DOCUMENT RESUME

ED 060 027 TM 001 110

AUTHOR Angoff, William H.; Stern, June
TITLE The Equating of the Scales for the Canadian and American Scholastic Aptitude Tests.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY College Entrance Examination Board, New York, N.Y.; Service for Admission to College and University, Ottawa (Ontario).
REPORT NO PR-71-24
PUB DATE Dec 71
NOTE 39p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Comparative Statistics; Correlation; Cross Cultural Studies; *Equated Scores; Item Analysis; *Mathematics; Statistical Analysis; Testing; *Test Interpretation; *Verbal Ability
IDENTIFIERS Canadian Scholastic Aptitude Test; *College Board Scholastic Aptitude Test; CSAT; SAT

ABSTRACT

This study examines a common set of equating items for their appropriateness to both American and Canadian populations and, to determine the equations for converting scores from the verbal and mathematical scales of the 1971 form of the Canadian Scholastic Aptitude Test (the future standard continuing scales of the CSAT) to the corresponding standard continuing scales of the Scholastic Aptitude Test. (AG)

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS
RDR-71-72, NO. 4

•

PROJECT REPORT
PR-71-24 DECEMBER 1971

The Equating of the Scales for the Canadian and American Scholastic Aptitude Tests

•

William H. Angoff and June Stern

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

The Equating of the Scales for the Canadian and American Scholastic Aptitude Tests

by

William H. Angoff

and

June Stern

December, 1971

This project was sponsored jointly by the Service for Admission to College and University and the College Entrance Examination Board.

Introduction

The scores earned on the verbal and mathematical sections of the College Board Scholastic Aptitude Test (SAT) are reported on a scale that extends from 200 to 800, a scale originally established by setting the mean and standard deviation of candidates tested in April 1941 at the arbitrary values of 500 and 100, respectively. The transformation from raw (formula) scores to scaled scores is a linear one, which preserves all moments of the raw score distribution above the second, that is, it retains the shape of the raw score distribution. The SAT scale is a continuing scale; scores on each form of the test are equated to scores on the preceding forms with the effect that, within the limits of equating error, a score of 563, for example, earned on any one form of SAT-verbal has the same meaning -- represents the same level of ability -- as a score of 563 on any other form of SAT-verbal.

The scores earned on the verbal and mathematical sections of the Canadian Scholastic Aptitude Test (CSAT) of the Service for Admission to College and University (SACU) are also reported on a 200-800 scale. Here, too, the transformation from raw (formula) scores to scaled scores is a linear one. Unlike the SAT-verbal scale, however, which is equated back to April 1941, the 1971 linear scale of the CSAT-verbal was defined on the form actually administered in January 1971 by setting the minimum raw score at 200 and the maximum raw score at 800. The two earlier forms of the CSAT-verbal (introduced in 1969 and in 1970) were scaled in the same way, independently for each form and not equated to each other. However, scores on all future CSAT-verbal forms will be equated to the CSAT-verbal scale for January 1971. Thus, the January 1971 scale becomes the basic reference scale for CSAT-verbal and corresponds directly to the April 1941 reference scale used for SAT-verbal.

Statements similar to these just made for the verbal tests are also appropriate for SAT-mathematical and CSAT-mathematical, respectively.

Although both the CSAT and the SAT are very similarly designed tests of verbal and mathematical aptitude, and although their scales are similar in appearance, it is clear from the foregoing, and also from the fact that no effort has ever been made to link the two systems together, that equal-appearing scores in the two systems -- say, again, 563 on CSAT-verbal and SAT-verbal -- might have quite different meanings and reflect quite different levels of ability. Accordingly, the present study was undertaken to examine a common set of equating items for their appropriateness to both populations and, following that examination, to determine the equations for converting scores from the verbal and mathematical scales of the 1971 form of the CSAT (the future standard continuing scales of the CSAT) to the standard continuing verbal and mathematical scales of the SAT.

Method

The test forms used in the present equating study were the form of the SAT administered on January 9, 1971 and the form of the CSAT administered on January 20, 1971. As an introduction to the design of the study it will be useful to review the administrative arrangements and content of the two tests. Both tests were administered in five separately timed sections: two operational verbal sections, two operational mathematical sections, and one nonoperational section administered in a time slot reserved for the presentation of pretest or equating items. Table 1 describes the item lengths and timing of the two tests and the order of appearance of the sections in the forms of the tests studied here.

The SAT-verbal contains four item types: antonyms, analogies, sentence completion, and reading comprehension. Typically, all four item types are represented in both of the timed sections of the SAT. The CSAT-verbal items

are of the same type as those in the SAT. Unlike the SAT, however, one of the two separately timed verbal parts of the CSAT contains exclusively discrete items; the other contains exclusively reading comprehension items.

Table 1

Administrative Arrangement and Content

of the CSAT and the SAT

SAT				CSAT			
Section		No. of Items	Timing (Min.)	Section		No. of Items	Timing (Min.)
I	Verbal	40	30	I	Math	20	30
II	Verbal	50	45	II	Verbal	55	30
III	Nonopnl	Variable	30	III	Nonopnl	Variable	30
IV	Math	35	45	IV	Math	20	30
V	Math	25	30	V	Verbal	20	30

Except for the data sufficiency items (the SAT has 18; the CSAT has none) the mathematical items in the SAT and CSAT are quite similar. The timing, however, is different. The SAT gives 1 1/4 minutes per mathematical item; the CSAT gives 1 1/2 minutes.

As indicated above, the nonoperational section of both tests (the third time slot in both the SAT form and the CSAT form used in this study) is used for pretesting items and also, in the SAT, for equating successive forms of the tests. In any formal administration of either test more than one nonoperational section may be administered in that time slot, with different students taking different nonoperational sections. In order to assure that randomly equivalent groups of students take one such section or another, the test booklets are distributed in "spiralled" arrangement. Thus, if m non-

operational sections are to be administered, every mth student takes the same such section. In the administration of the SAT, for example, it is usual for m to be as high as 25 or 30.

In preparation for the present study arrangements were made between the Service for Admission to College and University (SACU) and the College Entrance Examination Board (CEEB) to exchange 30-minute operational sections and for each of the two organizations to administer the sections of the counterpart organization in its own nonoperational position. Thus, in addition to the usual 30-minute pretests and 30-minute equating sections normally assigned to the nonoperational position in the January 1971 SAT, Educational Testing Service (ETS) also administered (in behalf of CEEB) one of the two 30-minute operational CSAT-verbal sections and one of the two 30-minute operational CSAT-mathematical sections taken from the January 1971 CSAT. Correspondingly, in addition to the usual 30-minute pretest sections normally assigned to the nonoperational position in the January 1971 CSAT, SACU also administered the 30-minute operational SAT-verbal section and the 30-minute operational SAT-mathematical section taken from the January 1971 SAT.

With this arrangement a full 60 minutes of testing material in the verbal area -- representing a total of 95 items -- was administered for purposes of equating to subsamples of both populations in connection with their operational testing. Thirty minutes of this material (40 items) was operational in the SAT (Section I) -- i.e., contributing to the student's formal score -- and nonoperational in the CSAT; thirty additional minutes (55 items) was operational in the CSAT (Section II) and nonoperational in the SAT. Under the same sort of arrangement a full 60 minutes of testing material in the mathematical area -- representing a total of 45 items -- was administered, for the purposes of

equating, to subsamples of both populations in connection with their operational testing. Thirty minutes of this material (25 items) was operational in the SAT (Section V) and nonoperational in the CSAT; thirty additional minutes (20 items) was operational in the CSAT (Section I) and nonoperational in the SAT. As indicated above, all nonoperational sections appeared in Section III of both tests. The operational SAT-verbal section is referred to as SAT-Va and the operational CSAT-verbal section is referred to as CSAT-Vc. The total of the two sections is referred to as Vt. Corresponding notation is used in the SAT- and CSAT-mathematical tests: Ma, Mc, and Mt.

The mathematical sections that were exchanged between the two organizations were quite similar in content, representing a mixture of arithmetic reasoning, algebra, and geometry. The CSAT-verbal section that was administered along with the SAT contained sentence completion items, analogies and opposites only. The SAT-verbal section administered along with the CSAT contained these same item types, but in addition, included reading comprehension items as well.

The procedure followed in developing the conversions between the two testing systems called for two general steps: 1) an examination of the 95 common items in the verbal area and the 45 common items in the mathematical area in order to identify items that appeared to have somewhat "different meaning" for American and Canadian students, and to eliminate such items from the group of common items to be used for equating the two tests; 2) the development of the necessary statistics, based on the remaining items, for equating the verbal and mathematical sections of the two tests.

