UNIVERSITY OF MICHIGAN
ENGLISH LANGUAGE INSTITUTE

TESTING & CERTIFICATION DIVISION

RESEARCH REPORTS

2001 – 01

MELAB/COMPUTER-BASED TOEFL STUDY

BARBARA DOBSON
IRENE HAN
AMY D. YAMASHIRO

MELAB/Computer-Based TOEFL Study
(March 2001)

Purpose

While there have been earlier studies comparing MELAB and TOEFL, the most recent, conducted in 1996, was based on MELAB scores and self-reported scores on the paper-and-pencil TOEFL. With the implementation of the Computer-Based TOEFL (CBT) came the need for a new study. This study was conducted to provide information on the relationship between MELAB scores and Computer-Based-TOEFL (CBT) scores. While it is not advisable to try to "translate" scores from one of these standardized examinations to scores on the other, it is useful to study how the same group of examinees performs on each of the two test batteries. Such a study can provide helpful practical information to admissions officers who work with both of these tests.

Participants

There were 110 participants in this study, 39 males and 71 females. They were self-selected volunteers from among the total group of people applying to take the MELAB in Ann Arbor, MI. They learned about the study from an informational flyer when they registered for a MELAB test at the Ann Arbor test center. There were two requirements for participation in the study: 1) the examinee must take or have taken the CBT within 30 days of the date of their MELAB administration; and 2) the candidate must agree to provide ELI-UM with a copy of that CBT score report for use in this study. Test candidates were allowed to participate only one time in the study. There was no restriction as to whether they took the CBT or the MELAB first. Sixty-two took CBT first; forty-eight took the MELAB first. The mean number of days between taking the two test batteries was 16 days. The earliest MELAB test date for the subjects was in September, 1998. The latest test was in November, 2000.

Tables 1 – 3 below provide additional background information about the subjects. As can be seen in Table 1, the participants came from 18 different countries, but 80% of the subjects were from five countries in Asia (Korea=24.5%, Thailand=16.4%, China=14.5%; Taiwan=12.7%; and Japan=10.9%).

Table 1: Countries of Origin

Country	Frequency	Percent
Argentina	1	.9
Azerbaijan	1	.9
China PRC	16	14.5
France	2	1.8
Guinea	1	.9
India	5	4.5
Indonesia	1	.9
Japan	12	10.9
Jordan	2	1.8
Korea	27	24.5
Lebanon	1	.9
Mexico	4	3.6
Pakistan	1	.9
Palestine	1	.9
Saudi Arabia	2	1.8
Switzerland	1	.9
Taiwan	14	12.7
Thailand	18	16.4
Total	**110**	**100.0**

Table 2 shows the native languages of the participants. Although 14 languages were represented, about two-thirds of the subjects were native speakers of Chinese, Korean, or Thai.

Table 2: Native Languages

Language	Frequency	Percent
Arabic	6	5.5
Azerbaijan	1	.9
Chinese	30	27.3
French	2	1.8
German	1	.9
Hindi	2	1.8
Indonesian	1	.9
Japanese	12	10.9
Korean	27	24.5
Malayalam	1	.9
Malinke	1	.9
Spanish	5	4.5
Telegu	2	1.8
Thai	18	16.4
Urdu	1	.9
Total	**110**	**100.0**

Participants each took one of 9 different forms of MELAB Part 1 (Composition), one of 4 different forms of MELAB Part 2 (Listening) and 6 different forms of MELAB Part 3 (Grammar/Cloze/Vocabulary/Reading). As is standard practice with operational MELABs, part scores (scaled) and a MELAB Final score (mean of the scaled part scores) were reported. It is these scaled scores that are used in this study for comparison with CBT scores.

Results and Analysis

Descriptive statistics for the MELAB Final scores and CBT total scores of the participants in this study are reported in the dark gray columns of Table 3 below. Also shown (in the light gray columns) are the corresponding statistics for the "total" population of CBT and MELAB test takers.

Table 3: Descriptive Statistics for Sample Population and "Total" Population

	"Total" CBT[1]	Sample CBT	"Total" MELAB[2]	Sample MELAB
Minimum	10	73	38	49
Maximum	300	267	99	91
Mean	215	181.6	75.8	70.3
Standard Deviation	46	42.9	10.4	9.2

As can be seen from the minimum and maximum values, the range of scores in the sample group is nearly as wide as the range for all who take CBT and MELAB (from below the 1st to approximately the 95th percentile rank for both test batteries).

[1] *TOEFL Test and Score Data Summary, 2000-2001 Edition*, p. 4
[2] *MELAB Technical Manual, 1996*, p. 17

It should be noted, though, that the sample group is noticeably different from the total CBT and total MELAB populations in terms of their proficiency level, as measured by both of the tests. The mean CBT score for these 110 subjects was 181.62 while the mean total score for the "total group" of 317, 708 examinees taking CBT between July, 1999 and June, 2000 was 215.[3] The mean MELAB score for the sample group was 70.00, whereas the mean for all first-time MELAB test-takers between 1991 and 1993 was 75.84. The mean CBT and MELAB scores of the sample group fall at less than the 30th percentile rank of the "total" populations taking these tests.

