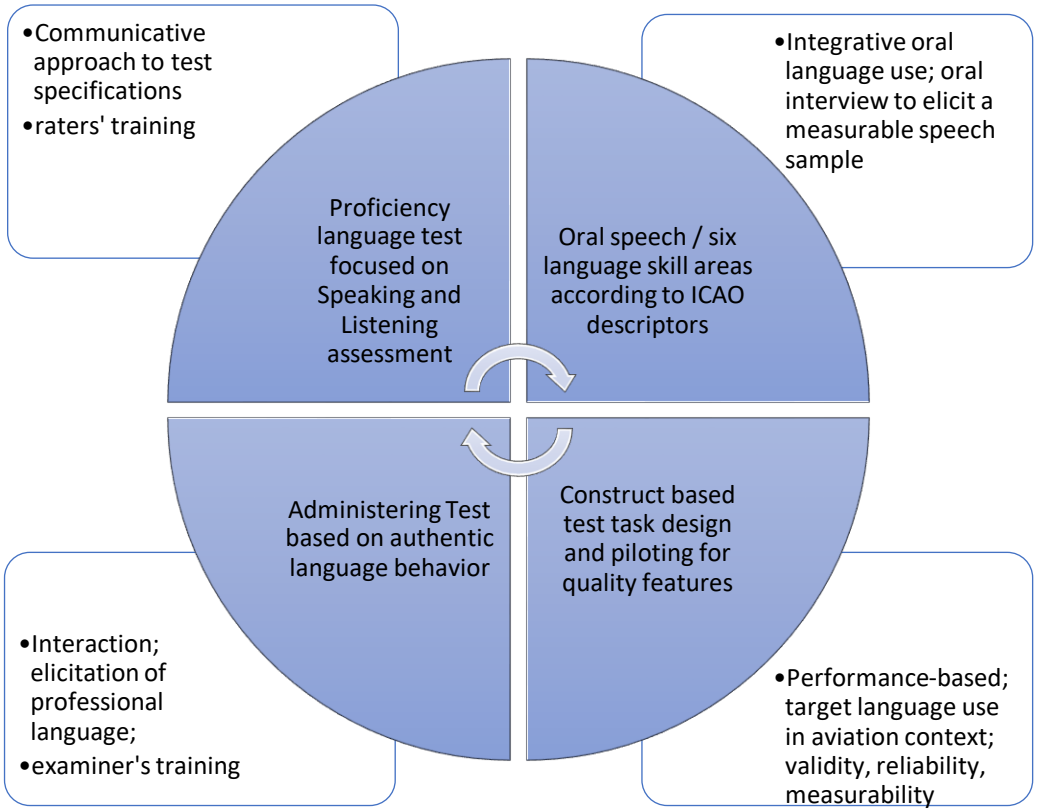


*Figure 1. Model of useful aviation English testing (UAET).*

The model above was created and consequently followed in order to design the testing system named Test of English for Aviation Personnel (TEAP). Following the principles of good practices and complying with the ICAO designated critical features of an appropriate testing system in aviation context, TEAP provides proficiency language tests for speaking and listening relevant to high-stake aviation language proficiency measurement in the broader context than ICAO standardized phraseology use (ICAO 2010: 6.3.2.4). The TEAP results are ratable against ICAO Rating Scale descriptors.

## 2. TEAP specifications

### 2.1. Test purpose

The Test of English for Aviation Personnel (TEAP) is a high-stake proficiency test aimed to assess speaking and listening proficiency of air traffic controllers and pilots in accordance with each component of the ICAO Language Proficiency Rating Scale (ICAO 2010) and the holistic descriptors in SARPS Annex 1. (ICAO 2010: 6.3.2.1, Attachment A to Annex 1). The TEAP assesses the overall ability of the aviation personnel to use plain English language in the context of operational aviation communication at Level 3 (Pre-Operational), Level 4 (Operational) and Level 5 (Extended). The achievement in the TEAP at Operational Level 4 or Extended Level 5 may be used as evidence of speaking and listening proficiency for licensing purposes in fulfilment of ICAO Annex 1 Requirements (ICAO 2010: 6.2.5.2).

Further, the TEAP is appropriate for licensing purposes and/or for purposes of periodic re-evaluation of language proficiency. The TEAP does not correspond to any training curriculum or training course. It is not possible to prepare directly for the TEAP, either. The TEAP assesses the test takers' ability to use plain English in aviation context. The target group is professional air traffic controllers and pilots who are license holders. The research in the form of questionnaires and direct interviews with aviation personnel was conducted during the year 2005. Initially, it dealt with air traffic controllers speech samples. Later, during the period of 2008–2010, the content was revised for the variety of candidates and availability of various speech samples (flight crew members, university graduates, helicopter pilots, PPL pilots, etc.).

