A Comparison of Somatotype Methods [1]

BARBARA HONEYMAN HEATH AND J. E. LINDSAY CARTER
5 Via Joaquin, Monterey, California and Physical Education Research Laboratory, San Diego State College, San Diego, California

ABSTRACT In order to compare Parnell's and Heath's somatotype methods, the authors independently somatotyped a series of 59 adult male and 61 adult female subjects, (1) using the criteria of Heath's method, (2) using the criteria of Parnell's method, and (3) taking into consideration tentatively adapted Parnell criteria in addition to Heath's criteria.

The authors conclude that when use similar rating criteria their mean differences are smaller, their overall correlations are similar, and their percentage agreements to a half-unit are higher (96%) than for comparisons reported by other investigators.

The study considers the potentially important relationships of measurements of subcutaneous fat to ratings of the first component. The similarity of distributions of subcutaneous fat measurements and of first component ratings in selected samples suggest important interrelationships among ratings of the first component, height/weight ratios and subcutaneous fat measurements.

The authors feel: (1) that Parnell's method fails to modify the basic weaknesses in Sheldon's somatotype method; and (2) that analyses of the anthropometric data basic to Parnell's method, if guided by the criteria of Heath's method, will further objectify and simplify Heath's method, will improve agreement among independent raters, and will increase the usefulness of somatotyping as a research instrument.

The authors of this study have had long experience with two somatotype methods, each designed to clarify, modify, objectify and validate somatotype methodology. Carter's ('64, '65a) study of the physiques of male and female physical education teachers in training in New Zealand provides the data required for somatotype rating by both methods — Heath's ('63) and Parnell's ('54, '58). It also furnishes an opportunity to compare two methods, together with the possibility of additional clarification and modification of somatotype methodology.

Heath's and Parnell's methods are based upon modifications of the Sheldonian ('40, '54) system of describing variations of human physique. Parnell ('54, '58) suggested that physical anthropometry used in conjunction with somatotype photographs could objectify somatotype ratings. He chose three sets of measurements: (1) bone diameters, (2) muscle girths, and (3) skinfolds or subcutaneous fat measurements. Figure 1 is a reproduction of Parnell's so-called *M.4 deviation chart*, constructed for recording and interpreting these three sets of measurements, and other data used in assigning somatotype ratings. (The construction and rationale of the M.4 chart is discussed further in the section, *Modification and Adaptation of* M.4 Data in, FINDINGS AND DISCUSSION — p. 95.) It is designed to include "all necessary information, thus dispensing with further tables of reference;" and to "correspond as closely as possible with Sheldon's estimate of somatotype" (Parnell, '58, p. 19). Examination of figure 1 shows that this "correspondence" is achieved by retaining Sheldon's arbitrary, closed 7-point scale for rating somatotype components; by adjusting the intervals on the fat scale to conform with Sheldon's reported means for the first component; by rating the third component so that "it corresponds closely to the median of the range for each component given in Sheldon's ('54) tables" (Parnell, '58 — p. 20); and in making ratings of the first and third components to conform with age-scaled interpretations of skinfolds and height/weight ratios. It is difficult to reconcile Parnell's M.4 chart procedures for arriving at somatotype ratings with his statement that his emphasis is on phenotype, "which can be measured at a given time, rather

[1] The authors are grateful to the Wenner-Gren Foundation for Anthropological Research and to the San Diego State College Foundation for grants, which in part supported this study.

ADULT DEVIATION CHART OF PHYSIQUE (Male standards)

NAME ______ AGE ______ DATE ______

OCCUPATION ______ Married/Single Ch: M ______ F ______ REF. No. ______

Fat:	(Skinfold)	Age	Total 3 skinfold measurements												
(mm.)	Over triceps	16–24	10	12	14	17	20	24	29	36	45	57	73	93	114
	Subscapular	25–34	12	14	17	20	24	30	38	48	60	74	94	114	+
	Suprailiac	35–44	13	16	19	22	27	35	44	55	68	87	109	+	+
	Total fat	45–54	14	17	20	23	29	37	47	61	74	95	118	+	+
Endomorphy estimate			1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7

