

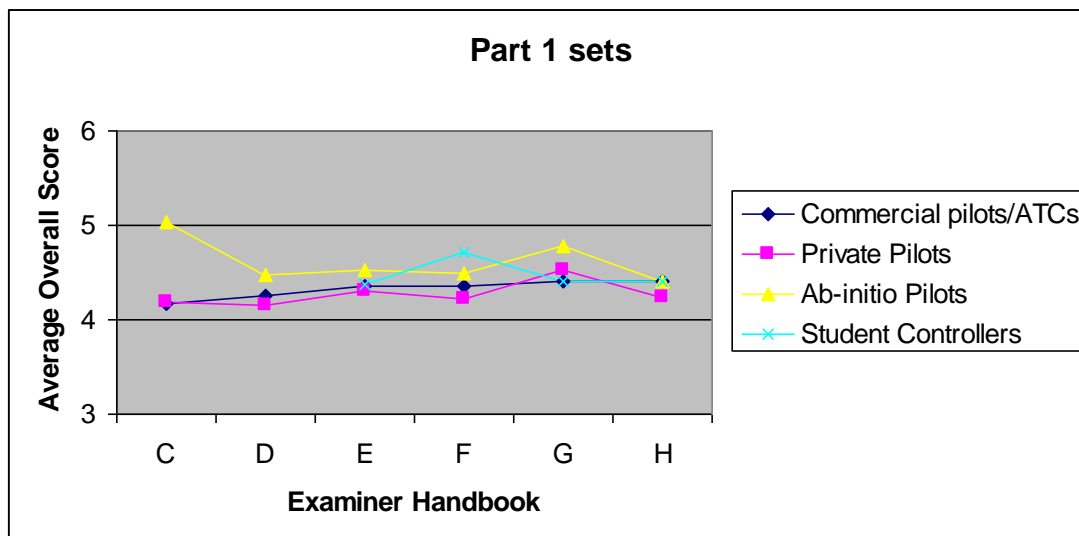
It is important to assess the performance of the different versions of TEA by analysing whether they performed in the same way across the TEA Examiner Handbooks. Significant differences in performance would suggest a degree of unreliability in the test. Each section of the test – Part 1, Part 2 and Part 3 – was studied individually. There was not enough data on Handbooks I & J, released in November 2011, for comparison.

Part 1

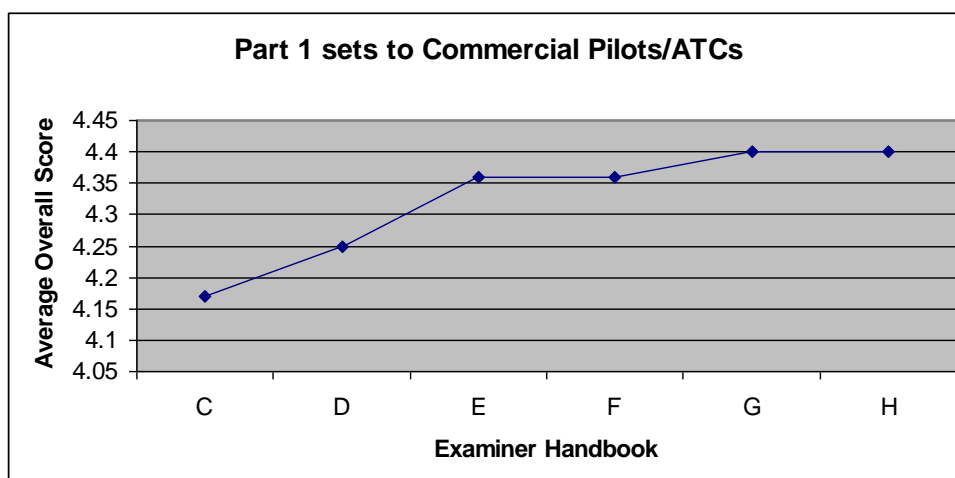
Part 1 has different question sets for different test populations. From TEA Version 2010, which went ‘live’ in April 2010, sets were added for Student Controller candidates giving 4 separate test populations in Part 1.