Examination of Item Data

In order to carry out the first of these two steps, two mutually exclusive random samples of the SAT candidates who took Vc were drawn (these

candidates also took Va as part of the operational SAT), and similarly, two mutually exclusive random samples of the CSAT candidates who took Va were drawn (these candidates also took Vc as part of the operational CSAT). Four additional samples were selected for mathematical, and in the same way: two mutually exclusive random samples of the SAT candidates who took Mc were drawn (these also took Ma operationally), and similarly, two mutually exclusive random samples of CSAT candidates who took Ma were drawn (these also took Mc operationally). Each of the individual American samples consisted of 450 cases; the Canadian samples, however, ranged in size from 405 to 420. The four SAT samples are referred to as AV1, AV2, AM1, and AM2; the four corresponding CSAT samples are referred to as CV1, CV2, CM1, and CM2. In addition, samples AVT, CVT, AMT, and CMT were formed. (The letters A and C denote American and Canadian students respectively; the letters V and M denote tests for which these data were to be analyzed: verbal and mathematical; the numerals 1 and 2 are used only to distinguish between random subsamples of students drawn from the same population and taking the same battery of tests. The letter T denotes a combined sample; thus AVT = AV1 + AV2.)

The first step in the analysis -- the examination of the performance of the American and Canadian students on the same items -- involved the use of a technique that will bear detailed description. In comparing two samples of individuals in this way, an item analysis is carried out for each of the two groups and a plot is made of the points represented by the paired Δ-values, Δ_{g_i} vs Δ_{h_i} *, one point for each of the items (i) under consideration for

*Delta (Δ) is an index of item difficulty for an item. $\Delta = 4z + 13$, where z is a normal deviate corresponding to p, the proportion of examinees answering the item correctly.

which Δ-values are available. The plot of these points normally appears in the form of an ellipse extending from lower left to upper right, and if the samples are drawn from the same type of population, the scatterplot of these points is a long, narrow one, often representing a correlation as high as .98 or .99. When the samples are different in level, the points will still fall in a long, narrow ellipse, but displaced vertically or horizontally, depending on which group is the abler one. Even when the groups differ in degree of dispersion, the points will still fall in the same type of ellipse, but the ellipse will be tilted at an angle more or less steeper than 45°, depending on which sample is the more dispersed. However, when the groups are different in <u>type</u>, the item difficulties will not fall in precisely the same rank order for the two groups, and the correlation represented by the delta-points will be lower than .98 or .99. The items falling at some distance from the plot may be regarded as contributing to the item x group interaction. They are the items that are especially more difficult for one group than for the other, <u>relative to</u> the other items.

The method developed for summarizing the significant features of each plot involves the determination of the major axis of the ellipse represented by the plot and the calculation of the perpendicular distance (D_i) from each point to that line. The standard deviation of the distribution of these distances is a function of the item x group interaction. As indicated above, the correlation coefficient represented by the ellipse represents the degree to which the items have the same rank order of difficulty in the two groups -- also a representation (inversely) of the item x group interaction.

The equation used for the major axis of the ellipse is a linear one, $h = Pg + Q$, where

$$P = \frac{(\sigma_h^2 - \sigma_g^2) \pm \sqrt{(\sigma_h^2 - \sigma_g^2)^2 + 4r_{gh}^2\sigma_g^2\sigma_h^2}}{2r_{gh}\sigma_g\sigma_h}$$

and

$$Q = M_h - PM_g \quad .$$

(It is recalled that the variables, g and h, are, respectively, the delta values for the two groups under consideration.) The formula for the perpendicular distance, D_i, of each point, i, in the plot to the line is given as:

$$D_i = \frac{Pg_i - h_i + Q}{\sqrt{P^2 + 1}} \quad .$$

In preparation for the examination of the delta plots, six item analyses were carried out for the verbal tests, one for each of the four separate verbal samples (AV1, AV2, CV1, and CV2) and one for each of the combined samples (AVT and CVT). Similarly, six item analyses were carried out for the mathematical tests, one for each of the four separate mathematical samples (AM1, AM2, CM1, and CM2) and one for each of the combined samples (AVT and CVT). Tables 2 and 3 summarize the data of the item analyses. Table 2, which summarizes the data for the verbal tests, makes it clear that the mean test scores for the Canadian examinees are higher by 1/5 to 1/4 of a standard deviation than are the means for the American examinees. This difference may be observed on the items that were originally Canadian, and also on the items originally American. As expected, the item data support this observation; the mean deltas for the American samples are higher than the mean deltas for the Canadian samples and indicate that the items are more difficult for the Americans than for the Canadians. It is also observed in both the Va and Vc items that the Canadian samples are more homogeneous than the American samples. This finding is also supported by the standard

Table 2

Means and Standard Deviations of Raw Scores and Items

for American and Canadian Samples

Verbal Tests

Sample	No. of Cases	American Items (Va); n=40						Canadian Items (Vc); n=55					
		Test Score Data		Item Data				Test Score Data		Item Data			
				Deltas		Biserials				Deltas		Biserials	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
American													
AV1	450	15.44	8.08	13.05	2.90	.492	.131	19.89	11.02	13.36	2.91	.518	.131
AV2	450	14.67	7.97	13.23	2.79	.482	.132	18.77	11.31	13.61	2.85	.534	.134
AVT	900	15.06	8.04	13.16	2.83	.487	.126	19.33	11.18	13.51	2.86	.522	.134
Canadian													
CV1	405	16.81	7.00	12.93	3.05	.455	.106	21.64	10.96	13.17	3.05	.555	.115
CV2	410	17.16	6.84	12.85	3.08	.443	.095	22.08	10.15	13.00	3.15	.534	.111
CVT	775*	16.92	6.97	12.91	3.06	.458	.097	21.78	10.61	13.13	3.08	.533	.110

*Because of the difficulty in establishing the unique identities of some cases in Samples CV1 and CV2, 40 cases were dropped in the process of combining these samples to form Sample CVT.

Table 3

Means and Standard Deviations of Raw Scores and Items

for American and Canadian Samples

Mathematical Tests

Sample	No. of Cases	American Items (Ma); n=25						Canadian Items (Mc); n=20					
		Test Score Data		Item Data				Test Score Data		Item Data			
				Deltas		Biserials				Deltas		Biserials	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
American													
AM1	450	9.90	5.93	13.15	2.39	.581	.129	6.69	5.35	13.56	2.51	.640	.073
AM2	450	10.25	5.71	12.98	2.55	.580	.103	6.67	4.98	13.53	2.57	.602	.111
AMT	900	10.07	5.82	13.07	2.44	.580	.110	6.68	5.17	13.54	2.54	.621	.081
Canadian													
CM1	420	12.16	4.68	12.23	3.28	.538	.093	9.03	4.61	12.50	2.92	.599	.128
CM2	415	12.47	4.86	12.14	3.05	.565	.104	9.27	4.67	12.36	2.86	.622	.097
CMT	775*	12.31	4.82	12.16	3.17	.566	.090	9.13	4.68	12.45	2.87	.616	.111

*Because of the difficulty in establishing the unique identities of some cases in Samples CM1 and CM2, 60 cases were dropped in the process of combining these samples to form Sample CMT.

deviations of deltas, which are larger for the Canadians than for the Americans.* To what extent these various observations can be generalized beyond the observed samples it is impossible to say. At the very most, the data indicate that the CSAT group tested in January 1971, from which the CSAT samples were randomly drawn, is more able on the verbal abilities measured by the Va and Vc items than the SAT group tested in January 1971. It is beyond the scope of these data to give evidence on the extent to which these groups are differently representative of their parent populations.

The data on the biserial coefficients are interesting in that they indicate that items originally American are slightly more discriminating for the Americans ($\bar{r}_{bis}$ = .49 for AVT) than for the Canadians ($\bar{r}_{bis}$ = .46 for CVT), while items originally Canadian are slightly more discriminating for the Canadians ($\bar{r}_{bis}$ = .55 for CVT) than for the Americans ($\bar{r}_{bis}$ = .52 for AVT). This finding in turn suggests the possibility that the items in the two tests may be homogeneous in content, but specific to the students for whom the items were intended.