Height (in.)	55.0	56.5	58.0	59.5	61.0	62.5	64.0	65.5	67.0	68.5	70.0	71.5	73.0	74.5	76.0	77.5	79.0	80.5
Bone: Humerus	5.34	5.49	5.64	5.78	5.93	6.07	6.22	6.37	6.51	6.65	6.80	6.95	7.09	7.24	7.38	7.53	7.67	7.82
(cm.) Femur	7.62	7.83	8.04	8.24	8.45	8.66	8.87	9.08	9.28	9.49	9.70	9.91	10.12	10.33	10.53	10.74	10.95	11.16
Muscle: Biceps	24.4	25.0	25.7	26.3	27.0	27.7	28.3	29.0	29.7	30.3	31.0	31.6	32.2	33.0	33.6	34.3	35.0	35.6
(cm.) Calf	28.5	29.3	30.1	30.8	31.6	32.4	33.2	33.9	34.7	35.5	36.3	37.1	37.8	38.6	39.4	40.2	41.0	41.8

First estimate of mesomorphy	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	
Correction for fat (T.F. mm.)	12	15	18	22	27	33	40	48	57	68	83	100	120	140
Age: 16–24	+0.5	+0.5	+0.25	+0.25	0	−0.25	−0.5	−1	−1.5	−2	−2.5	−3	−4	
25–34	(+0.5)	+0.5	+0.25	+0.25	0	−0.25	−0.5	−0.75	−1.25	−1.75	−2.25	−2.75	−3.5	−4
35+	(+0.5)	(+0.5)	+0.25	+0.25	0	−0.25	−0.25	−0.5	−1	−1.5	−2	−2.5	−3	−3.5
Mesomorphy (corrected estimate)	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	

Weight	Wt. lb.	H.W.R.	Age													
Present			18	12.1	12.3	12.5	12.7	12.9	13.1	13.3	13.5	13.7	13.8	14.0	14.2	14.4
H. K. W.			23	11.7	12.0	12.2	12.5	12.8	13.0	13.2	13.4	13.6	13.8	14.0	14.2	14.4
Usual			28	11.5	11.8	12.1	12.4	12.6	12.8	13.0	13.3	13.5	13.7	13.9	14.2	14.4
At 18 years			33	11.3	11.7	12.0	12.3	12.5	12.7	12.9	13.2	13.4	13.6	13.9	14.1	14.4
At 23 years			38	11.2	11.5	11.8	12.1	12.4	12.6	12.8	13.1	13.3	13.6	13.9	14.1	14.4
Recent change			43+	11.1	11.4	11.7	12.0	12.3	12.6	12.8	13.1	13.3	13.6	13.9	14.1	14.4
Ectomorphy				1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7

Fig. 1 M.4 deviation chart with equivalent skinfold measurements for Harpenden calipers, from Parnell ('58, p. 21).

than on Sheldon's somatotype" (Parnell, '58 — p. 20). Apparently the M.4 chart was constructed on the basis of studies of more than 2,000 male and almost 700 female students at the Universities of Oxford and Birmingham, a study of about 800 school children in City of Oxford schools, and some small selected samples.

Heath's ('63) method has opened the rating scale at both ends, eliminated extrapolations for age, and established a linear relationship between somatotype ratings and height/weight ratios. The somatotype ratings are phenotypes. In cross-sectional studies the elimination of corrections for age opens the way to considering the probability that several somatotypes, or phenotypes, are possible for each individual. Longitudinal data (Heath — unpublished) show that there *are* phenotypic changes in both childhood and adult life, and that they can be described meaningfully by somatotype ratings. Heath's method requires discriminating photo-anthroposcopy checked against her empirical table of somatotypes distributed in linear relationships to height/weight ratios (table 1).

The wide use of both Parnell's and Heath's methods in a variety of investigations warrants a comparison of the two. Parnell ('54, '57, '58) reports his own and other studies. Cortes ('61), Donnan ('59) and Carter ('64, '65a, b, c) report their use of Parnell's method. Heath's method is applied in a number of studies (Allen,

TABLE 1

Distribution of somatotypes on the criterion of height/$\sqrt[3]{weight}$, for both sexes at all ages, which maintains a linear relationship between height/weight ratios and somatotype component ratings