### 2.2. Test construct

The language test construct is considered as *a hypothesized ability or mental trait* which cannot necessarily be directly observed or measured (ICAO 2010). The TEAP has been developed as a direct test of oral English language proficiency based on the constructs of language and communicative competences applied for aeronautical radiotelephony communication (ICAO 2010: 3.3, 3.4). The competencies are assessed through observable oral language performance in the areas of fluency, comprehension and interaction (ICAO 2010: 2.3). For assessment purposes, such language performance is recorded and stored as oral speech samples.

The TEAP construct is defined within the context of the aeronautical communication unique features which identify conditions and constraints, mental contexts of interactants, mode of language activities and type of task texts (ICAO 2010: 3.2.2). So, the test construct in speaking is defined as an individual's ability to use plain English as a natural language in a spontaneous, creative and non-coded manner. The plain English is a non-phraseology language used by participants in radiotelephony communications when standardised phraseology is not appropriate (ICAO 2010: 6.2.8.4). The test construct in listening is defined as an individual's ability to comprehend details of a message/information being received through an aural channel with and without background noise. The background noise simulates difficulty of comprehension which might be a case in voice-only settings of real-world aeronautical communication. The construct of overall language proficiency to be measured is based on the ICAO profile of a proficient speaker (ICAO 2010: 4.5.3).

The candidates receive an overall rating equivalent to that of the lowest rating achieved in any of the six skills presented in the ICAO Rating Scale (ICAO 2010: 4.5.11, 6.3.2.11). The language proficiency of the TEAP candidates is identified within different constructs in accordance with six skills: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, Interaction.

*Pronunciation* is a skill that aims at adjusting an individual's phonological competence in accordance with observable correction/feedback on success of communication. The candidates who lack this skill demonstrate pronunciation not intelligible enough for safe aeronautical communication (e.g. due to strong accent contaminated by native language communication); so they may be assessed at Pre-Operational Level 3. The candidates who demonstrate the pronunciation intelligible enough for safe aeronautical communication (e.g. though influenced by native language, but not interfering with the ease of comprehension) may be assessed at Operational Level 4. The candidates displaying the pronunciation slightly influenced by native language, but fully intelligible for safe aeronautical communication or

nearly English native pronunciation may be assessed at Extended Level 5 (ICAO 2010: 4.6.2).

*Structure* is a skill that shows the appropriate application of grammar in potential aviation communication. The candidates who lack a good command of basic grammatical structures and sentence patterns and therefore are not able to communicate efficiently due to errors resulting in misunderstandings may be assessed at Pre-Operational Level 3. Those who demonstrate a good command of basic grammatical structures and sentence patterns creatively used for effective communication (though infrequent errors may occur especially when attempting complex structures or in an unexpected turn of events the communication is overall effective) may be assessed at Operational Level 4. The candidates who have consistent control of basic grammatical structures and sentence patterns and use of complex structures (though possible mistakes in complex structures sometimes may occur, but they do not interfere with meaning) may be assessed at Extended Level 5 (ICAO 2010: 4.6.3).

*Vocabulary* is a skill that refers to an individual's lexical competence, at range and with appropriate speed of access to the vocabulary required in a given situation (ICAO 2010: 2.6.3). The candidates who demonstrate limited and sometimes inappropriate vocabulary as well as the frequent inability to clarify or paraphrase unknown words and as a result inaccurate communication may be assessed at Pre-Operational Level 3. The candidates demonstrating sufficient accuracy and range of appropriate vocabulary and ability to paraphrase unknown words and as a result accurate and efficient communication may be assessed at Operational Level 4. The candidates displaying sufficient accuracy and range of appropriate vocabulary with possible phrasal verbs or idiomatic language and ability to efficaciously communicate easily paraphrasing unknown words may be assessed at Extended Level 5 (ICAO 2010: 4.6.4).

*Fluency* is a skill addressing an individual's ability to produce unrehearsed speech at an appropriate pace and with a guidance of a listener through the oral discourse (ICAO 2010: 2.6.3). The candidates demonstrating oral speech with limited stretches of language and frequent silent pauses inappropriate for effective communication may be assessed at Pre-Operational Level 3. The candidates demonstrating production of language stretches at fairly constant speech rate and ability to provide effective communication (though with limited discourse markers/connectors or occasional slowness due to natural language processing) may be assessed at Operational Level 4. The candidates demonstrating speech rate and oral discourse organisation approaching natural fluency for effective communication (though not much higher than 100 words/min as recommended by ICAO) may be assessed at Extended Level 5 (ICAO 2010: 4.6.5).