Table 3, which summarizes the data for the mathematical tests, repeats some of the findings reported in Table 2. As in the verbal tests, the Canadian students score higher than their American counterparts by almost half a standard deviation on both sets of items, those originally Canadian as well as those originally American. As expected, the means of the item deltas confirm these findings; the mean deltas for the American students are higher than those for the Canadian students, indicating that the items are more difficult

*In general, standard deviations of item difficulties are larger for more homogeneous groups and smaller for more heterogeneous groups.

for them than for the Canadian students. Also, as in the verbal tests, the Canadians are more homogeneous as a group than the Americans on both the Ma and Mc subtests. This finding is also supported by the item data, which show a larger standard deviation of deltas for the Canadians. The data on item biserials are less clear for the mathematical items than they are for the verbal items. Although the item biserial correlations for mathematical are higher for the Americans on the items originally American, they are about equal for the two national groups on the items originally Canadian.

Figures 1-3 present the delta plots for the verbal items. Figure 1 is the plot of all 95 items for one random sample of American students versus the other (AV1 vs. AV2); Figure 2 is the plot for the same set of 95 items for one random sample of Canadian students versus the other (CV1 vs. CV2), and Figure 3 is the plot for the same set of items for the combined sample of Americans versus the combined sample of Canadians (AVT vs. CVT). It is clear from these figures that the cross-national plot (Figure 3) shows a greater dispersion than does either of the within-national plots (Figures 1 and 2). These observations are supported by the correlations represented by these plots (.976 for AVT vs. CVT, as compared with .990 for AV1 vs. AV2 and .991 for CV1 vs. CV2) and also by the values of σ_D (.458 for AVT vs. CVT, as compared with .287 for AV1 vs. AV2 and .296 for CV1 vs. CV2).

Figures 4-6 present the corresponding delta plots for the 45 mathematical items: Figure 4 is a plot of the deltas for one random sample of Americans versus another (AM1 vs. AM2); Figure 5 is a plot of the deltas for one random sample of Canadians versus another (CM1 vs. CM2), and Figure 6 is a plot of the deltas for the combined American sample versus the combined Canadian sample (AMT vs. CMT). Here too the cross-national plot (Figure 6) is more dispersed

Figure 1

Delta Plot of Verbal Items for
American Sample #1 vs American Sample #2
(AV1 vs AV2)

r = .990

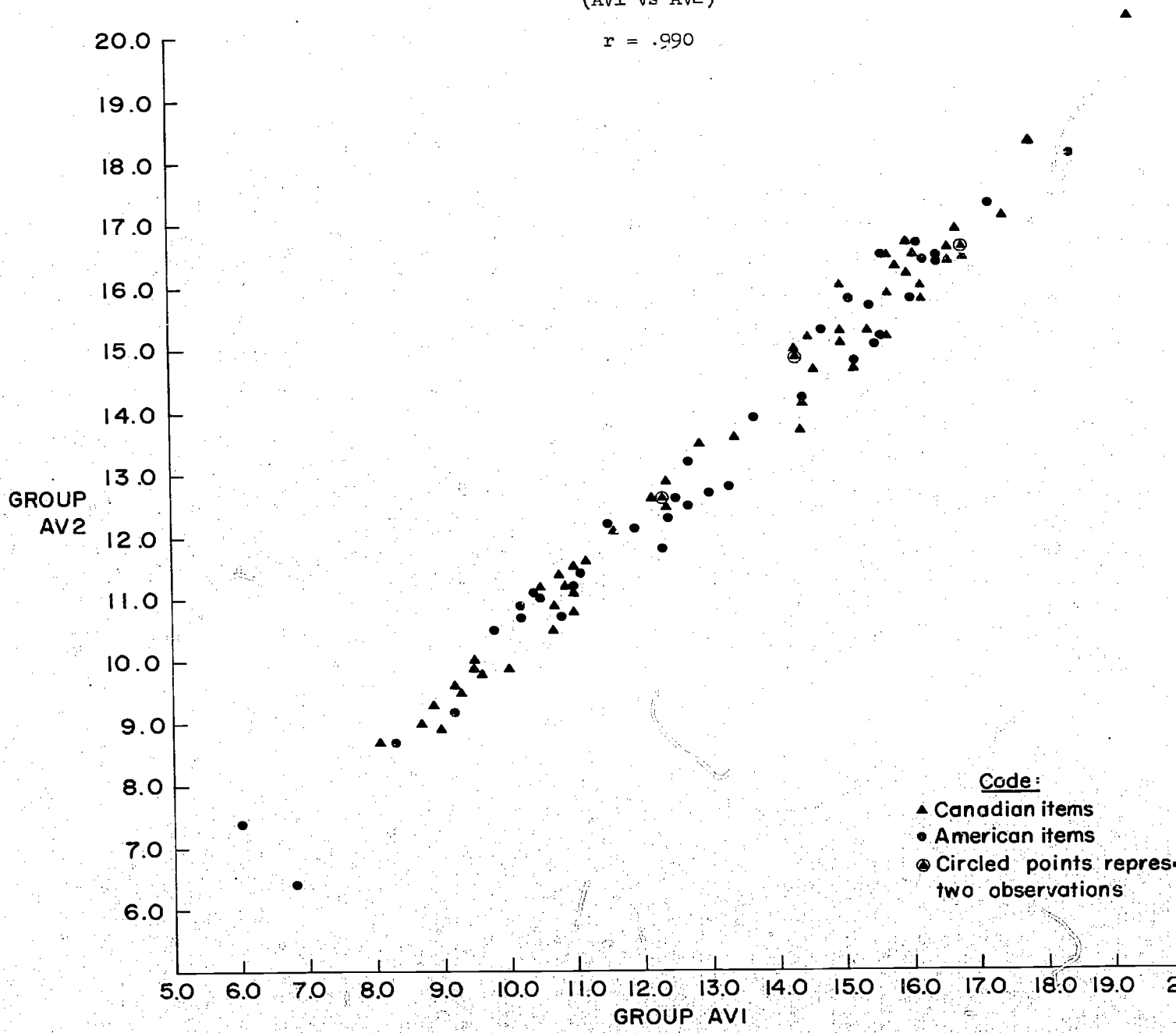


Figure 2

Delta Plot of Verbal Items for
Canadian Sample #1 vs Canadian Sample #2
(CV1 vs CV2)

