



**CaMLA English Placement Test (EPT)
Forms D–F: Development Report**

Daniel Walter
Jasmine Hentschel

September 2013

CONTACT INFORMATION

All correspondence and mailings should be addressed to:

CaMLA

Argus 1 Building
535 West William St., Suite 310
Ann Arbor, Michigan
48103-4978 USA

T: +1 866.696.3522

T: +1 734.615.9629

F: +1 734.763.0369

info@cambridgemichigan.org

www.CambridgeMichigan.org

Table of Contents

1. Introduction	1
2. History of the EPT	1
3. Revision Project.....	1
3.1. Test Design	1
3.2. Compiling the Pilot Test Forms.....	2
3.3. Pilot Testing Design	3
4. Pilot Testing Results.....	3
4.1. Test Taker Demographic Characteristics	3
4.2. Pilot Testing Descriptive Statistics	5
5. Structure of the New EPT	5
5.1. Content of the New EPT	5
5.2. Test Information Functions	6
6. Interpreting EPT Scores and Determining Cut Scores.....	7
References.....	7
Appendix A. Sample Items	8
Appendix B. Score Concordance Table from the Prior Versions of the EPT to the New EPT Forms.....	10

1. Introduction

The CaMLA English Placement Test (EPT) assesses general receptive language proficiency by measuring performance in listening comprehension, grammatical knowledge, vocabulary range, and reading comprehension and is aimed at language learners whose English language proficiency ranges from high beginner to low-advanced, or levels A2 to C1 on the Common European Framework of Reference (CEFR) (Council of Europe, 2001).¹ The EPT is designed to quickly and reliably group ESL students into different ability levels. In addition, the test is used by universities, colleges, language programs, and businesses all over the world to evaluate students' and employees' ability to use English. In 2012 CaMLA revised the EPT in response to feedback from test users and created three new forms. Launched in January 2013, each of the parallel new forms contains 80 multiple-choice questions and takes 60 minutes to complete. Scoring can be done with a punched stencil using the provided answer sheets, or purchasers may use their own answer sheets.

This report describes the development of the new CaMLA EPT, also known as EPT forms D, E, and F. It provides a brief history of the EPT, a description of the revision project, evidence for the validity of the new test forms, and advice on how to set cut scores.

2. History of the EPT

The EPT was originally created in 1972 to help determine the placement of incoming students in the six-level Intensive English Program (IEP) at the University of Michigan. All items for the EPT were selected based on statistics obtained from pilot testing on the multi-level IEP students and the test assessed test takers across a range of ability levels. In 1987, the EPT was made available for use at other institutions. Since then, it has been a well-established and popular product in a variety of social, educational, and occupational contexts.

3. Revision Project

The goal of the EPT revision project was to renew the content of the test while keeping the test format and score interpretation consistent so that users of older versions could move seamlessly to the new test forms.

3.1. Test Design

The EPT is typically used for placement into language programs or as part of the hiring process. Test users need quick turnaround of results; they also need to administer the test often and at short notice. In some contexts the test is used more than once with the same test takers – perhaps after a period of study. In addition to this flexibility in administration, test users need a robust, reliable instrument that is not time-consuming. The EPT meets these requirements: it is a paper-and-pencil test of 80 items that can be administered in 60 minutes. It is a test of general receptive language proficiency, assessing listening comprehension, use of English (vocabulary and grammar), and reading comprehension. The items are situated in a variety of language domains: educational, social, occupational, and personal. There are three unique forms, constructed so that they are parallel in content and in difficulty.

In order to provide a comprehensive measure of receptive language proficiency, a range of item types was included. Each item type targets different language interactions and contexts, enabling the test takers to demonstrate a range of receptive language skills (see Appendix A for examples of each item type).