*Comprehension* is a next component skill that addresses an individual's ability to recognise and understand speech which may vary in discourse and topical complexity, degree of detail and speed of understanding, delivery style and conditions of reception (ICAO 2010: 2.6.3.1). The candidates demonstrating comprehension limited to routine communications conducted in optimum conditions (they fail to understand even after seeking clarification) may be assessed at Pre-Operational Level 3. The candidates demonstrating ability to ultimately comprehend routine and unusual communications applying, if needed, clarification strategies (especially when confronted with a linguistic or situational complication) may be assessed at Operational Level 4. The candidates displaying high degree of detailed accuracy in their understanding of both standard and non-standard communications even when encountered with linguistic or situational complication may be assessed at Extended Level 5 (ICAO 2010: 4.6.6).

*Interaction* is the sixth component skill addressing an individual's ability to respond quickly and appropriately, to take conversational initiatives and volunteer new information as well as to resolve misunderstandings and respond to an examiner's/interlocutor's feedback (ICAO 2010: 2.6.3). The candidates lacking ability to interact concisely and efficiently and demonstrating mis-/non-understandings resulting in frequent breakdowns of communication may be assessed at Pre-Operational Level 3. The candidates demonstrating ability to initiate and maintain exchanges and in unexpected situations to communicate the fact of possible misunderstandings (by checking, asking for confirmation, clarifying) may be assessed at Operational Level 4. The candidates displaying ability to interact with high levels of comprehension and fluency and good control over the conduct and direction of the conversation may be assessed at Extended Level 5 (ICAO 2010: 4.6.7).

### 2.3. Test format

The TEAP specifications are compliant with the fundamental constraints specific to the context of the ICAO language proficiency testing requirements (ICAO 2010: 6.2.8), namely:
a) test focus – on speaking and listening proficiency;
b) test content – relevant to work roles of pilots and air traffic controllers, 'work-related topics/context and routine work situation', not standard phraseology based;
c) test tasks – similar to real-life activities, related to aviation operations/radiotelephony communications, e.g. questions and answers, problem-solving exchanges, etc.

The distinction between communicative competence and actual performance means that the TEAP should contain tasks that require actual performance as well as tasks or item types that measure language knowledge. Such task types would allow test takers to demonstrate their knowledge in action (ICAO 2010: Doc 9835, 6.2.6.1). Regarding the importance of real-life simulation through the test tasks, an oral interview is used for the TEAP speaking component. The oral interview aims at measuring speaking ability and it is close to real life interaction. Therefore, it is widely used in testing speaking in order to make its results the most reliable. In general, an oral interview in the form of one-to-one conversation is viewed as the best method to test oral proficiency because the process is 'a realistic reflection of real-life conversation' (A. Lazaraton 2002). The TEAP candidates are presented with a variety of topics (through questions and prompts) and have an opportunity to perform at their best. Following a standardised examiner/interlocutor interview frame, a specially trained examiner/interlocutor facilitates the interaction, asks questions, instructs and sets the tasks.

The TEAP has its test blueprint, namely a detailed plan that provides a basis for developing an entire test. The test blueprint designed for the TEAP has been adapted from the blueprint for a useful test described by L. Bachman and A. Palmer (1996). It includes the following elements: 1) the characteristics that pertain to the structure of the test: the number of parts/tasks, time allotment for each part, the salience of parts/tasks, the sequence of parts/tasks, the relative importance of parts/tasks, number of tasks per part, and 2) task specifications: purpose, definition of construct, setting, time allotment, instructions, characteristics of input and expected response, scoring method.

The TEAP blueprint has been designed in accordance with the TEAP purpose and the TEAP constructs mentioned above. It was followed by the TEAP item creators and test developers as a framework for reference to develop the TEAP prototype trialled in June 2006. Since then, the blueprint has been reviewed and upgraded as required:

<table>
<tr><td colspan="2">TEST FORMAT</td></tr>
<tr><td>Number of sections/ parts<br><br>Doc 9835, 6.3.2.1 ; 6.3.2.4</td><td>Two components – Listening Comprehension and Oral Interaction<br>Listening Test consists of 3 sections and is delivered in a pencil and paper format.<br>Speaking Test consists of 3 parts and is delivered in the format of oral face-to-face interview with one voice-only task.</td></tr>
<tr><td>Number of tasks/ questions per section/ part</td><td>Listening Test: Section 1 includes 2 tasks, points 1-14.<br>Section 2 includes 1 task, 15-20 points.<br>Section 3 includes 1 task, points 21-30.<br>Speaking Test (Oral Interview):<br>Part 1 Warming-up: guided by examiner’s/interlocutor’s questions.<br>Part 2 Task A – Conversation guided by a printed prompt on a cue card.<br>Task B – Voice-only task followed by comments based on a cue card.<br>Part 3 Conversation: guided by 2 visual prompts and follow-up questions.</td></tr>
<tr><td colspan="2">Listening Test / in total: 23 min</td></tr>
<tr><td>Purpose<br>SARPS Annex 1, Attachment A to Annex 1</td><td>To assess aural comprehension ability against the language proficiency criteria contained in ICAO Rating Scale and the ICAO Holistic Descriptors of operational language.</td></tr>
<tr><td>Time allotment<br>Up to 23 min</td><td>Pre-test phase:<br>Introductory instruction – 2 min<br>Filling in the answer sheet cover page – 2 min<br>Reading the answer sheet – 2 min<br>Test-taking phase: writing answers in the answer sheet – 15 min
<table>
<tr><td>Section 1</td><td>Section 2</td><td>Section 3</td></tr>
<tr><td>Task 1 1-9<br>Task 2 10-14</td><td>Task 3: 15-20</td><td>Task 4: 21-30</td></tr>
</table>
After-test phase: checking/refining the answers – 2 min</td></tr>
<tr><td>Instructions<br>Doc 9835, 6.3.5.3 -4</td><td>The general instruction for the candidates at the time of test administration is presented orally by the test administrator in front of the candidates just before the test is taken.<br>The task instruction is not a part of the input to which the candidates are expected. The task instruction is printed in English in the answer sheet and is followed by an example.<br>The instruction language is simple.<br>Text type/length: short radiotelephony simulated and authentic transmissions featuring interacting speakers; narration by a single native speaker.</td></tr>
</table>

| | |
|---|---|
| Characteristics of input/ texts | |
| Task/ Item types<br><br>Doc 9835, 6.2.8.9 ; 6.3.2.5 | *Labelling* of statements with letters P (pilot) or C (controller)<br>*Completion* of the statement by putting down one-two words or numbers<br>*Multiple choice* tasks with 3 options indicated by A, B, C (circling the chosen letter)<br>*Multiple choice* tasks with 3 options – 'Right'– 'Wrong'– 'Doesn't say' – indicated by A, B, C (circling the chosen letter)<br>The comprehension checks are presented in the same order as the details of information are sequenced in the recordings. |
| Expected response | In Sections 1 and 2 a candidate is expected to:<br>o understand gist and details of a radiotelephony communication between an English speaking (with at least two different accents) controller and pilot in routine and non-routine situations and in the optimum and sophisticated (background noise) conditions. In Section 3 a candidate is expected to:<br>o identify specific factual information contained in a narration of a single speaker.<br>Corrections are permitted. |
| Scoring method | Objective method: the candidate's answers are compared with key/right answers; number of correct answers is calculated as a sum of scores. The scoring of Listening Test is performed by a marker using arithmetic calculation of correct answers. The results of the scoring are indicated in a designated place on the answer sheet cover page.<br>Each correct answer weights 1 score, maximum score for the test is 30, which is equivalent to 100%.<br>Range of scores correlated with levels 3, 4, 5 of ICAO Rating Scale:<br>Level 3 – 60% -75% (score 18-22)<br>Level 4 – 76% -90% (score 23-26)<br>Level 5 – 91% -100 % (score 27-30)<br>There is no pass mark. |

| Speaking Test /in total: 22 min | |
|---|---|
| Purpose<br>Doc 9835, 4.6 | To assess speaking ability (to use plain English in aeronautical communication) against the language proficiency criteria contained in the *ICAO Rating Scale* |
| Definition of constructs<br><br>Doc 9835, 6.2.8.4 | The test construct in speaking is defined as an individual's ability to use plain English as a natural language in spontaneous, creative and non-coded manner. The plain English is a non-phraseology language used by participants in radiotelephony communications when standardised phraseology is not appropriate. |
| Time allotment<br><br>Up to 22 min. | Pre-interview phase:<br>Checking ID, introductory instruction, filling in and signing the interview protocol – 2 min;<br>Part 1 – 4-5 min; Part 2 – 7-8 min; Part 3 – 6-7 min |