r = .991

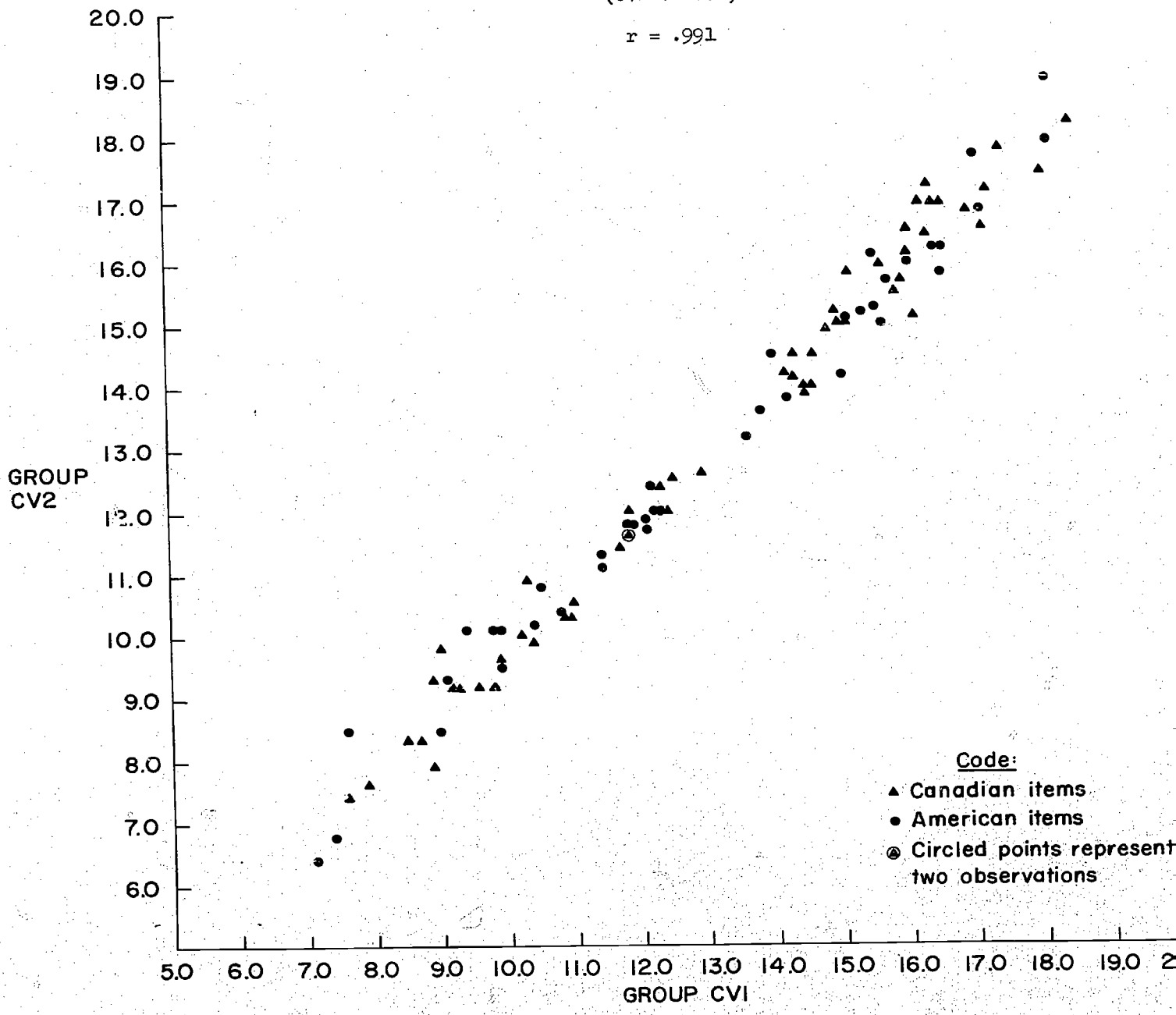


Figure 3

Delta Plot of Verbal Items for
Total American Sample vs Total Canadian Sample
(AVT vs CVT)

r = .976

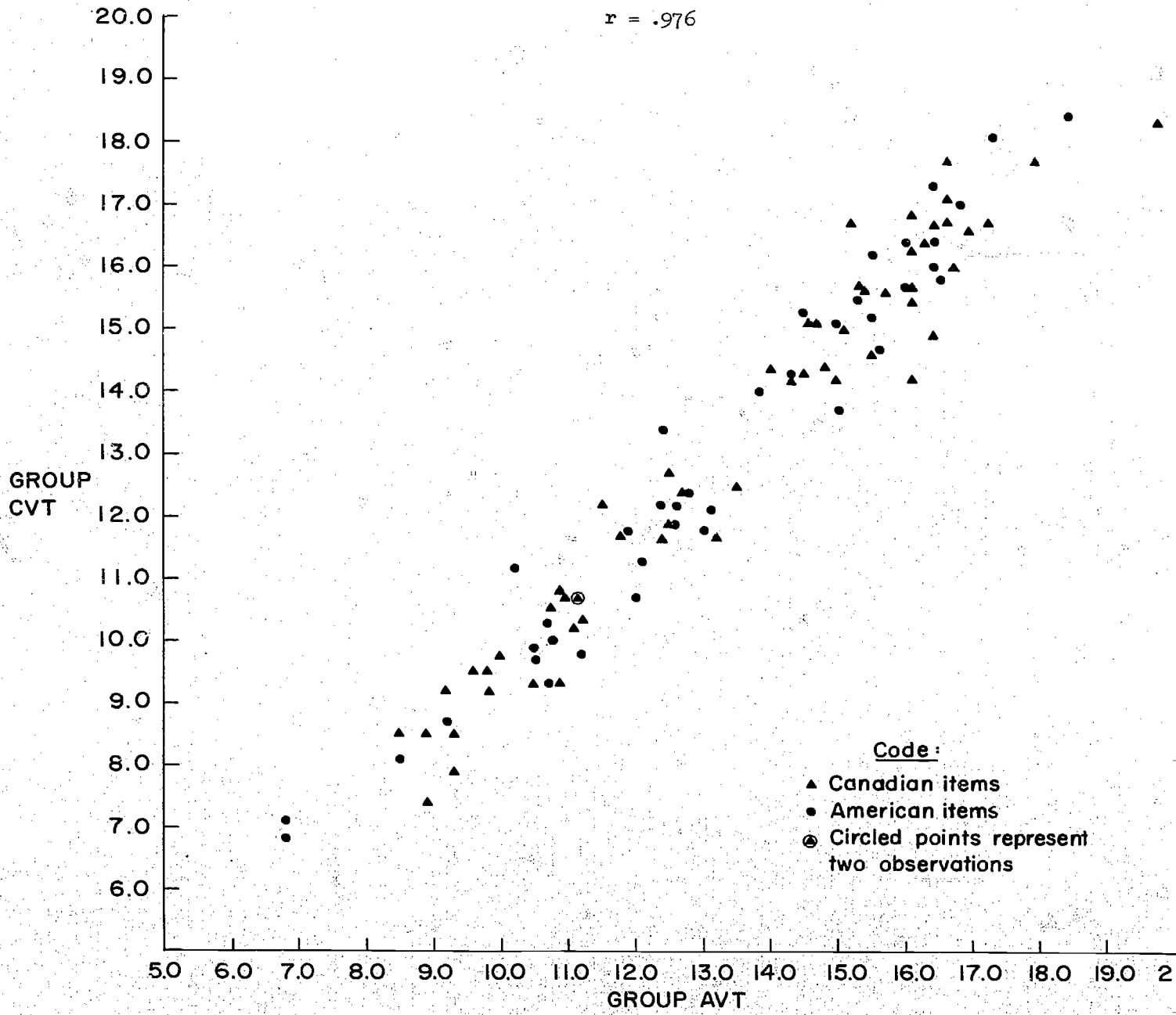


Figure 4

Delta Plot of Mathematical Items
American Sample #1 vs American Sample #2
(AM1 vs AM2)

r = .991

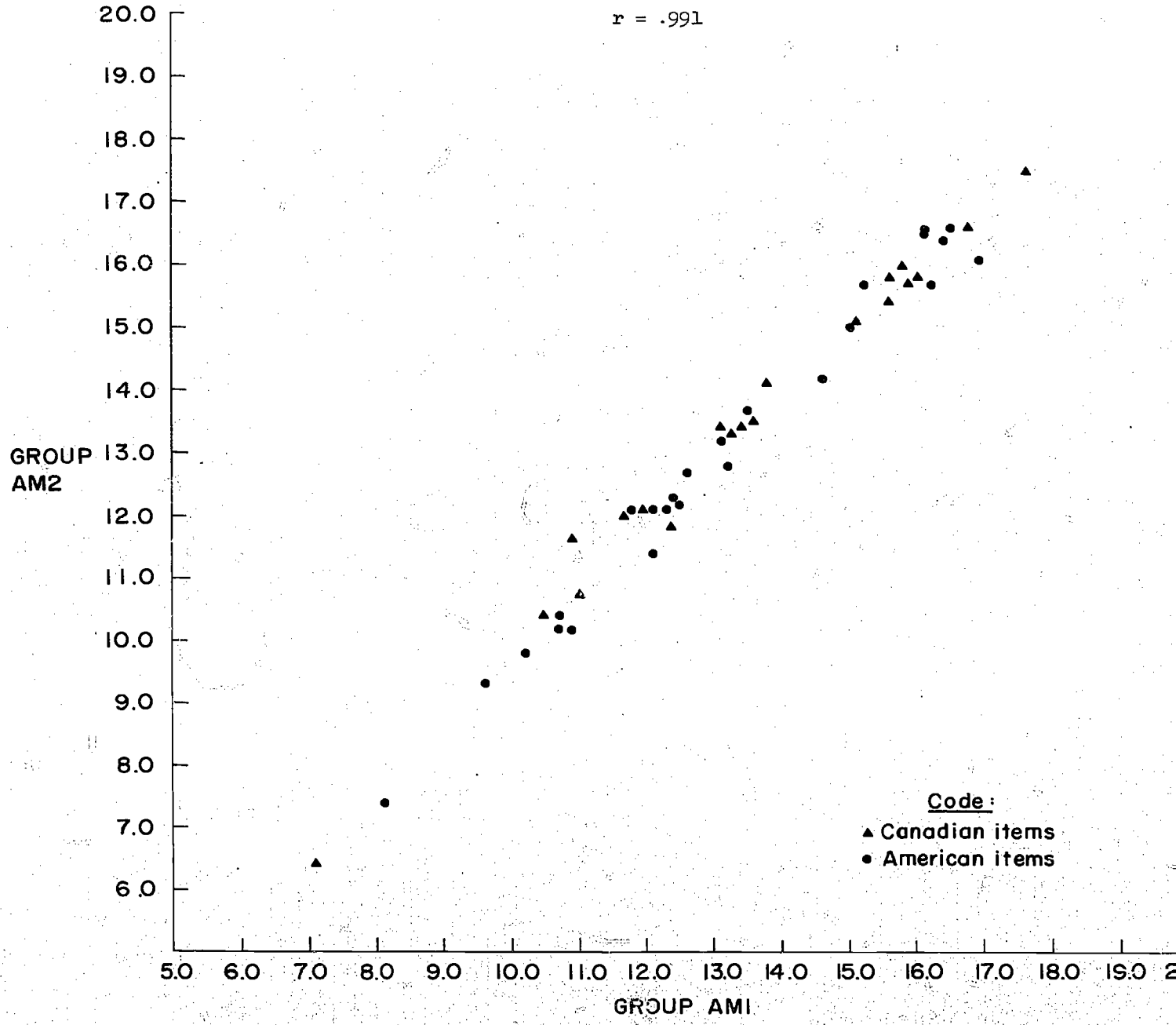


Figure 5

Delta Plot of Mathematical Items
Canadian Sample #1 vs Canadian Sample #2
(CM1 vs CM2)