¹ Descriptions of these levels can be found at:
www.cambridgemichigan.org/exams/general/cefr

Table 1: Description of Item Types

Language Skill	Item Type	Item Type Description
Listening	Listening question	Listen to a short question and select the best response from 3 answer choices
	Listening dialogue	Listen to a conversation between 2 speakers and answer a question about the exchange by choosing from 3 answer choices
Language in Use	Grammar	Read a short dialogic exchange between 2 speakers in which part of a turn has been omitted and select which of 4 answer choices best completes the exchange
	Vocabulary	Read a single sentence from which one word has been omitted and select which of 4 answer choices correctly completes the sentence
Reading	Sentence-level reading	Read a single sentence and then answer a comprehension question by choosing from 4 answer choices
	Reading passages	Read a passage and answer reading comprehension questions about the passage by choosing from 4 answer choices

Table 1 summarizes the item types. The listening section assesses the test takers’ ability to understand dialogic speech that varies in formality and also to understand the pragmatic force of single utterances in order to provide an appropriate response. The grammar section also assesses the test takers’ ability to understand and complete dialogic speech. The vocabulary section assesses the test takers’ vocabulary breadth and depth, sometimes expecting test takers to know secondary meanings of words. The reading section assesses both close reading of single sentences and the reading of longer texts for understanding and to draw inferences.

Table 2 shows the listening and reading skills assessed by the EPT.

Table 2: Listening and Reading Subskills Assessed by the EPT

Skill	Subcategory
Global	• Understanding main idea
	• Identifying speaker’s/writer’s purpose
	• Synthesizing ideas from different parts of a conversation/interview/written text
Local	• Identifying supporting detail
	• Understanding vocabulary
	• Synthesizing details
	• Recognizing restatement/paraphrase
Inferential	• Understanding rhetorical function
	• Making an inference
	• Inferring supporting detail
	• Understanding pragmatic implications

3.2. Compiling the Pilot Test Forms

In order to maximize the success of the pilot test forms, the project team constructed a preliminary item-bank with items that had been previously pre-tested (i.e. for which item-level statistics were available). The bank comprised items covering a range of item difficulties and content characteristics. These were revised to standardize their format, refresh the content, and improve distractors. The bank was then augmented with new items written in accordance with the EPT language construct. All items went through a three-stage review process that included two content review stages and a fairness and bias check using CaMLA’s Fairness Guidelines.

After the review of all items was complete, four different EPT pilot test forms were created. Each pilot test form was compiled using 91 items from the bank and 15 items from the original EPT Form A. The latter were embedded in each pilot test form to facilitate the equating of old and new forms. The forms were constructed to test language in various domains (educational, occupational, public, or personal), a broad range of topics, and several different item sub-skills. The distribution by item type for each pilot test form is shown in Table 3.

Table 3: EPT Pretesting from Item Breakdown

Item Type	Number of Items on Each Pilot Form ²	Total number of items piloted
Listening question	15	45
Listening dialog	15	60
Grammar	25	85
Vocabulary	25	85
Sentence-level reading	10	25
Reading passages	16	64 (8 passages)
Total	106	364

² This includes the EPT Form A link items.

3.3. Pilot Testing Design

The pilot testing phase enabled CaMLA to identify items that did not perform satisfactorily so that they could be removed from the item bank before compilation of the final test forms. Item-statistics from pilot testing were also used to ensure that the final EPT test forms would be at approximately the same level of difficulty.

Each pilot test form was administered in several locations. CaMLA recruited test centers with English language learners from a wide range of ability levels and first-language backgrounds to ensure that the pilot test population was representative of the EPT’s target test population; some were previous EPT users and others were interested in using the new EPT in the future. In addition to the new pilot test forms, the old EPT form A was administered at several pilot test centers. In order to ensure that the pilot test data were comparable across all administrations, all institutions were given an administration manual that specified how to give the test under standardized conditions. Statistics from administrations of Form A and from the 15 embedded Form A items on each of the four pilot forms were used to create the concordance table of scores for old and new versions of the EPT.

Test forms were piloted at 13 institutions across the United States and Canada. Demographic data, including sex, date of birth, and the native language of each candidate, was provided by the test takers.

4. Pilot Testing Results

As presented in Table 4, a total of 480 test takers participated in pilot testing of the EPT. This section provides their demographic characteristics and information about how they performed on the test.