	Part	Handbook	Test Population	Average Overall Score
Original TEA	1	A/B	All	4.00
↓	1	C	Commercial pilots/ATCs	4.17
	1	C	Private Pilots	4.18
	1	C	Ab-initio Pilots	5.04
	1	D	Commercial pilots/ATCs	4.25
	1	D	Private Pilots	4.15
	1	D	Ab-initio Pilots	4.47
Version 2010	1	E	Commercial pilots/ATCs	4.36
↓	1	E	Private Pilots	4.30
	1	E	Ab-initio Pilots	4.53
	1	E	Student Controllers	4.38
	1	F	Commercial pilots/ATCs	4.36
	1	F	Private Pilots	4.22
	1	F	Ab-initio Pilots	4.50
	1	F	Student Controllers	4.71
	1	G	Commercial pilots/ATCs	4.40
	1	G	Private Pilots	4.53
	1	G	Ab-initio Pilots	4.78
	1	G	Student Controllers	4.40
	1	H	Commercial pilots/ATCs	4.40
	1	H	Private Pilots	4.23
	1	H	Ab-initio Pilots	4.41
	1	H	Student Controllers	4.40

The data above and the graphic below shows that the variation between Part 1 sets in different handbooks has little difference in overall score outcomes.



One would expect a general rise in scores as time progresses (as new handbooks are introduced) since the LPRs were introduced several years ago and, in the meantime, training and testing plans have been implemented, raising the standard of language proficiency generally. The results for Commercial Pilots/ATCs – the largest test population – show a general trend of improvement.



The average overall score for Handbook C (5.04) stands out as an outlier. Compared to its partner Handbook (D at 4.47), there appears to be a significant difference between the performance of the sets. On closer inspection of the data, however, it was found that Handbook C was used to test 156 Dutch candidates whereas Handbook D was used to test only 11 Dutch candidates. Given that the average overall score of Dutch candidates is 5.51, and that the testing of student pilots was limited in the early days of TEA testing (only 439 tests were delivered to ab-initio pilots with Handbooks C & D), it is easy to see how the data can become skewed by a 'batch' of testing in a country with a distinct proficiency level (high or low).

It is interesting to note that the average overall scores for ab-initio pilots is generally higher than for other test populations. This is to be expected since they are generally younger candidates who have grown up with a greater education and exposure to English than their older counterparts.

Part 2

Part 2 has different CDs in every Examiner Handbook.

	Part	Handbook	CD #	Average Overall Score
Original TEA	2	A/B	All	4.04
↓	2	C	9	4.20
	2	C	10	4.25
	2	C	11	4.56
	2	C	12	4.21
	2	D	13	4.30
	2	D	14	4.24
	2	D	15	4.24
	2	D	16	4.19
Version 2010	2	E	17	4.67
↓	2	E	18	4.65
	2	E	19	4.62
	2	E	20	4.61
	2	F	21	4.30
	2	F	22	4.28
	2	F	23	4.37
	2	F	24	4.38
	2	G	25	4.43
	2	G	26	4.40
	2	G	27	4.41
	2	G	28	4.43
	2	H	29	4.44
	2	H	30	4.40
	2	H	31	4.30
	2	H	32	4.13

It is positive to find that no test CD appears to overly benefit or penalise the candidates it was delivered to. Pre-testing aimed to achieve such results. Given that the average Comprehension score across all TEA tests is 4.6, it would appear that the CDs are not influencing the rating of candidates across different versions significantly. Since TEA

Version 2010 was operationalised, the batches of CDs (4 per handbook – *see Report 08 – Item Development & Version Content* for more details) have been constructed and trialled differently and that appears to have resulted in a more uniform performance within batches (see the results for CDs 25-28: 4.43, 4.40, 4.41, 4.43). See *Report 09 - Assessing the Reliability of test CDs during Development* for further data on test construction.

Part 3

There are 6 themes in Part 3 of every Examiner Handbook and research was conducted into the performance of each handbook in deciding overall scores.

	Part	Handbook	Part 3 Sets	Average Overall Score
Original TEA	3	A/B	1 - 12	4.03
↓	3	C	13 - 18	4.30
	3	D	19 - 24	4.25
Version 2010	3	E	25 – 30	4.35
↓	3	F	31 – 36	4.32
	3	G	37 – 42	4.42
	3	H	43 - 48	4.34

The Part 3 sets of different handbooks do not appear to help or hinder candidates significantly with the range of average overall scores from Part 3s of Version 2010 handbooks being 0.1 of a band score (4.32 to 4.42).