r = .994

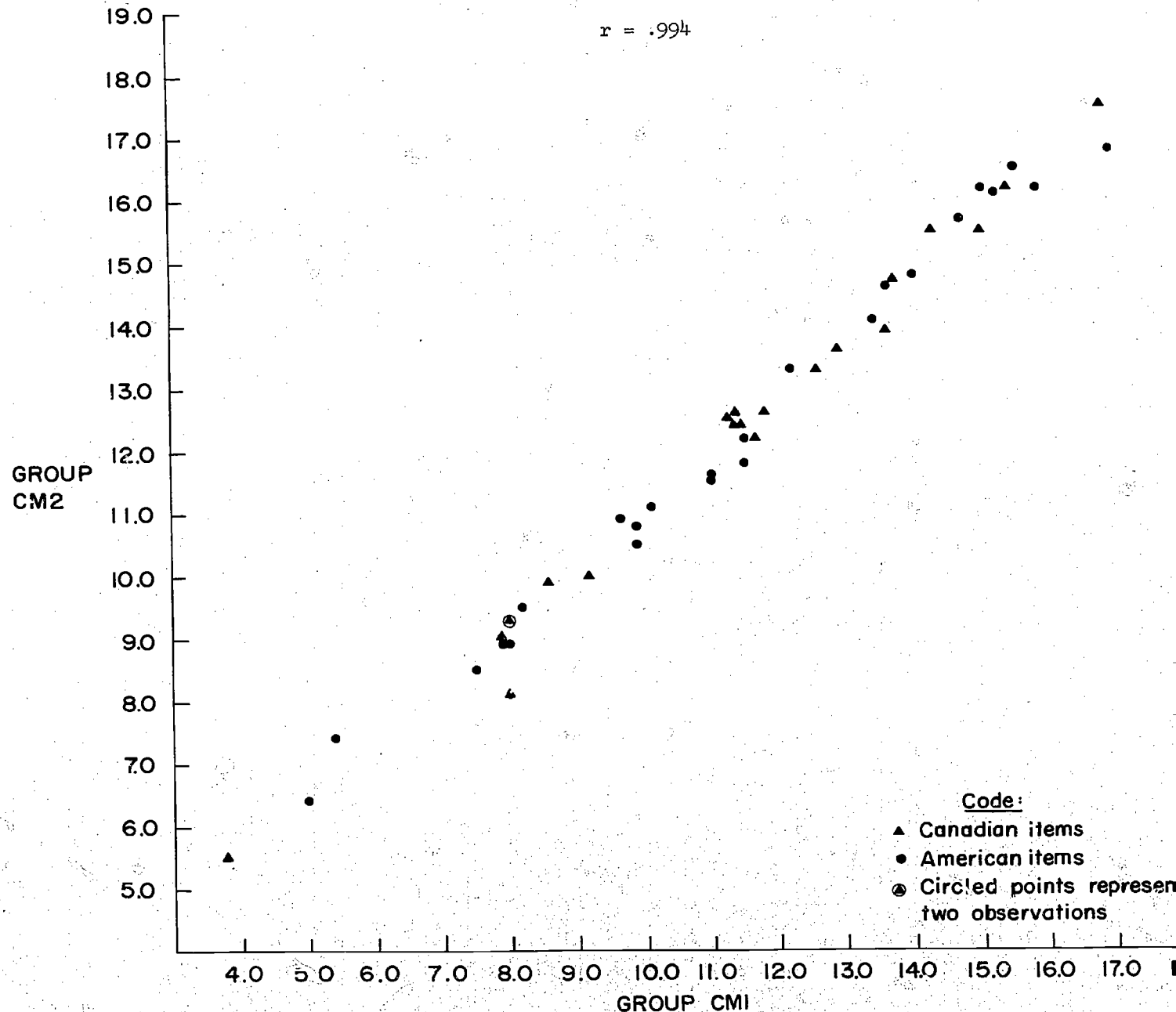
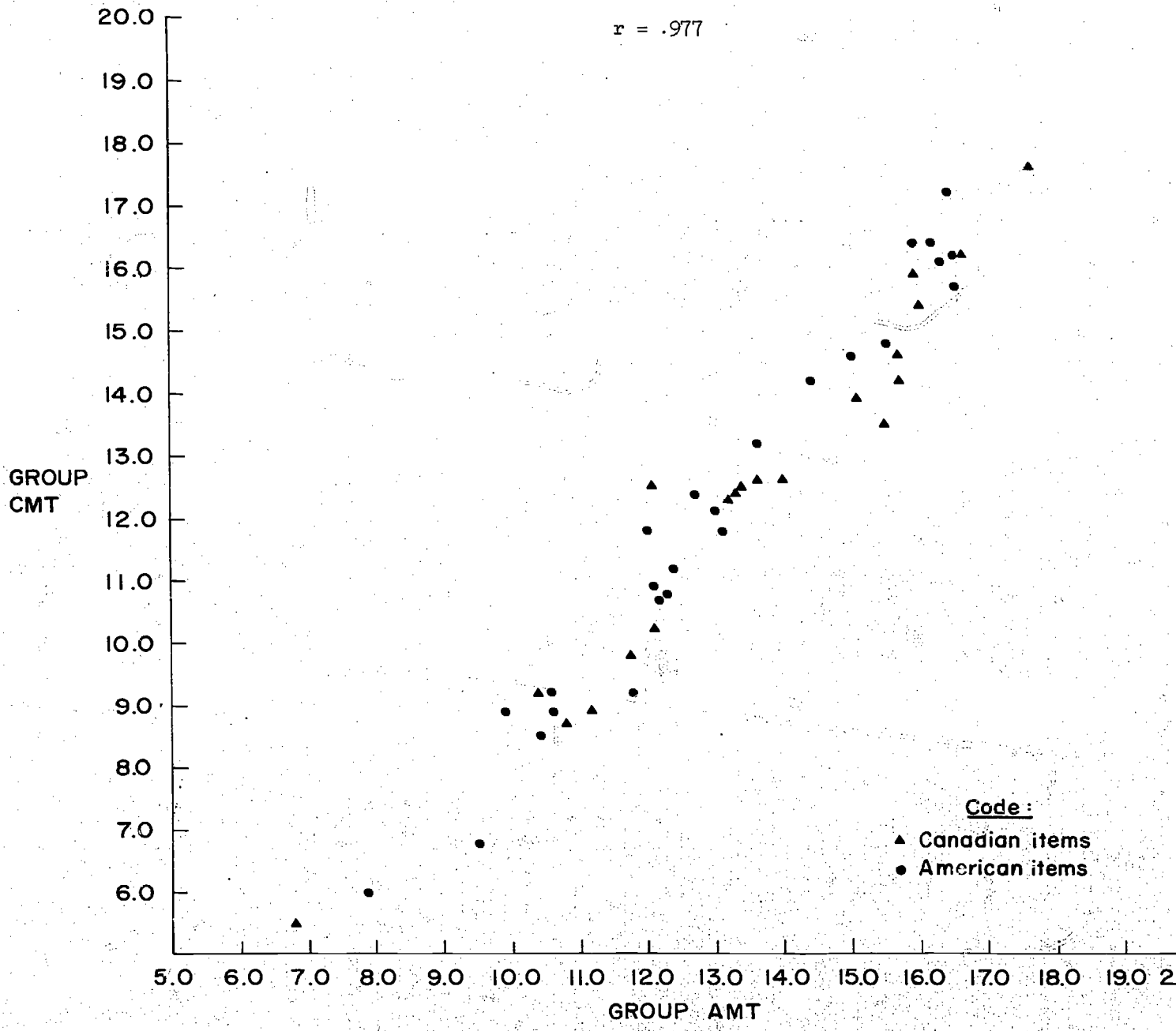


Figure 6

Delta Plot of Mathematical Items
Total American Sample vs Total Canadian Sample
(AMT vs CMT)

r = .977



than either of the within-national plots (Figures 4 and 5), and, as expected, the statistics based on the plots bear out these observations. (The correlation for AMT vs. CMT is .977, as compared with the correlations of .991 for AM1 vs. AM2 and .994 for CM1 vs. CM2; the σ_D for AMT vs. CMT is .415, as compared with the σ_D of .233 for AM1 vs. AM2 and .242 for CM1 vs. CM2.) The six distributions of D-values, corresponding respectively to Figures 1-6, are given in Table 4.