Table 4: Number of Test Takers for Each Pilot Version

Form	Number of Test Takers	Number Who Also Took Form A
Form 1	109	-
Form 2	95	-
Form 3	205	40
Form 4	71	15

4.1. Test Taker Demographic Characteristics

Test taker first languages

The EPT pilot testing population comprised speakers of 29 different first languages (see Table 5). The range of test takers’ first languages indicates that the test taking population is broad enough to be generalizable to a wide variety of contexts.

Table 5: Native Languages of EPT Pilot Test Takers

Arabic	Hungarian	Punjabi
Bengali	Indonesian	Russian
Bulgarian	Italian	Sinhalese
Chinese	Japanese	Spanish
Creole	Korean	Tagalog/Filipino
Farsi/Persian	Malayalam	Thai
French	Mongolian	Turkish
Greek	Nepali	Ukrainian
Gujarati	Pashtu	Vietnamese
Hindi	Portuguese	

In Figure 1, the six most frequently occurring first languages (those with more than ten speakers in the test taking population) are represented. The remaining 23 languages comprise the ‘other’ category, which makes up 13.26% of the total test taking population. The figure shows that approximately 68% of test takers were speakers of Arabic or Chinese. Spanish, Korean, French, and Japanese were the other major language groups represented. According to the 2012 Open Doors report (Institute of International Education, 2012), these six languages represent 50.10% of the overall population that studied abroad in the US in 2011/2012. Since some educational institutions use the EPT for placement decisions for foreign students, this indicates the EPT pilot testing population is highly representative of those populations.

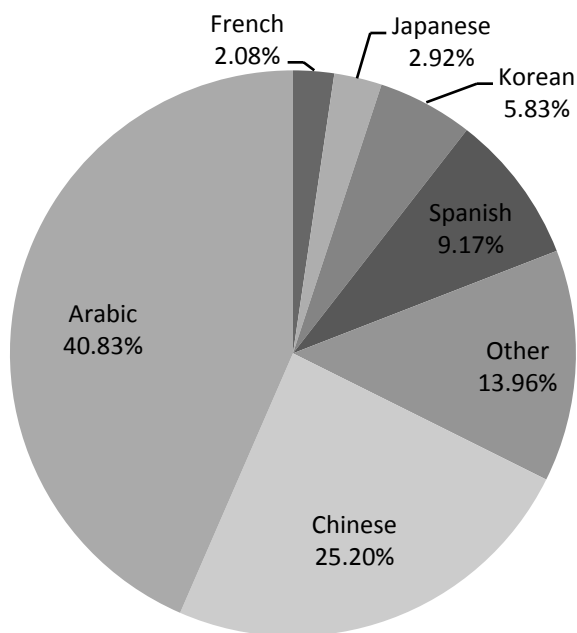


Figure 1: First Language of EPT Pilot Test Takers

Age Distribution

Table 6 shows the wide age distribution of pilot test takers, with peaks evident around the expected ages—17-19, when people typically enter undergraduate education, and 22-25, when people typically pursue graduate degrees.

Table 6: Age Breakdown of EPT Pilot Test Takers

Age Group	Number	% of Total
13–16	5	1.04
17–19	124	25.83
20–22	106	22.08
23–25	84	17.50
26–29	50	10.42
30–39	57	11.88
≥ 40	37	7.71
Missing data	17	3.54
Total	480	100.00

Gender Distribution

Table 7 below shows the distribution of test taker gender. Cases where information was not given were treated as missing data. The data shows that at least 52% of test takers were male and at least 34% were female. In many ESL contexts, it is common to find more female than male test takers yet, in this case, a majority of the pilot testing took place in ESL instruction courses in which the gender balance happened to be more heavily weighted towards men. In particular, the Arabic and Chinese test takers were primarily male. Though this is a slightly unusual gender distribution, there were still enough female test takers that the population can be considered representative.