Ratio index	Somatotypes						
11.40	951						
11.60	941, 851						
11.80	841, 751, 481						
12.00	741, 471						
	651, 561						
12.20	731, 371						
	641, 461	551					
12.40	721, 271, 631	732, 372, 642					
	*361, 541, 451	462, 552					
12.60	711, 171, 621	*362, 542, 452					
	*261	722, 272, 632					
12.80		712, 172, 622	633, 363				
		532,*352,*262	543, 453				
		442					
13.00		612, 162	623,*263, 443				
		*252, 522	533,*353				
13.20			*253, 523,*433	*444			
			613, 163,*343	534, 354			
				*254			
13.40				614, 524			
				434,*344			
				424, 334			
13.60				514, 154	*345		
				244	525, 435		
13.80					335, 245		
					425, 515		
14.00					235, 325		
					145, 415	336, 426	
14.20					225	236, 326	
						416	
14.40						316, 136	
						226	327
14.60						126, 216	227
							317
14.80							127
							217
15.00							117

* Somatotypes accounting for over one-half of the males, Sheldon ('54, p. 30).

'65; Behnke, '57; Clarke, '61; Heath, '61, '63, '65; Irving, '59; Olsen, '61; Seltzer, '64; Tanner, '64; Walker, '62).

The primary purpose of this study is to compare and evaluate two pairs of independent somatotype ratings of one series of subjects. The first pair consists of Heath's ratings applying her method alone and Carter's ratings applying Parnell's M.4 method. The second pair consists of Heath's ratings plus use of the M.4 chart data, and Carter's ratings plus use of Heath's method.

The secondary purpose is evaluation of the influence of M.4 data on Heath's ratings and the influence of applying Heath's method to Carter's ratings.

This study is motivated by the desire further to modify Heath's method, in order to aid objectification of somatotype ratings, and to simplify methodology for easier reproduction and wider use.

PROCEDURE

Subjects. One hundred twenty physical education students at the University of Otago, New Zealand — 59 males, 61 females, ages 17–23 years. For other data regarding this series, see Carter ('64, '65a).

METHODS

The subjects were photographed in the three standard somatotype poses. The males were unclothed except for athletic supports. The females were only nonelastic brassieres and brief panties. Age, height, weight, weight history, subcutaneous fatfolds, muscle girths, and bone diameters were measured by one of the authors (Carter) at the time of photography. Intra-observer reliabilities for the anthropometric measurements ranged from, $r = +0.94$ to $r = +0.99$ (Carter and Rendle, '65b).

Carter made M.4 deviation chart ratings on the combined male and female sample and assigned Sheldonian ratings on the male portion of the series before Heath had seen the series and data. After consultation with Heath, Carter made ratings, using Heath's method plus the M.4 data. The photographs, together with age, height, weight and height/weight ratios were then sent to Heath for her independent ratings. Upon completion of these ratings, Heath was introduced to the M.4 deviation chart, with which she was entirely unfamiliar. She then re-rated the series in the light of the additional information. Thus, in all, five different ratings were made on the male portion of the series, and four on the female portion.

Comparison of rating methods. Means, minima and maxima for each of the components rated by different methods are presented in table 2.

The two major comparisons selected for further analysis are: (1) Heath's rating minus the M.4 rating (H – M.4), and (2) Heath's rating plus M.4 data minus Carter's using M.4 data plus Heath's method *(H + M.4* minus C, *M.4 + H).* Comparisons were made on the bases of absolute differences between means, the t-test of significance of the differences between means of two correlated samples, the percentage agreement between series of ratings, and the Pearson product-moment correlation between the ratings. For the two-tailed t-test a t (df = n – 1) $\pm$ 2.66 for the male and female samples, and t (df = n – 1) $\pm$ 2.62 for the total sample, were required to reject the null hypothesis at 0.01 level of confidence. A summary of these comparisons is contained in table 3.

Table 4 shows a tabulation of changes by components when the raters both used M.4 data in arriving at ratings.

FINDINGS AND DISCUSSION

The following general points should be kept in mind when evaluating the analysis of the data:

1. The subjects in the sample studied represent a select group, not typical of the general population. There is relatively low variability of the component ratings. The low variability makes some comparative tests more demanding, wherefore any relationships demonstrated may be expected to "improve" (statistically) when the samples are more variable. Product-moment correlations may be limited in value in a sample in which the majority of ratings for each component fall within one and one-half to two rating units of one another.