Although these differences are clear and indisputable, certain qualifications are in order. The two sets of within-national samples, it is recalled, were selected randomly, and the differences between them are as small as the method of random sampling and the sample size permit. If systematic samples (e.g., samples drawn from somewhat different geographical areas within the United States or within Canada) had been drawn, we would expect the resulting delta plots to be considerably more dispersed than the observed random within-national samples. In some instances, even larger differences in dispersion of deltas could result from geographical sampling for within-national groups than for the random cross-national samples observed here. Therefore, to draw the general conclusion that between-national comparisons are more dispersed than within-national comparisons is not warranted by these data.

In any case, the dispersions observed in the cross-national plots are too small to warrant the conclusion that the American items, for example, are inappropriate for the Canadians, or vice versa. However, since this issue is of itself not central to the purpose of the present study -- which was to find the equations to be used for converting the scaled scores on the SAT to the scale for the CSAT -- it was not pursued in detail. Instead, the distributions of D-values for the cross-national plots (AVT vs. CVT and AMT vs. CMT) were

Table 4

Distributions of D-Values for the Item Delta Plots

	Plots of Verbal Items			Plots of Mathematical Items		
D-Value	AV1 vs. AV2	CV1 vs. CV2	AVT vs. CVT	AM1 vs. AM2	CM1 vs. CM2	AMT vs. CMT
1.2			1			
1.1			1			
1.0			1			1
0.9			0			0
0.8			3			1
0.7		1	0		1	1
0.6	1	1	5	1	0	0
0.5	5	2	6	0	1	2
0.4	5	6	4	2	1	5
0.3	11	4	5	2	4	2
0.2	9	15	8	8	3	3
0.1	8	16	6	5	10	5
0.0	14	17	10	9	4	6
0.1	12	8	8	6	7	6
0.2	9	5	13	4	9	1
0.3	12	7	4	6	3	5
0.4	4	4	6	1	1	0
0.5	2	5	5	0	1	2
0.6	1	2	1	1		1
0.7	2	1	3			2
0.8		1	2			1
0.9			1			0
1.0			1			0
1.1			1			1
No. of Items		95			45	
Mean		0.0			0.0	
S.D.	.287	.296	.458	.233	.242	.415
Plot Correlation	.990	.991	.976	.991	.994	.977

examined to exclude from the set of items used in equating the CSAT to the SAT those items that were considered to be relatively inappropriate to Americans or inappropriate to Canadians. These items were taken to be the "outliers" (i.e., those with high positive or high negative D-values), contributing heavily to the item x nationality interaction. Accordingly, the items removed from the verbal equating sections were those for which $-0.55 \geq D_i \geq 0.55$; the items removed from the mathematical equating section were those for which $-0.52 \geq D_i \geq 0.52$. This process resulted in the removal of 20 of the 95 Verbal items (9 Canadian and 11 American items) and 10 of the 45 Mathematical items (4 Canadian and 6 American items).

Equating

All available Canadian examinees who had taken Section Va as their nonoperational section and all available Canadian examinees who had taken Section Ma as their nonoperational section, exclusive of those who had been part of the item analysis samples (CV1, CV2, CM1, or CM2), were selected for the equating process. This selection resulted in Canadian samples of 3493 for the verbal equating and 3315 for the mathematical equating. Corresponding American samples of 3495 and 3320 (approximately the sizes of the Canadian samples) were selected at random from the available American examinees who had taken Section Vc or Mc, exclusive (as in the choice of the Canadian samples) of all those who had been part of the item analysis samples (AV1, AV2, AM1 or AM2).

The method used for equating the Canadian and American scales involved first the determination of the equation relating the raw (R - W/4) scores on the operational American SAT to the raw (R - W/4) scores on the operational Canadian

CSAT and substituting into that equation the relationship between raw scores and scaled scores for each of the two testing programs to yield a final relationship between American scaled scores and Canadian scaled scores.

The initial conversion between raw scores on the two systems required first an estimation of the mean and variance for the combined American-Canadian sample on each of the two tests in accordance with a method developed by Tucker (Angoff, 1971, p. 580), as follows:

$$\hat{M}_{x_t} = M_{x_\alpha} + b_{xv_\alpha}(M_{v_t} - M_{v_\alpha}) \quad , \tag{1}$$

$$\hat{s}^2_{x_t} = s^2_{x_\alpha} + b^2_{xv_\alpha}(s^2_{v_t} - s^2_{v_\alpha}) \quad , \tag{2}$$

$$\hat{M}_{y_t} = M_{y_\beta} + b_{yv_\beta}(M_{v_t} - M_{v_\beta}) \quad , \tag{3}$$

and

$$\hat{s}^2_{y_t} = s^2_{y_\beta} + b^2_{yv_\beta}(s^2_{v_t} - s^2_{v_\beta}) \quad , \tag{4}$$

where

x or X = the test (CSAT) taken by Group α (Canadian),

y or Y = the test (SAT) taken by Group β (American),

v = the equating test, i.e., scores on the items taken by both Group α and Group β and used in the equating,

and

t = the combined group, $\alpha + \beta$.

The notation, b_{xv} , represents the usual coefficient of regression of variable x on variable v: $b_{xv} = r_{xv}\sigma_x/\sigma_v$. (Similarly, $b_{yv} = r_{yv}\sigma_y/\sigma_v$.) The estimated values, $\hat{M}_{x_t}$, $\hat{M}_{y_t}$, $\hat{s}_{x_t}$, and $\hat{s}_{y_t}$, were then substituted in the equation,

$$\frac{Y - \hat{M}_{y_t}}{\hat{s}_{y_t}} = \frac{X - \hat{M}_{x_t}}{\hat{s}_{x_t}} \quad , \tag{5}$$

to yield the equation,

$$Y = aX + b \quad , \tag{6}$$

converting raw scores on Test X to the scale of the raw scores on Test Y. In this equation, $a = \hat{s}_{y_t}/\hat{s}_{x_t}$ and $b = \hat{M}_{y_t} - a\hat{M}_{x_t}$.

In order to calculate the estimated values given in equations (1) to (4) for the verbal tests, the correlation between the operational test and the 75-item equating section,* as well as the related means and standard deviations, were prepared for each of the two verbal samples, the sample consisting of the 3495 American students and the sample consisting of the 3493 Canadian students. These statistics are given in Table 5. Distributions of the operational verbal and equating verbal tests for the SAT and CSAT are given in Table 6.

The application of the data in Table 5 to equations (1) to (4) resulted in the following estimated values:

$$\hat{M}_{x_t} = 30.6009 \quad ,$$

$$\hat{M}_{y_t} = 35.5304 \quad ,$$

$$\hat{s}^2_{x_t} = 237.0424 \quad ,$$

and

$$\hat{s}^2_{y_t} = 289.9611 \quad .$$

*Recall that 20 verbal "outlier" items had been removed from the original group of 95 items administered to both the American and the Canadian examinees.

Table 5

Correlations Between Operational and Equating Tests

for American and Canadian Samples

Verbal Tests

	American Sample	Canadian Sample
No. of Cases	3495	3493
Operational Test*		
Mean	31.8710	33.9969
Std. Dev.	17.5004	14.1405
Equating Test**		
Mean	25.8489	32.5663
Std. Dev.	15.1437	13.3905
Correlation:		
Opnl. vs. Equating	.9431	.9572

*The operational SAT-verbal contained 90 items: the operational CSAT-verbal contained 75 items.