| | |
|---|---|
| Instructions<br><br>Doc 9835, 6.3.5.3 | *The general instruction* for the candidates at the time of test administration is presented orally by the examiner.<br>*The task instruction* is not a part of the input to which the candidates are expected to respond directly.<br>The instructions are provided in English orally by an examiner/interlocutor and prompts are printed in English on the cue cards (Part 2). |
| Characteristics of input<br><br>Doc 9835, 6.2.6; 6.3.6<br><br>Task types<br><br>Doc 9835, 6.2.7; 6.2.8.9; 6.3.2.5 Doc 9835, 6.3.2.7; 6.3.2.9 | Channel/format: live verbal interaction (dialogue), longer response (monologue).<br>The interview is recorded. It is a one-to-one *oral interview* guided by an examiner's/interlocutor's questions and supported by printed, aural and visual prompts.<br>Rationale:<br>In order to elicit a sample of unrehearsed spontaneous speech an examiner/interlocutor applies interlocution strategies to guide a candidate's oral discourse to become measurable against the *ICAO Rating Scale*. Thus, the interview is semi-structured and relatively flexible in its content that allows for adaptation and customisation appropriate to the level of oral discourse and operational environment of an individual candidate.<br><br>The interview consists of three parts:<br>*Part 1* – a warm up segment guided by an examiner's/interlocutor's questions;<br>*Part 2* – Task A – the candidate is given a printed card and is asked to speak on a specific topic (unusual/emergency situation) related to his/her area of professional activity; the candidate's speaking is guided by 3 prompts printed on a card and, if needed, by an examiner's/interlocutor's questions; the candidate is given 1 minute to focus on the topic and prompts. Task B – voice-only segment: the candidate listens to short exchanges on unusual/emergency situation and is asked to comment on it; the candidate's response is guided by a cue card.<br>*Part 3* – the candidate is asked to describe two photos with an aviation situation related to their area of professional activity; the photos are given in sequence and may be guided by an examiner's/interlocutor's questions.<br><br>Language communicative functions assessed are specified in ICAO Doc 9835, Appendix B 'Language of Aeronautical Radiotelephony Communications': The candidates are supposed to use plain English on various topics that are related to radiotelephony communications without replicating radiotelephony communications specifically. In particular, these are the topics of non-standard/emergency situations. |
| Expected response<br><br>Doc 9835, 6.2.7.7; 6.3.2.8-10; 6.3.2.1 | The candidates gain no credit for rehearsed language.<br>A candidate is expected:<br>In Part 1 – to interact with an examiner/interlocutor by responding immediately, accurately and appropriately to an examiner's/interlocutor's questions on common and work-related questions; to speak briefly about themselves and their work place/area of aviation operations. The candidates are also given an opportunity to speak spontaneously on issues related to their personal life and their professional activities.<br>In Part 2 – to demonstrate an oral discourse guided by printed and aural prompts on an unusual/emergency situation. The candidates are expected to speak spontaneously and accurately on issues related to non-standard situations according to the topic in the cue cards/sound file.<br>In Part 3 – to demonstrate the oral discourse guided by visual prompts.<br>The candidates are expected to give detailed description of a photo and to supplement the description with their comments/opinion/ on the situation/event depicted. |

| | |
|---|---|
| | Throughout the whole interview for *minimum operational level* the candidates are expected to be able to:<br>- speak fluently on common, concrete and work-related topics,<br>- produce oral speech in stretches of language at an appropriate tempo,<br>- confirm, negotiate, clarify if needed,<br>- understand a communicator by responding immediately, informatively and appropriately,<br>- use fillers, connectors appropriately,<br>- take turn to maintain/initiate interaction,<br>- paraphrase if needed,<br>- continue to communicate effectively in unexpected turn of events or when confronted with linguistic or situational issue,<br>- keep fluency and use clarification strategies,<br>- communicate accurately with proper pronunciation, intonation, range of vocabulary and good control of basic grammatical structures,<br>- provide information relating to present, past or future events,<br>- provide information concerning necessity, feasibility, capacity,<br>- express agreement/disagreement, appreciation, opinions,<br>- describe events, people, places, sequence of events, procedure or process,<br>- compare, explain, justify, assess, present, instruct, advise, approve, permit, etc. |
| Scoring method<br><br>Doc 9835, 6.3.4.1-3 | Subjective method of scoring is based on individual judgment of a specially trained rater. Speaking component is assessed through rating oral speech samples against six descriptors of language profiles in ICAO Rating Scale. The assessment is conducted by at least two qualified raters. They use standard protocols/assessment sheets to record data of inaccuracy, errors, etc. They are guided by the instructions on the rating process. |
| Final score for the test as a whole<br><br>Doc 9835, 4.5.11; 6.3.5.2; 6.3.2.11 | The candidates are awarded a rating within one of the six levels contained in the *ICAO Rating Scale* in each of its discrete features of language: pronunciation, structure, vocabulary, fluency, comprehension and interaction. The candidates receive an overall rating equivalent to that of the lowest rating achieved in any one of the six features. |

*Table 1. TEAP blueprint.*

## 3. Methods and materials

### 3.1. Baseline study

For the purposes of a new test, a baseline study was conducted in order to know the language related attitudes of aviation personnel in English non-native country regarding the new ICAO language requirements. The investigation was conducted in the period of 2005–2006 as an informal pilot survey of the TEAP target group. It included questionnaires, interviews and test results analysis. The test target population samples were selected at random from the aviation personnel available (representatives of the CAA, the State Flight Academy, the Ukrainian State Air Traffic Service Enterprise training centre, Ukrainian Helicopters Airline, South Airlines of Ukraine, Antonov Aviation Research Bureau). The respondents were asked to fill out an open question questionnaire in Ukrainian. The interviewees were mainly

senior air traffic controllers, instructors, recent graduates of the State Flight Academy, representatives of air traffic trade unions in Ukraine, pilots and flight engineers, pilot-instructors and aviation business managers.