Table 7: Gender Breakdown of EPT Pilot Test Takers

Gender	Number	% of Total
Female	166	34.58
Male	254	52.92
Missing data	60	12.50
Grand Total	480	100.00

Table 8: Descriptive Statistics by Pilot Form

Form	# of test takers	Min	Max	Avg	St. Dev.	St. Error	Reliability
1	109	5	98	50.81	19.40	4.51	0.95
2	95	32	103	74.04	17.40	4.23	0.94
3	205	11	96	53.57	17.45	4.49	0.93
4	71	24	104	69.75	19.97	4.20	0.96

4.2. Pilot Testing Descriptive Statistics

Table 8 presents the descriptive statistics for each pilot test form. Test takers varied in ability both within and across forms. The average scores for test takers who took forms 2 and 4 were higher than those who took forms 1 and 3, implying slightly different ability levels for the populations who took each form. However, higher and lower ability level test takers were found in the population that took each form. The reliability achieved by each pilot test form was excellent and the standard error estimates were acceptable given the generally low-stakes purpose of the test.

The pilot test data was analyzed using a 1 Parameter Logistic (or Rasch) Item Response Theory model. A primary benefit of the Rasch model is that it places test takers and items on a common scale, allowing for comparability of ability levels and item difficulties across the pretesting forms as well as providing a performance link to earlier versions of the EPT. The project team established the following parameters for identifying successful items: they had to fall between -2 and +2 in equated IRT difficulty and also had to achieve a minimum biserial correlation of 0.30.

Table 9 presents the success rate of the pilot items by item type. It shows that on average, 84% of the items met the parameters for success. This represents a very good success rate. The items that met these

Table 9: Pretested Item Success Rates

Item Type	Total	Successful	Rejected	Success Rate
Listening question	45	37	8	82%
Listening dialogue	60	51	9	85%
Grammar	85	65	20	76%
Vocabulary	85	72	13	85%
Sentence-level reading	25	22	3	88%
Reading Passage	64	55	9	86%

performance criteria were used to compile the three new CaMLA EPT forms D, E, and F.

5. Structure of the New EPT

5.1. Content of the New EPT

Each form of the new CaMLA EPT contains 80 items. Test takers are allowed a maximum of 60 minutes to complete all questions. Each correct answer carries equal weight within each section and there are no points deducted for wrong answers. The item distribution on the new EPT test forms is shown in Table 10 below.

Table 10: Item Structure of New CaMLA EPT Forms

Section	Item Type	Number of Items
Listening	Listening questions	10
	Listening dialogues	15
Grammar Vocabulary Reading	Grammar	20
	Vocabulary	20
	Sentence-level reading	5
	Reading passages	10
Total Number of Items		80

The new EPT forms were compiled using the successful pilot items. The IRT data allowed the project team to compile three new test forms of parallel difficulty, with items targeted across a range of difficulty levels. In addition to item difficulty, items were selected to achieve an even balance of domains, topics, and item sub-skills across the forms.

5.2 Test Information Functions

As stated in sections 1 and 2, the new EPT forms have been designed to provide meaningful information about test takers across a wide range of ability levels, typically high-beginner to low-advanced. Two concepts help in understanding how much information a test provides at a particular ability level: the Item Information Function (IIF) and the Test Information Function (TIF). The IIF is a graphic representation of the item response theory parameter estimates of one item. Test taker ability level is presented on the x-axis where a higher number signifies higher ability. The item information level is presented on the y-axis and provides an indication of how much information the item can provide about test takers at each ability level. What the IIF displays is how well the item is able to differentiate between test takers across the ability spectrum. At the ability levels near the center of the curve, the item is most informative and can most effectively place a test taker's ability level. The item provides less information about a test taker whose ability level is further away from that center because the item is either too easy or too difficult for him/her. An example IIF from the EPT is represented below in figure 3. It looks like a normal curve centered at the ability level of a test taker who is expected to have a 50% chance of answering the question correctly. It shows that this item tests best around an ability level of -1, but it still provides a substantial amount of information about test takers in the -3 to +2 ability level range.

