

Trialling the ELPAC test

1. Trialling after July 2007

The **ELPAC** first final test versions were constructed in June and July 2007 and released on the **ELPAC** test server as planned - on 31 July 2007. **ELPAC** development will, of course, not stop with the release of the first test versions. New items and tasks will be written and trialled.

1.1. Trialling new Paper 1 items

At the end of the test each candidates will be encouraged to complete a short additional section. This is how new items will be trialled. It will take a maximum of 10 minutes and candidates will make their own decision whether to take the additional section or not.

The **ELPAC** system will employ an algorithm to ensure that each item is trialled on a representative sample of the target population. All items will be analysed, reviewed and updated by the **ELPAC** project team. Where necessary, items will be re-trialled. The construction of future test versions will follow the same criteria as the first test versions to ensure that the test versions are as parallel as possible.

1.2. Trialling new Paper 2 items

Trialling will be arranged separately from **Paper 1** and will be co-ordinated with **ELPAC** test users. All tasks will be analysed, reviewed and updated by the **ELPAC** project team. Where necessary, tasks will be re-trialled. The construction of the future test versions will follow the same criteria as the first test versions to ensure that the test versions are as parallel as possible.

2. Trialling during the ELPAC development stage

The **ELPAC** development stage was completed in July 2007 with the release of the first final test versions for both Test Papers. The trialling exercises were conducted in more than 20 European countries with over 800 controllers. Bio data from candidates and feedback comments from candidates, Test Administrators, **Paper 1** Markers and **Paper 2** Examiners were collected and used for the test / item / task analysis. A number of changes were implemented during the development stage, taking into consideration not only the statistical findings but also feedback comments.

Detailed information and also evaluation of the trialling exercises can be found under the menu item "**ELPAC** validation".

2.1. Paper 1

There were five trialling exercises (November 2005, March 2006, October 2006, March 2007 and May 2007), each trialling two test versions of **Paper 1**. All trialling exercises were followed by a thorough item / test analysis which included statistical findings and feedback from stakeholders (test takers, administrators, developers, etc).

The first test versions (triated in November 2005) had a slightly different format (7 test parts) and did not provide enough time for the test takers to read and answer test questions. That is why the test format was changed. The time to read and answer questions increased several times during the trialling exercises. All test versions trialled had high reliability. The versions trialled in May 2007 show a slightly lower reliability (yet still acceptable) but this seems to be caused by a smaller number of candidates.

Test version	Reliability (Cronbach's Alpha)	Number of test takers
Version A – November 2005	0.800	105
Version B – November 2005	0.870	96
Version C – March 2006	0.935	130
Version D – March 2006	0.940	128
Version G – October 2006	0.934	204
Version H – October 2006	0.929	204
Version L – March 2007	0.921	196
Version M – March 2007	0.923	196
Version P – May 2007	0.892	52
Version Q – May 2007	0.875	52

It has to be noted that not all test versions trialled consisted of new items - many items had to be re-trialled. Some items (and marking keys) were updated and the time to read and answer questions was increased. Certain audio files were amended. All this led to changes in the test and thus everything had to be trialled again.

Test items also went through standard setting where English language and ATC subject matter experts who were not involved in the item writing assigned levels of difficulty and provided more feedback on individual items.

The test developers then collated all information from trialling analyses and standard setting exercises and selected items which performed well. The final test versions were constructed according to the following criteria:

- item difficulty
- content
- length of individual snippets
- specific language functions assigned to test items

Parallel test versions were produced, taking into consideration all of the above listed criteria.

2.2. Paper 2

There were four trialling exercises (May/June 2006, October 2006, March 2007 and May 2007), each trialling several versions of **Paper 2**. All trialling exercises were followed by a thorough task analysis which included analysis of the submitted recordings and feedback from stakeholders (test takers, administrators, examiners, developers, etc).

Special care was given to the analysis of the recordings as the examiners received only limited training before each trial.

Task 1 provides a specific context to tower, approach or en-route controllers and was trialled in this format. Task 2 was trialled with controllers of all ATC functions and then, based on the outcomes, individual tasks were assigned to be suitable either for all controllers or for controllers of particular experience (tower, approach, en route).

All tasks were reviewed and all updates re-trialled. Then, a thorough analysis was conducted again. As with **Paper 1**, **Paper 2** tasks were also standard set. Again a group of English language and ATC subject matter experts not involved in task preparation was used.

Only after all tasks were trialled, re-trialled and standard set did the test developers collate information from all the analyses and construct the final test versions, making them as parallel in level of difficulty as possible.