The survey results indicated that nearly 96% of the target population wanted to learn English to meet ICAO requirements by March 5, 2008. If, at that time, there had been an English language test for licensure endorsement about 90% of them would have taken the test. Approximately 80% expected being tested for memorised English language knowledge. More than 95% had no previous experience of being interviewed in English for the purpose of direct assessment of their spoken English proficiency. Only 1/3 was interested in English for daily communication. 2/3 of the respondents demonstrated their oral speech ability at ICAO level 2. Most of them were highly motivated to learn plain English. They clearly wished to learn English for practical use in their professional activity, but they hesitated to identify the role of plain English alongside the standard phraseology. The survey revealed the general attitude to English proficiency as rehearsed coded speech behaviour. Moreover, the results of the survey showed that there was a definite need for a systematic study of plain English as a means of communication in situations when the standard phraseology is not sufficient. Furthermore, the test task types and test administration should provide ratable speech samples of unrehearsed spontaneous true language performance to be assessed against the *ICAO Rating Scale* (ICAO 2010: 6.2.8.4).

Based on the study results, the ICAO recommendations and with the reference to UAET model presented above, the objectives of the TEAP design and development have been defined as follows:

- to develop a valid, reliable and practical tool for assessment of a candidate's ability to operate in English in the context of aeronautical radiotelephony communications in international civil aviation;
- to develop an assessment tool reflecting actual language use;
- to develop an assessment tool acceptable to both pilots and controllers;
- to develop an assessment tool for English language proficiency in the context of aviation in a manner acceptable to the Ukrainian civil aviation authorities, professional registration bodies and employers;
- to provide assessment mechanism appropriate to the high-stake testing of English for aviation personnel for licensing purposes;
- to provide testing of English for aviation personnel and quality of testing service delivered at international standards;
- to provide testing of English for aviation personnel that is fair to all candidates.

As a result, the TEAP is a high-stake test for aviation language (ICAO 2010: 3.2). Following the ICAO requirement for such types of tests, the TEAP quality has been verified.

### 3.2. Piloting for quality

ICAO recommendations emphasise that 'fairness' should be the overriding concern of high-stakes test developers. In language testing, fairness is interpreted in terms of validity and reliability. Practicality is a third fundamental test consideration (ICAO 2010: 6.2.3.3). Based on our own studies and long-term practice of testing aviation personnel, one more quality feature has been suggested for aviation language test – measurability (O. Petrashchuk 2013). Therefore, high-stake aviation language tests should be evaluated in terms of their validity, reliability, practicality and measurability based on collected data:

a) *Validity* indicates the degree to which a test measures what is supposed to measure. To this end, testers should gather and provide evidence to support the conclusions that are made about an individual's English language proficiency based on the individual's performance on a test;
b) *Reliability* refers to the stability of a test. Evidence should be provided to prove that the test can be relied upon in order to produce consistent results. Although no test will achieve a perfect reliability, one should look for tests with the highest possible reliability;
c) *Practicality* refers to the balance between the resources required to develop and support a test (including the funds and the expertise) and the resources available to do so;
d) *Measurability* of speech samples against the ICAO Rating Scale indicates the degree to which test-takers language elicited and performance demonstrated is appropriate to the ICAO Rating Scale descriptors of language proficiency expected within 6-skill areas: pronunciation, structure, vocabulary, fluency, comprehension, and interaction.

In order to prove the quality of TEAP, the test was piloted during the years 2006-2009 on samples of 1114 professional aviation personnel working all over Ukraine. The piloting was specifically focused on evaluation of two test components – Listening Test and Speaking Test. Both quantitative and qualitative methods of analysis were applied. Below there are two brief reports illustrating some procedures undertaking for piloting purposes.

### 3.3. Report on TEAP validity

The TEAP validity was evaluated at both social and cognitive levels. Social aspect of the validity was provided by modeling 'co-operative' communication through one-to-one interview. Cognitive aspect of the test validity concerned the extent to which the cognitive processes employed by candidates were similar to those in the real-world communication.