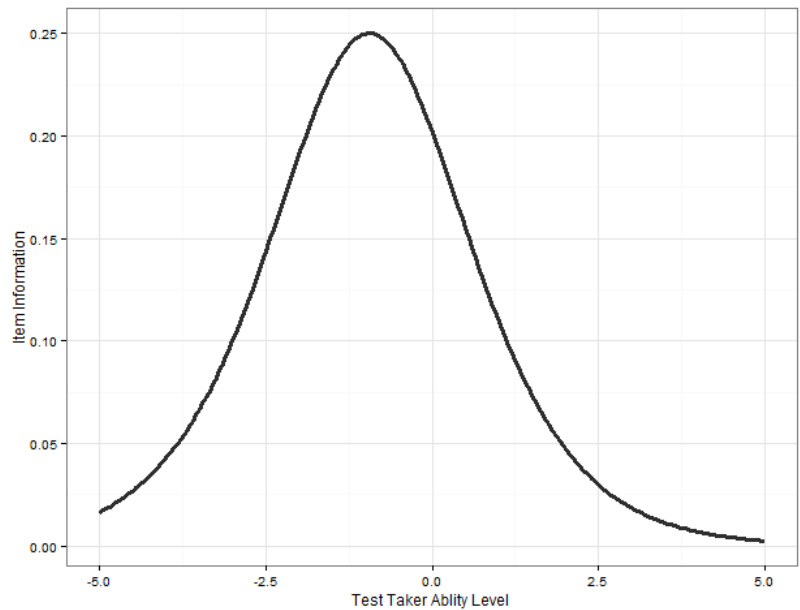


Figure 2: Example Item Information Function (IIF) for Item 4 of Form D

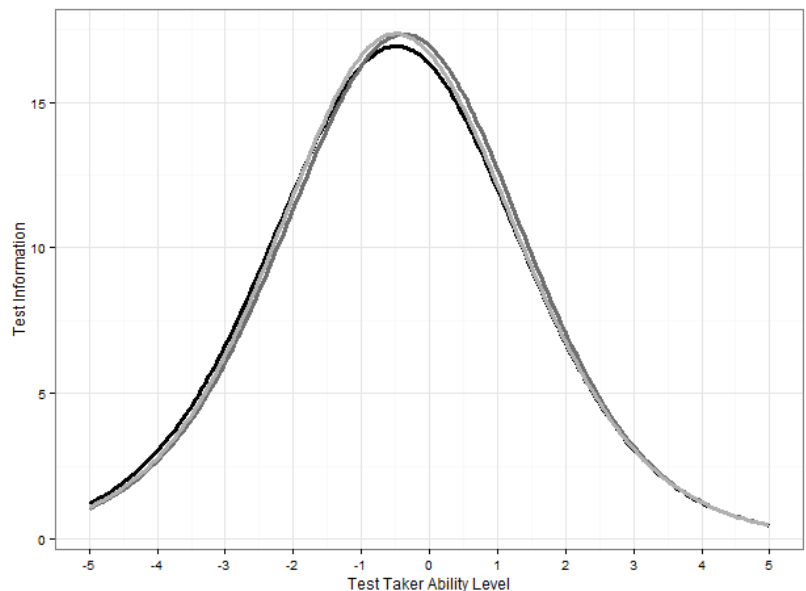


Figure 3: Test Information Functions for Forms D, E, and F

The TIF displays similar information to the IIF, showing the candidate ability levels most accurately measured by an entire test. The TIF is produced by adding the values of the IIF's at each candidate ability level. Figure 2 shows estimated TIF's for the new EPT forms, D, E, and F.

All three tests have a maximum TIF at the ability level -0.3. That means the test provides the most information about a test taker at that ability level, which corresponds to a score of 44 on forms D, E, and F. The graph indicates that all three forms are able to most precisely locate ability levels for candidates whose ability is in the -3 to +2.5 range. This corresponds roughly to the score range from 7 to 76 on forms D, E, and F of the new EPT. These results confirm that the EPT measures robustly across a wide proficiency range. The fact that the graph shows nearly indistinguishable TIF's for the three forms suggests parallel measurement across the three forms.

6. Interpreting EPT Scores and Determining Cut Scores

Test taker data from forms D, E, and F, as well as form A, were run through proprietary software that converts scores and probabilities into ability estimates for candidates. Ability estimates for each of the new forms were lined up against the ability estimates for Form A to produce equivalence tables showing how raw scores on forms D, E, and F correspond to scores on the previous EPT. The score correspondence table for all forms can be found in Appendix B below.