2. Four analysis techniques are used in this study for data interpretation (mean difference, t-ratio, percentage agreement, and product-moment correlation), plus

TABLE 2

Descriptive statistics on New Zealand Somatotype Series rated by different methods

Rating method	Sex	First component M	First component Min.	First component Max.	Second component M	Second component Min.	Second component Max.	Third component M	Third component Min.	Third component Max.
Heath by H	Male = 59	3.34	2.0	5.0	5.15	3.5	6.5	2.33	1.0	5.0
	female = 61	4.39	3.0	6.0	4.35	3.0	5.5	2.48	0.5	6.0
	m+f = 120	3.86	2.0	6.0	4.75	3.0	6.5	2.41	0.5	6.0
Parnell M.4 by C	male	3.61	2.5	5.5	5.17	3.0	6.5	2.81	1.0	5.0
	female	4.46	2.5	5.5	3.86	2.0	5.5	2.66	1.0	6.0
	m+f	4.04	2.5	5.5	4.50	2.0	6.5	2.73	1.0	6.0
M.4+Sheldon by C	male	3.45	2.5	5.0	5.15	3.5	6.5	2.79	1.0	5.0
Heath+M.4 by C	male	3.47	2.5	5.5	5.20	3.5	6.5	2.49	1.0	4.5
	female	4.61	2.5	6.0	4.02	2.5	5.5	2.52	1.0	6.0
	m+f	4.05	2.5	6.0	4.60	2.5	6.5	2.51	1.0	6.0
Heath+M.4 by H	male	3.42	2.0	5.5	5.17	4.0	6.5	2.44	1.0	5.0
	female	4.45	2.5	5.5	4.23	3.0	5.0	2.42	0.5	5.5
	m+f	3.95	2.0	5.5	4.69	3.0	6.5	2.43	0.5	5.5

NOTE: The effect of Heath's open rating scale is reflected in minima recorded for ratings by Heath. In the female series, Heath made ratings of one-half-unit in the third component. Neither Sheldon's nor Parnell's method allows a rating less than one — or more than seven.

TABLE 3

Comparisons between different methods of somatotype rating. N.Z. P.E. Data

Method	Component	Sex	D	t-ratio	Agreement ± one-half-unit	r_{xy}
					%	
H — M.4 Heath — Parnell's M.4 chart	first	male	−0.27	−3.21 [1]	77.9	0.58
		female	−0.07	−1.00	82.8	0.41
		m+f	−0.18	—	80.0	0.74
	second	male	−0.02	—	88.1	0.88
		female	+0.49	+7.00 [1]	68.8	0.80
		m+f	+0.25	—	78.3	0.89
	third	male	−0.48	−9.80 [1]	76.3	0.92
		female	−0.18	−4.00 [1]	95.1	0.94
		m+f	−0.32	—	85.8	0.92
$H_{+M.4} — C_{M.4+H}$	first	male	−0.05	—	94.8	0.88
		female	−0.16	−3.56 [1]	95.1	0.86
		m+f	−0.10	—	95.0	0.92
	second	male	−0.03	—	100.00	0.80
		female	+0.21	+4.57 [1]	93.3	0.87
		m+f	+0.09	—	96.7	0.89
	third	male	−0.04	—	96.6	0.70
		female	−0.10	−2.38	98.3	0.95
		m+f	−0.08	−2.42	97.5	0.88

[1] t-ratio significant at the 0.01 level.

frequency distributions. A comprehensive analysis appears to be warranted in view of the limitations of each technique considered separately.

3. With respect to adequate reliability of ratings, our own experience — and that of other criterion raters, as reported in Sheldon et al. ('40), Parnell ('54, '58), Tanner ('54, '64), Hunt and Barton ('59), Damon et al. ('60) — suggest that the following criteria are important: (a) the mean differences for a given component ought to be less than one-fifth of a rating unit; (b) the difference should be statis-

TABLE 4

Changes by components with M.4 data available (%)

Rater	Component	Sex	− 1	− ½	0	+ ½	+ 1	Changes
$H_{+ M.4}$	first	male		13.6	59.3	25.4	1.7	40.7
		female		16.4	52.5	31.1		47.5
		m+f		15.0	55.8	28.3	1.6	44.9
	second	male		11.9	72.9	15.3		27.1
		female	1.6	16.4	70.5	11.5		29.5
		m+f	0.8	14.2	71.7	13.3		28.3
	third	male	1.7	1.7	71.2	23.7	1.7	28.8
		female		19.7	73.8	6.6		26.2
		m+f	0.8	10.8	72.5	15.0	0.8	28.3
$C_{M.4 + H}$	first	male	6.8	16.9	74.6	1.7		25.4
		female	1.6	3.3	62.3	29.5	3.3	37.7
		m+f	4.2	10.0	67.5	15.8	1.7	32.5
	second	male		8.5	78.0	13.6		22.1
		female		3.3	72.1	19.7	4.9	27.9
		m+f		5.8	75.0	16.7	2.5	25.0
	third	male	6.8	49.2	42.4	1.7		57.6
		female		21.3	78.7			21.3
		m+f	3.3	35.0	60.8	0.8		39.2
$C_{M.4 + S}$	first	male		33.9	64.4	1.7		35.6
	second	male		6.8	89.8	3.4		10.2
	third	male		8.5	86.4	5.1		13.6
	Total	male		16.4	80.2	3.4		19.8