**The verbal equating test consisted of 75 items (after the removal of 20 "outlier" items).

Table 6

Frequency Distributions for Operational and Equating Sections of the SAT and CSAT

Verbal Tests

Raw (Formula) Score	American Sample		Canadian Sample	
	Operational SAT	Equating Section	Operational CSAT	Equating Section
84 - 86	2			
81 - 83	3			
78 - 80	11			
75 - 77	11			
72 - 74	20		1	1
69 - 71	29	7	10	10
66 - 68	52	9	30	14
63 - 65	56	19	50	35
60 - 62	61	28	54	45
57 - 59	67	35	65	57
54 - 56	116	52	115	79
51 - 53	118	72	137	110
48 - 50	146	88	169	127
45 - 47	145	97	207	203
42 - 44	178	134	210	206
39 - 41	221	181	253	206
36 - 38	205	201	261	297
33 - 35	201	218	265	295
30 - 32	202	244	325	310
27 - 29	212	267	270	318
24 - 26	214	251	235	298
21 - 23	223	237	229	227
18 - 20	204	254	156	193
15 - 17	175	221	158	162
12 - 14	161	196	103	110
9 - 11	157	210	72	88
6 - 8	115	156	57	51
3 - 5	92	150	43	30
0 - 2	48	85	9	15
-3 - -1	33	50	6	4
-6 - -4	16	29	3	1
-9 - -7	1	3		1
-12 - -10		1		
No. of Cases	3495	3495	3493	3493
Mean	31.8710	25.8489	33.9969	32.5663
Std. Dev.	17.5004	15.1437	14.1405	13.3905

These values were applied to equation (5) to yield the equation,

$$Y = 1.1060\,X + 1.6857 \quad , \tag{7}$$

which permitted the conversion of scores from the raw score scale of the CSAT-verbal to the raw score scale of the SAT-verbal. The derivation of the conversion from the CSAT-verbal reporting scale to the SAT-verbal reporting scale is developed as follows:

The linear equation,

$$S_c = AX + B \quad , \tag{8}$$

is the equation by which raw scores on the form of the CSAT used in January 1971 (X) are converted to the CSAT reporting scale (Sc). Similarly, the linear equation,

$$S_a = A'Y + B' \quad , \tag{9}$$

is the equation by which raw scores on the form of the SAT used in January 1971 (Y) are converted to the SAT scale (Sa). Expressing equation (8) in terms of $X \left[X = \frac{S_c - B}{A}\right]$ and equation (9) in terms of $Y \left[Y = \frac{S_a - B'}{A'}\right]$, and substituting in equation (6) results in the equation, $\frac{S_a - B'}{A'} = a\left[\frac{S_c - B}{A}\right] + b$, which, when simplified, becomes

$$S_a = \frac{aA'}{A} S_c + A'b + B' - \frac{aA'B}{A} \quad . \tag{10}$$

Equation (10) is a linear equation with slope equal to aA'/A and intercept equal to $A'b + B' - \frac{aA'B}{A}$, and may be used to convert verbal or mathematical scores from the January 1971 converted score scale for the CSAT to corresponding scores on the SAT scale.

In order to derive the numerical conversion from the CSAT-verbal scale to the SAT-verbal scale, the following numerical values for the slopes and intercepts of equations (6), (8), and (9) were applied to the constants in equation (10):

$a = 1.1060$; $b = 1.6857$ [from equation (6)] ,

$A = 7.2533$; $B = 256.0000$ [from equation (8)] , and

$A' = 6.3656$; $B' = 229.2008$ [from equation (9)] .

The resulting conversion equation for the verbal test is, therefore,

$$S_a = 0.9706\, S_c - 8.5528 \quad . \tag{11}$$

The inverse equation, for converting the SAT-verbal scale to the CSAT-verbal scale, is:

$$S_c = 1.0303\, S_a + 8.8119 \quad . \tag{11a}$$

The following equivalencies were determined from equations (11) and (11a):

From Equation (11):

CSAT-verbal Score	Equivalent SAT-verbal Score
800	768
700	671
600	574
500	477
400	380
300	283
200	(186)*

From Equation (11a):

SAT-verbal Score	Equivalent CSAT-verbal Score
800	(833)*
700	730
600	627
500	524
400	421
300	318
200	215

More detailed conversion tables for the verbal tests are provided in Appendix I.

*Scores lower than 200 or greater than 800 on the SAT are reported as 200 or 800, respectively. The same procedures will be employed on all future forms of the CSAT-

Directly parallel procedures were followed in deriving the equation for converting scaled scores on the CSAT-mathematical to scaled scores on the SAT-mathematical. In order to calculate the estimated values given in equations (1) to (4) for the mathematical tests, the correlation between the operational test and the 35-item equating section,* as well as the related means and standard deviations, were prepared for each of the two mathematical samples, the sample consisting of the 3320 American students and the sample consisting of the 3315 Canadian students. These statistics are given in Table 7. Distributions of the operational mathematical and equating mathematical tests for the SAT and CSAT are given in Table 8.

The application of the data in Table 7 to equations (1) to (4) resulted in the following estimated values:

$$\hat{M}_{x_t} = 15.7479 \quad ,$$

$$\hat{M}_{y_t} = 26.2079 \quad ,$$

$$\hat{s}^2_{x_t} = 88.7126 \quad ,$$

and

$$\hat{s}^2_{y_t} = 162.9277 \quad .$$

These values were then applied to equation (5) to yield the equation,

$$Y = 1.3552X + 4.8663 \quad , \tag{12}$$

which permitted the conversion of scores from the raw score scale of the CSAT-mathematical to the raw score scale of the SAT-mathematical. In order to derive the conversion from the CSAT-mathematical scale to the SAT-mathematical

*Recall that 10 mathematical "outlier" items had been removed from the original group of 45 items administered to both the American and the Canadian examinees.

Table 7

Correlations Between Operational and Equating Tests

for American and Canadian Samples

Mathematical Tests

	American Sample	Canadian Sample
No. of Cases	3320	3315
Operational Test*		
Mean	22.8536	18.2477
Std. Dev.	13.1620	8.3968
Equating Test**		
Mean	12.0973	16.7493
Std. Dev.	8.5261	7.2081
Correlation: Opnl. vs. Equating	.9348	.9219

*The operational SAT-mathematical contains 60 items; the operational CSAT-mathematical contains 40 items.

**The mathematical equating test consisted of 35 items (after the removal of 10 "outlier" items).

Table 8

Frequency Distributions for Operational and Equating Sections of the SAT and CSAT

Mathematical Tests

Raw (Formula) Score	American Sample		Canadian Sample	
	Operational SAT	Equating Section	Operational CSAT	Equating Section
58 - 59	8			
56 - 57	8			
54 - 55	19			
52 - 53	13			
50 - 51	31			
48 - 49	39			
46 - 47	42			
44 - 45	70			
42 - 43	78			
40 - 41	81		4	
38 - 39	105		33	
36 - 37	98		23	
34 - 35	135	8	80	25
32 - 33	156	24	100	24
30 - 31	170	46	125	91
28 - 29	164	82	175	122
26 - 27	162	91	173	149
24 - 25	180	129	200	210
22 - 23	167	161	194	260
20 - 21	190	156	255	301
18 - 19	188	221	306	302
16 - 17	176	212	304	338
14 - 15	153	234	311	368
12 - 13	132	261	276	313
10 - 11	179	286	257	267
8 - 9	146	293	198	213
6 - 7	100	263	128	136
4 - 5	121	263	89	104
2 - 3	81	220	48	53
0 - 1	62	218	25	29
-2 - -1	35	112	9	6
-4 - -3	19	32	2	3
-6 - -5	7	7		1
-8 - -7	5	0		
-10 - -9		1		
No. of Cases	3320	3320	3315	3315
Mean	22.8536	12.0973	18.2477	16.7493
Std. Dev.	13.1620	8.5261	8.3968	7.2081

scale, the following numerical values from the slopes and intercepts of equations (6), (7), and (8) were applied to the constants in equation (9):

$a = 1.3552$; $b = 4.8663$ [from equation (6)] ,

$A = 13.9550$; $B = 242.0000$ [from equation (8)] , and

$A' = 8.5922$; $B' = 273.2077$ [from equation (9)] .