These concordance numbers are provided to help users familiar with the previous EPT to adapt their existing cut points to the new EPT forms. CaMLA strongly encourages organizations using EPT for the first time to make their own cut points based on their particular needs and program structure. Organizations will require different cut scores based on the proficiency of students and number of levels those students need to be streamed into. EPT users must determine which EPT cut scores are relevant in their context for the test to effectively classify students.

One strategy an organization might use for determining appropriate cut scores on the EPT is to administer the test to students whose ability levels in that context are already known (i.e. students who have just enrolled in a lower-intermediate level English language learning classes at a university). If the majority

of test takers obtained scores in the range of 35-42, for example, a decision could be made to use that EPT score range in the future to assign students to the lower-intermediate course. By examining the EPT scores obtained by students in context, a set of acceptable and marginal scores can be developed for classifying future students. A similar approach could be used in an occupational setting by administering the test to non-native English-speaking employees whose proficiency has already been deemed acceptable by their institution.

References

- Council of Europe (2001) *Common European Framework of Reference for Languages: learning, teaching, assessment*, Cambridge: Cambridge University Press.
- Institute of International Education (2012). *Open Doors 2012: Fast Facts* Retrieved from the IIE website <http://www.iie.org/Research-and-Publications/Open-Doors/Data/Fast-Facts>

Appendix A. Sample Items

Samples of each item type are presented below. The key to each item is indicated with an asterisk.

Listening Question

[Audio only: *Can we talk about the group project tomorrow?*]

1.
 - a. Yes, I think it's sold out.
 - b. Yes, it was a big success.*
 - c. Yes, the store opens soon.

Listening Dialog

[Audio Only:

W: *Can you tell me how much it's going to cost to fix the car?*

M: *I think you're looking at a couple of thousand dollars in total.*

W: *(angrily) What!?! Are you serious? That's just ridiculous!*]

Why is the woman angry?

- a. The man can't fix the car.
- b. The car is not ready.
- c. The cost is not reasonable.*

Grammar

"What's on the agenda for next week?"

"The new designs will _____ by then, so we should discuss those."

- a. finalize
- b. be finalizing
- c. have finalized
- d. have been finalized*

Vocabulary

Susan loved the new jeans she bought but they were too _____ for her to wear.

- a. tense
- b. close
- c. tight*
- d. miniature

Sentence-Level Reading

It was only when Stephen went abroad to college that he learned how valuable it was to have his family living nearby.

What did Stephen realize when he went to college?

- a. He enjoyed living overseas.
- b. His family wanted him to live at home.
- c. His family missed him.
- d. His family was important to him.*

Reading Passages

Invasive Weeds

A *weed* can be defined as any plant considered undesirable or a nuisance. Often the term is applied to unwanted plants found in settings such as gardens, lawns, agricultural fields, and parks. It can also be applied to unwelcome plants in forests, wooded areas, and other natural habitats.

Weeds that grow and reproduce rapidly, crowding out other plants, are known as *invasive weeds*. Invasive weeds are generally nonnative species, often introduced by humans, whether intentionally or accidentally. They harm the environment by outcompeting native plants, altering animal habitats, and increasing soil erosion.

Controlling invasive weeds can be difficult and expensive. For small- or medium-sized infestations, introducing insects or diseases that attack the weeds can be effective. Machines or people working by hand can also dig up invasive weeds, taking care not to scatter the seeds in the process. Often, when an invasive weed infestation is discovered, it has become so large that containment may not be practical. Herbicides—chemicals that kill the plants—may need to be sprayed, either by ground crews or from aircraft. Often the most successful approach is to use a combination of these techniques.

What is the main idea of the text?

- a. Weeds grow more slowly than other plants.
- b. Weeds can be easily controlled.
- c. Most weeds are native species.
- d. Some weeds are bad for the environment.*

In the second sentence of the passage, what does **the term** refer to?