As far as a construct validity of TEAP is concerned, it was verified through analysis of the oral discourse derived from speech samples of oral interviews at levels 3-5. The analysis was based on a qualitative approach to the validation of oral language tests (A. Lazaraton 2002). The samples were selected at random at each level of actual performance. In total 30 samples of levels 3 and 4, and 9 samples of level 5 were selected and analysed. The analysis was aimed at collecting evidence of how well the test performance can be interpreted as a meaningful measure of the TEAP constructs identified for each language profile of the ICAO Rating Scale for the levels 3, 4 and 5. Minimum level of the construct validity was set high because of flight safety issue. The trialled samples of oral discourse proved high degree of validity by consistency with the constructs to be measured and measurement results of corresponding language performances at levels 3, 4, 5. The TEAP construct validity was provided by trained examiners/interlocutors who had to elicit a ratable speech sample, thereby increasing the likelihood of drawing valid inferences about the construct from the candidates' performance (see L. Bachman/ A. Palmer 1996).

The term *face validity* refers to the test surface credibility or public acceptability. This aspect was studied and ensured during the test design and piloting. Face validation was proved by feedback obtained from selected aviation personnel working in Ukraine and involved into discussion of training and testing needs. The personnel consisted of 150 subject matter experts, top management, the CAA experts and representatives of ATC Trade Union of Ukraine, Ukrainian State Aeronavigation Provider, state ANTONOV airline. One of the outcomes of high level of TEAP face validity was the test endorsement by the Ukrainian CAA for the licensing purposes.

### 3.4. Report on TEAP reliability

Reliability refers to the degree that the test produces consistent and fair results (ICAO 2010: 6.3.3.1). In order to verify the TEAP reliability a few evaluation methods were used. In the period of 2006–2007 the TEAP Listening and Speaking components were verified by parallel testing method and a test-retest method. Items of the Listening Test were analysed statistically; the oral interview procedure was verified by trial-marking sessions. All the trialling methods were applied on a large-scale sampling. The parallel testing is given here as a single example due to restriction of page space. According to this research method, there were two similar tests chosen – TEAP and IELTS. IELTS original tasks from Listening and Speaking sections were selected. The selected candidates sat for both tests during two days. The parallel testing was carried out in 10 cities of 6 regions of Ukraine and involved 914 candidates in the listening test and 898 candidates participated in speaking test (16 candidates refused to sit for the oral interview). 3 examiners/interlocutors, 15 test administrators and 8 raters conducted the test.

All candidates sat for the tests on the same single day. The results obtained in both test systems in listening sections were calculated statistically. The interviews in both tests were recorded and rated blindly by two raters. The documentation of oral interviews included 79 examiner protocols and 57 protocols of rating oral interviews. In total, nearly 75 hours of interviews were recorded. The total score was awarded in a different way (according to each test system rules). The total score for IELTS spoken performance was an average one, and the total score for TEAP spoken performance was the lowest one out of six. As a result, a number of samples awarded the same level of language proficiency. The reliability of TEAP speaking component was identified through equivalency of levels measured by different scales:

| English language proficiency equivalency | | TEAP reliability coefficient |
|---|---|---|
| IELTS speaking component | TEAP speaking component | |
| Intermittent user | L 1 | .89 |
| Extremely limited user | L2 | .87 |
| Limited user | L3 | .78 |
| Modest user | L3+ | .76 |
| Competent user | L4 | .91 |
| Good user | L4+ | n/a |
| Very good user | L5 | n/a |
| Expert user | L6 | n/a |

*Table 2. Equivalency of levels measured by different scales.*

In conclusion, it turned out that the levels of language proficiency assessed by similar tests, though designed for different purposes, in the most of samples correlated positively. The average reliability coefficient for TEAP oral interviews was 0.84 which is considered high for speaking tests.

## 4. Training examiners and raters

It has been indicated earlier that 'the ideal method for testing oral proficiency is a face-to-face conversation, so the interviewing process is a reasonably close, if not an absolutely realistic reflection of real-life conversation' (J.L. Clark 1980). As mentioned above, the TEAP Speaking Test is conducted as an oral proficiency interview that is a 'special case of conversation that are examiner-directed, it is obvious why conversational techniques such as the interview constitute a pragmatic speaking task' (J. Oller 1979). Oral proficiency interview consists of 15–25-minute structured conversation during which an examiner tries to elicit from the examinee a rich sample of speech by using a variety of questions types and covering a wide range of topics and situations (L. Bachman 1990). The main focus in TEAP examiners/interlocutors training is to teach the staff to apply appropriate interviewing/interlocution strategies to elicit from test-takers the target language construct. TEAP system suggests initial and then annually recurrent examiner/interlocutor training (ICAO 2010: 6.3.8).