The resulting conversion equation for the mathematical test is, therefore,

$$S_a = 0.8347\ S_c + 113.0219 \quad . \tag{13}$$

The inverse equation, for converting the SAT-mathematical scale to the CSAT-mathematical scale, is:

$$S_c = 1.1980\ S_a - 135.4042 \quad . \tag{13a}$$

The following equivalencies were determined from equations (13) and (13a):

From Equation (13):

CSAT-math Score	Equivalent SAT-math Score
800	781
700	697
600	614
500	530
400	447
300	363
200	280

From Equation (13a):

SAT-math Score	Equivalent CSAT-math Score
800	(823)*
700	703
600	583
500	464
400	344
300	224
200	(104)*

More detailed conversion tables for the mathematical tests are provided in Appendix II.

*Scores lower than 200 or greater than 800 will be reported as 200 and 800, respectively, on all future forms of the CSAT. The same procedures are employed on all forms of the SAT.

The incidental observation is made that the verbal and mathematical tests of the CSAT and the SAT are all quite similar in average difficulty. The estimated formula score means for the verbal tests for the combined SAT-CSAT verbal sample both represent a mean item difficulty of about 0.4 (0.41 for CSAT-verbal and 0.39 for SAT-verbal) in terms of the same combined verbal equating group. The estimated formula score means for the mathematical tests show a slightly greater difference; the mean item difficulty on CSAT-mathematical estimated for the combined mathematical equating group is 0.39 while the mean item difficulty on SAT-mathematical for the same group is slightly greater -- 0.43 -- but still very much in the same range. Although the reason for this difference can only be a matter for speculation, it is possible that part of it, at least, is attributable to the more generous timing allowances in the CSAT.

Ratios of the number of items to the estimated standard deviation for the combined samples were also obtained. These ratios were 4.87 for CSAT-verbal and 5.29 for SAT-verbal, and 4.25 for CSAT-mathematical and 4.70 for SAT-mathematical. These figures suggest the possibility that the mathematical tests yield more discriminations and may therefore be more reliable per unit test than the verbal tests. The figures also suggest the possibility that the tests of the CSAT are slightly more reliable than those of the SAT per unit test.

Conversion Results

From equations (11) or (11a) and (13) or (13a), and from the conversion tables that follow those equations it is clear that, although the Canadian and American scales are similar in the region of the 500-score, there are detailed differences. The equivalent SAT-verbal scores are lower than the corresponding CSAT-verbal scores throughout the scaled score range, the differences ranging from -14 points at the CSAT-verbal score of 200 to -32 points at the CSAT-verbal

score of 800. At 500 the difference is midway between these extremes, -23 points.

The differences in the mathematical scales are greater, however, and for the most part in the opposite direction. At the CSAT-mathematical scaled score of 200, the difference is as much as 80 points; thus, a student with a score at the 200-score level on CSAT is estimated to be equivalent in ability to a student with a 280 score on SAT. At the opposite end of the scale, the difference is much smaller, but in the other direction; a student at the 800-score level on the CSAT scale would have a score equivalent to a score less than 800 on the SAT scale, namely 781. As expected, the difference at 500 is about midway between these; a score of 500 on CSAT-mathematical is equivalent to a score of 530 on SAT-mathematical.

Summary

In this equating study the form of the CSAT administered in January 1971 was equated to the form of the SAT also administered in January 1971. The design for equating involved the administration of a one-half hour operational verbal section and a one-half hour operational mathematical section of the CSAT as "experimental" (nonoperational) tests to random samples of the January 1971 SAT group. Similarly, a one-half hour operational verbal section and a one-half hour operational mathematical section of the SAT were administered as "experimental" (nonoperational) tests to random samples of the January 1971 CSAT group. As a result of these symmetrical "cross-over" administrations, random samples of the American and the Canadian groups took the same two sections of verbal test materials. Other random samples of the American and Canadian groups took the same two sections of mathematical test materials. These common sections, representing 95 verbal items and 45 mathematical items,

were then item analyzed, and item difficulty (delta) plots were made to identify "outlier" items -- items that were not equally appropriate for the two groups of examinees. As a result of these analyses, 20 verbal and 10 mathematical items were identified as "outliers" and removed from the equating sections. Data on new samples were developed with the remaining 75 verbal equating items and 35 mathematical equating items to equate raw scores on that form of the CSAT to corresponding scores on that form of the SAT. With these equations in hand and with the equations converting each set of raw scores to its own reporting scale, conversion equations were developed relating the verbal and mathematical scales for the CSAT to the corresponding scales for the SAT.

Reference

Angoff, W. H., "Scales, Norms, and Equivalent Scores," in R. L. Thorndike (Ed.) Educational Measurement (2nd ed.). Washington: American Council on Education, 1971, pp. 508-600.

Appendix I

Conversions Between CSAT-Verbal Scaled Scores and SAT-Verbal Scaled Scores

From Equation (11): $S_a = 0.9706\ S_c - 8.5528$

CSAT-Verbal Score	Equivalent SAT-Verbal Score	CSAT-Verbal Score	Equivalent SAT-Verbal Score
800	768	490	467
		480	457
790	758	470	448
780	749	460	438
770	739	450	428
760	729	440	419
750	719	430	409
740	710	420	399
730	700	410	389
720	690	400	380
710	681		
700	671	390	370
		380	360
690	661	370	351
680	651	360	341
670	642	350	331
660	632	340	321
650	622	330	312
640	613	320	302
630	603	310	292
620	593	300	283
610	584		
600	574	290	273
		280	263
590	564	270	254
580	554	260	244
570	545	250	234
560	535	240	224
550	525	230	215
540	516	220	205
530	506	210	(195)*
520	496	200	(186)*
510	486		
500	477		

From Equation (11a): $S_c = 1.\ S_a + 8.8119$

SAT-Verbal Score	Equivalent CSAT-Verbal Score	SAT-Verbal Score	Equivalent CSAT-Verbal Score
800	(833)*	490	514
		480	503
790	(823)*	470	493
780	(812)*	460	483
770	(802)*	450	472
760	792	440	462
750	782	430	452
740	771	420	442
730	761	410	431
720	751	400	421
710	740		
700	730	390	411
		380	400
690	720	370	390
680	709	360	380
670	699	350	369
660	689	340	359
650	679	330	349
640	668	320	339
630	658	310	328
620	648	300	318
610	637		
600	627	290	308
		280	297
590	617	270	287
580	606	260	277
570	596	250	266
560	586	240	256
550	575	230	246
540	565	220	235
530	555	210	225
520	545	200	215
510	534		
500	524		

*[illegible] 200 or greater than 800 on the SAT are reported as 200 or 800, respectively. The same pro-

Appendix II

Conversions Between CSAT-Mathematical Scaled Scores and SAT-Mathematical Scaled Scores

From Equation (13): $S_a = 0.8347\ S_c + 113.0219$

CSAT-Math Score	Equivalent SAT-Math Score	CSAT-Math Score	Equivalent SAT-Math Score
800	781	490	522
		480	514
790	772	470	505
780	764	460	497
770	756	450	489
760	747	440	480
750	739	430	472
740	731	420	464
730	722	410	455
720	714	400	447
710	706		
700	697	390	439
		380	430
690	689	370	422
680	681	360	414
670	672	350	405
660	664	340	397
650	656	330	388
640	647	320	380
630	639	310	372
620	631	300	363
610	622		
600	614	290	355
		280	347
590	605	270	338
580	597	260	330
570	589	250	322
560	580	240	313
550	572	230	305
540	564	220	297
530	555	210	288
520	547	200	280
510	539		
500	530		

From Equation (13a): $S_c = 1.1980\ S_a - 135.4042$

SAT-Math Score	Equivalent CSAT-Math Score	SAT-Math Score	Equivalent CSAT-Math Score
800	(823)*	490	452
		480	440
790	(811)*	470	428
780	799	460	416
770	787	450	404
760	775	440	392
750	763	430	380
740	751	420	368
730	739	410	356
720	727	400	344
710	715		
700	703	390	332
		380	320
690	691	370	308
680	679	360	296
670	667	350	284
660	655	340	272
650	643	330	260
640	631	320	248
630	619	310	236
620	607	300	224
610	595		
600	583	290	212
		280	200
590	571	270	(188)*
580	559	260	(176)*
570	547	250	(164)*
560	535	240	(152)*
550	523	230	(140)*
540	512	220	(128)*
530	500	210	(116)*
520	488	200	(104)*
510	476		
500	464		

*Scores lower than 200 or greater than 800 will be reported as 200 and 800, respectively, on all future