- a. settings
- b. plant
- c. weed*
- d. nuisance

According to the information in paragraph 2, what is one characteristic of invasive weeds?

- a. They are harmful to humans.
- b. They grow very quickly.*
- c. They need special soil conditions.
- d. They are generally native to the area.

In paragraph 2, why does the author mention soil erosion?

- a. to give an example of how weeds affect the environment*
- b. to introduce a problem caused by some animals
- c. to compare the effects of two types of weeds
- d. to explain that some native plants are harmful

What does the author conclude about fighting invasive weeds?

- a. Using several methods together is best.*
- b. Containment is the most practical option.
- c. Chemicals are not a safe option.
- d. Herbicides are most effective on small infestations.

Appendix B. Score Concordance Table from the Prior Versions of the EPT to the New EPT Forms

Score Concordance Table for the New EPT

Old Version			New Version		
Form A	Form B	Form C	Form D	Form E	Form F
1	0	1	1	1	1
2	0	2	2	2	2
3	0	3	3	3	3
4	1	4	4	4	4
5	2	5	5		5
6	3	6	6	5	6
7	4	7	7	6	
8	5	8	8	7	7
9	6	9	9	8	8
10	7	10	10	9	9
11	8, 9	11	11	10	10
12	10	12	12	11	11
13	11	13	13		12
14	12	14		12	13
15	13	15	14	13	
16	14	16	15	14	14
17	15	17	16	15	15
18	16	18	17	16	16
19	17		18		17
20	18	19	19	17	18
21	19	20	20	18	19
22	20	21		19	
23	21	22	21	20	20
24	22	23	22	21	21
25	23	24	23		22
26	24	25	24	22	23
27	25	26	25	23	24
28	26	27		24	
29	27	28	26		25
30	28	29	27	25	26
31	29	30	28	26	27

Score Concordance Table for the New EPT

Old Version			New Version		
Form A	Form B	Form C	Form D	Form E	Form F
32	30	31	29	27	28
33	31	32		28	
34	32	33	30	29	29
35	33	34	31		30
36	34	35	32	30	31
37	35	36	33	31	32
38	36	37		32	
39	37	38	34	33	33
40	38	39	35		34
41	39	40	36	34	35
42	40	41		35	36
43	41	42	37	36	
44	42	43	38		37
45	43	44	39	37	38
46	44	45	40	38	39
47	45	46		39	40
48	46	47	41	40	
49	47	48	42		41
50	48		43	41	42
51	49	49		42	43
52	50	50	44	43	
53	51	51	45		44
54	52	52	46	44	45
55	53	53		45	46
56	54	54	47	46	47
57	55	55	48	47	
58	56	56	49		48
59	57	57		48	49
60	58	58	50	49	50
61	59	59	51	50	
62	60	60	52		51
63	61	61		51	52
64	62	62	53	52	53
65	63	63	54	53	

Score Concordance Table for the New EPT

Old Version			New Version		
Form A	Form B	Form C	Form D	Form E	Form F
66	64	64	55	54	54
67	65	65			55
68	66, 67	66	56	55	56
69	68	67	57	56	
70	69	68	58	57	57
71	70	69			58
72	71	70	59	58	59
73	72	71	60	59	
74	73	72	61	60	60
75	74	73		61	61
76	75	74	62		62
77	76	75	63	62	63
78	77	76	64	63	
79	78			64	64
80	79	77	65		65
81	80	78	66	65	66
82	81	79	67	66	
83	82	80		67	67
84	83	81	68	68	68
85	84	82	69		69
86	85	83		69	
87	86	84	70	70	70
88	87	85	71	71	71
89	88	86	72		72
90	89	87		72	
91	90	88	73	73	73
92	91	89	74	74	74
93	92	90	75		75
94	93	91		75	
95	94	92	76	76	76
96	95	93	77	77	77
97	96	94	78	78	78
98	97	95			
99	98	96	79	79	79

Score Concordance Table for the New EPT

Old Version			New Version		
Form A	Form B	Form C	Form D	Form E	Form F
100	99	97	80	80	80
	100	98			
		99			
		100			