Perhaps it is not surprising that the test takers in this study are, on average, less proficient than the average CBT or MELAB test taker. It seems reasonable to assume that lower-proficiency candidates might be more likely to take two different English proficiency tests than would candidates with a higher proficiency and, presumably, more confidence that the score from a single test would be enough to meet the entrance requirements for a program they hope to enter.

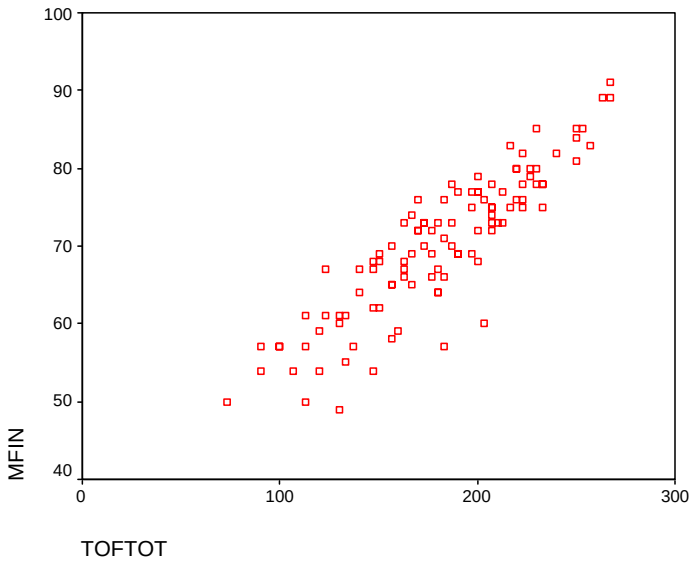
The subjects' CBT total scores were correlated with their MELAB Final scores, and the results are shown in Table 4.

Table 4: Correlation of MELAB Final and CBT Total Scores

Total N	Correlation (MELAB / CBT final scores)	Significance level
110	.89	.01

Figure 1, below, graphically illustrates this relationship. There is more variability at the lower end of the plot than at the higher end.

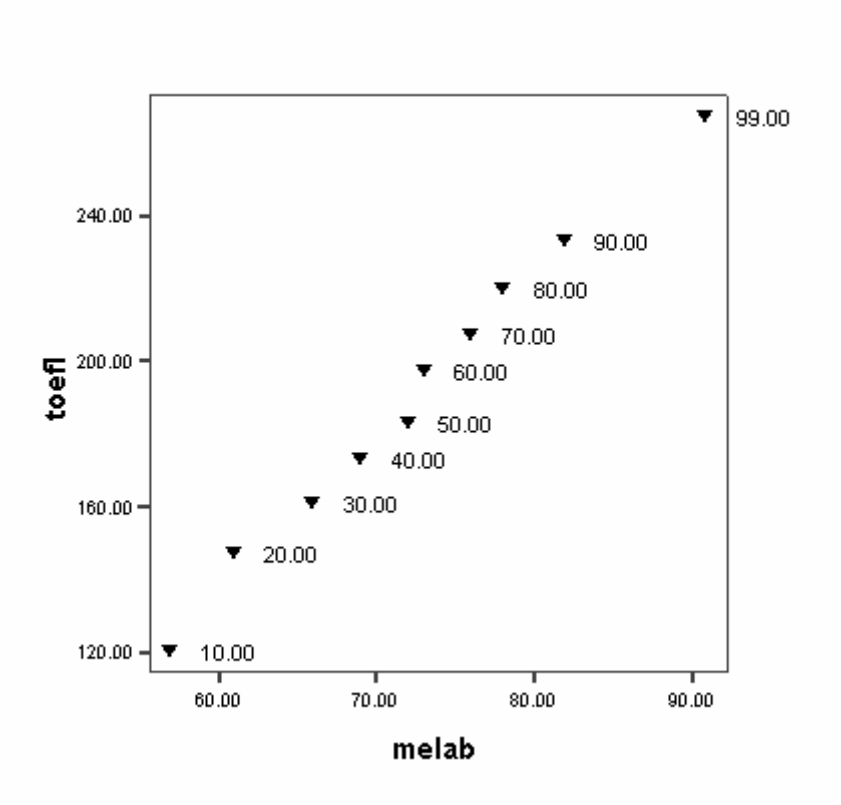
Figure 1: Scatter Plot of CBT Total Scores and MELAB Final Scores



[3] *TOEFL Test and Score Data Summary, 2000-2001 Edition*, p. 4

Figure 2, below, which is a plot of the score pairs at the 10th, 20, ….80, 90th percentiles of the sample group, indicates that the relationship is strongly linear throughout the range of scores.