The TEAP rating procedure is conducted by objective and subjective methods. The objective method is applied for the listening part results which are scored by matching them with the answer key. The range of listening test scores, corresponding to Levels 3, 4, 5 of the ICAO Rating Scale, has been identified as follows: Level 3: 60%-75% of correct answers (scores 18–22); Level 4: 76% -90% of correct answers (scores 23-26); Level 5: 91% – 99 % of correct answers (scores 27–30). It should be noted that the assessment of comprehension to which listening test results contribute is not to the detriment of the assessment of interaction (ICAO 2010: 6.3.2.6).

On the other hand, the subjective method of assessment is applied to the TEAP speaking part (oral interview) where rating process is based on human judgment. The rating is conducted after the test using a recording of a given sample of the test-taker's language performance. Reports on the candidates' language performance are based on an examiner/interlocutor protocol and on a rater protocol/assessment sheet. The recordings are assessed blindly by at least two remote raters (ICAO 2010: 6.3.4.2). Therefore, the main priority for appropriate raters training is on teaching them to compare the sound of language performance with the description of the language performance according to the ICAO Rating Scale, and to identify full or partial coincidence or lack of the coincidence between what is heard and what is described. The main challenge for raters is to pay particular attention to the boundaries between the Levels 3 and 4 as well as Levels 4 and 5. In order to achieve the main goals of training of both examiner/interlocutors and raters their training is based on the test language constructs. The language construct-based training enables to synchronise their approaches with language performance as a subject matter – one (interlocutor) elicits what then is assessed by another one (rater) (O. Petrashchuk 2017).

## Conclusion

Language related human factor in aviation is a hot point for aviation personnel training and testing. ICAO new language requirements focus on a communication instrument – plain language that is required in non-routine emergency situations. Flight safety should be regarded as a priority for a language proficiency test design and administration. Hence, the following recommendations are presented:

- The test should be a proficiency test of speaking and listening;
- The test should be based on the ICAO Rating Scale and holistic descriptors;

- The test should test speaking and listening proficiency in a context appropriate to aviation;
- The test should test language use in a broader context than in the use of ICAO phraseologies alone.

The suggested UAET above allowed to incorporate into TEAP specifications all ICAO recommendations for aviation test design in order to conduct accurate language assessment. The article emphasises the challenging process of designing TEAP. Each aviation testing system should include language and human related elements which should provide validity and reliability of the test tasks and measurability of speech samples. Last but not least is the fact that the fairness of assessment is a crucial point of high-stake tests. Human-related element of TEAP is presented by examiner/interlocutor and rater's assessment. Hence, TEAP is a performance-based procedure that is supposed to elicit and then to assess language proficiency of pilots and controllers in accordance with the ICAO Rating Scale.

## References

Bachman, L. (1990), *Fundamental considerations in language testing*. Oxford University Press: New York.

Bachman, L./ A. Palmer (1996), *Language testing in practice*. Oxford University Press. New York.

Borowska, A.P. (2017), *Avialinguistics. The Study of Language for Aviation Purposes.* Peter Lang: Frankfurt am Main.

Brown, H.D. (2004), *Language assessment: principles and classroom practices*. Longman.

Clark, J.L. (1980), *Towards a common measure of speaking proficiency,* (in:) J.L. Frith (ed.) "Measuring Spoken Language Proficiency". Georgetown University Press: Washington DC, 35–49.

Cushing, S. (1994), *Fatal Words: Communications Clashes and Airplane Crashes*. University of Chicago Press.

ICAO (2010), *Manual on the Implementation of ICAO Language Proficiency Requirements.* Doc. 9835 AN/453. Montréal: ICAO, 2d ed.

Lazaraton, A. (2002), *A Qualitative Approach to the Validation of Oral Language Tests.* (in:) M. Milanovic/ C. Weir (ed.) "Studies in Language Testing" No. 14. UCLES: CUP.

Oller, J. (1979), *Language Tests at School*. Longman: London.

Petrashchuk, O. (2012), *Test of English for aviation personnel to meet ICAO language proficiency requirements*, (in:) "Proceedings of the National Aviation University" No. 3, 160–164.

Petrashchuk, O. (2013), *Aviation English Test Development*, (in:) "Proceedings of the National Aviation University" No. 4, 155–160.

Petrashchuk, O. (2017), *Construct based approach to examiner/rater training in aviation language testing.* (in:) M. Grygiel/ M. Rzepecka (ed.) "Specialist communication in Education, Translation and Linguistics", Uniwersytet Rzeszowski, 131–141.

Weir, C.J. (1990), *Communicative language testing*. Prentice Hall International: London.