Figure 2: Correlation of MELAB and CBT Total Scores at Equipercentile Points



The high correlation (.890, uncorrected for attenuation) between the two sets of scores suggests that the two test batteries measure English proficiency levels in a similar manner. Still, this information is not sufficient justification for "translating" or "converting" scores from one test into scores for the other. The CBT and the MELAB are different in content and format and cannot be assumed to be measuring the same construct.

Those interested in establishing guidelines for using the MELAB as an admission criterion might find the attached Concordance Table and Range Comparison Table helpful in establishing *initial* guidelines for using MELAB scores. It must be emphasized, though, that local validation studies should be conducted to examine whether initial guidelines that are used prove to be appropriate.

The table labeled *Concordance Table: Total Score* was created by matching MELAB and CBT total scores based on the common percentile rank for the participants in this study. For example, a CBT Total score of 221 is at the same percentile rank for the sample group as a MELAB Final score of 78. Both scores share a percentile rank of 80 *for the sample group*.[4] In cases where there was more than one CBT score for a particular MELAB score, the CBT scores were averaged. The comparisons between CBT and Paper-Based TOEFL (PBT) scores are taken from a concordance table prepared by ETS[5]. An example of how to read the Concordance Table is given below the table. The Range Comparison Table provides another look at how MELAB, CBT, and PBT scores relate, and highlights the fact that direct "translations" of a score on one test to a score on another test is not appropriate.

This report was prepared by Barbara Dobson, Irene Han, and Amy Yamashiro, March 2001. The authors would like to gratefully acknowledge the contributions of Tony Bouttavong, Maria Huntley, and Alethia Ware, MELAB Administrative Staff, to gathering the data used in this study and Sarah Briggs, Research Associate, for initiating the study and conducting preliminary analyses of the data.

For further information about this study, please contact the MELAB Office (melabelium@umich.edu), Barbara Dobson (bdobson@umich.edu), Sarah Briggs (slbriggs@umich.edu), or Amy Yamashiro (amydyama@umich.edu).

[4] Again, it should be noted that this sample group is not as proficient as the total population that takes MELAB and CBT. A CBT score of 220 is at only the 49th percentile rank for the "total" population of CBT takers. A MELAB score of 78 is at only the 56th percentile rank for the "total" population of MELAB takers.

[5] *TOEFL Concordance Table*, ETS, 1998

MELAB/CBT STUDY, March 2001
University of Michigan, ELI, Testing Division
MELAB Contact Info: Tel: (734)764-2416, <melabelium@umich.edu>

CONCORDANCE TABLE: TOTAL SCORE

MELAB Total, TOEFL Computer-based Total (CBT)**, & TOEFL Paper-based Total** (PBT)

Concordance Table: Total Score

MELAB	CBT	PBT
{92}	{273}	{640}
91	267	630
89	266	
88	261	620
85	253	
84	250	600
83	241	
82	233	577
81	232	
80	229	567
79	225	
78	221	560
77	214	
76	207	540
75	202	
74	200	533
73	192	520
72	185	
71	183	513
70	180	510
69	176	
	{173}	{500}
68	169	
67	165	
66	160	
65	157	
64	150	
63	149	
62	147	
61	144	
60	135	
59	131	
57	122	
55	113	
54	103	
51	93	
50	90	
49	75	

Range Comparison

MELAB	CBT	PBT
{94+}	{287+}	{660+}
{92-93}	{273-283}	{640-657}
87-91	260-270	620-637
84-87	250-260	600-617
82-83	237-247	580-597
78-81	220-233	560-577
76-77	207-220	540-557
73-75	190-203	520-537
69-72	173-187	500-517
65-68	157-170	480-497
61-64	140-153	460-477
59-60	123-137	440-457
55-58	110-123	420-437
54	97-107	400-417
50-51	83- 93	380-397
49	70- 80	360-377

{} denotes estimated scores based on other studies

The concordance table was created by matching MELAB and CBT total scores based on the common percentile rank for the participants in this study. In cases where there were more than one CBT score for a particular MELAB score, the CBT scores were averaged. This table is intended to help those interested in establishing *initial* guidelines for using MELAB scores as an admission criterion. (Local validation studies should be conducted to examine whether initial guidelines used prove to be appropriate.)

Estimated MELAB scores are based on a MELAB/TOEFL Score Table presented in an ELI Testing Division Memorandum (Briggs, 1996).

The TOEFL CBT and PBT score comparisons and estimated scores are based on the *TOEFL Concordance Table* prepared by ETS (1998).

To use the above Concordance Table, you can start from a particular MELAB score, for example, 80, and then find the corresponding TOEFL CBT (229) and PBT (567). Or conversely, you can start from a particular TOEFL CBT score, such as 220, and then find the corresponding MELAB score (78). Please remember this table is not for converting scores, rather it is intended to assist Admissions Officers in interpreting and using MELAB scores.