tically insignificant; (c) the percentage agreement ± ½-unit ought to be approximately 90% or better; (d) the correlation (r_{xy}) ought to approach + 0.90.

When the mean difference is significant and the percentage agreement is low, a high correlation indicates that a correction to the mean may be a useful procedure in comparative studies. When a mean difference approximates one-half unit, individual corrections may be made and the percentage agreement re-calculated to determine whether the level reached is acceptable. When the means are similar and the percentage agreement is high, low correlations may be due to low component variability.

Table 2 shows that in both the male and female series the same relative component dominance is maintained for all rating methods. But within these patterns Heath tends to rate the first and third components lower, and to rate the second component higher in the female series than is the case for other rating methods.

Attention is now directed to comparisons of Heath's and Parnell's methods of somatotype rating, to the question of objectivity of independent raters, and to the influence of M.4 data upon the ratings.

Heath vs. M.4. How do the ratings compare when a group of athletically oriented young men and women are somatotyped according to Heath's method and according to Parnell's M.4 chart? Table 3 shows that Heath's means are significantly lower than M.4 means for the first component in males, higher for the second component for females, and lower in the third component for both males and females. The differences are approximately one-fifth to one-half rating unit.

The percentage agreements ± ½-unit are relatively low for the both the first (80.0%) and second (78.3%) components, and moderate (85.8%) for the third component. In these comparisons 18.6% of the total differences exceed ± ½-unit. However, when systematic differences are allowed for, the agreements improve markedly. When we take into account the mean difference of + ½-unit for the second component for females, the agreement increases to 86.9%, and the total agreement for the second component increases to 87.5%. Similarly, when we take into account the mean differences of − ½-unit for the third component in males, the agreement for the males increases to 96.6%, and the total agreement to 95.8%.

Product-moment correlations between ratings by the M.4 chart and by Heath's method are low for the first component ($r = +0.74$), but relatively high for the second ($r = +0.89$) and the third ($r = +0.92$) components.

Parnell ('54) cites agreement to one-half rating unit for 90% of the component ratings for comparison of M.4 chart ratings and ratings based on Sheldon's photometric tables; and 87.3% agreement between M.4 chart ratings and photoscopic ratings. Absolute mean differences ranged from 0.06 to 0.17 in the first, from 0.02 to 0.22 in the second comparison. In a later study, Parnell ('58) found discrepancies exceeding one-half unit in 3.9% of the first component ratings, in 10.6% of the second component ratings, and in 9.7% of the third component ratings. He also notes that, "The present M.4 method and photoscopic interpretation differ slightly more, not particularly in the first and third components, but in the second component discrepancies exceeding a half unit occurred in about one-third" (p. 22).

Tanner's ('64) data made possible a similar comparison between M.4 ratings and a criterion rating (Tanner's) on first and third component ratings of a select male athletic group — 62 British Empire Games wrestlers and weightlifters. For the first component the mean difference = 0.08 (T − M.4); $r = +0.78$; percentage agreement ± ½-unit = 82.3%. For the third component the mean differences = − 0.42 (T − M.4); $r = +0.87$; percentage agreement ± ½-unit =72.0%. If we assume a mean difference of ½-unit, the corrected percentage agreement is 87.0%.

While the differences between Heath and M.4 for males are larger than those between Tanner and M.4 for the first component, the differences between both Heath and Tanner and M.4 are similar for the third component. Since Tanner and M.4 are purportedly rating the same thing on the same scales, one would expect Tanner to be closer to M.4 than Heath to M.4. By using regression equations based upon anthropometry, Damon et al. ('62) found that they could predict component ratings to ½-unit in 80.0% of the cases, and to within one unit in 97.0% of the cases, for both white and Negro men chiefly in the third decade of life.

It is apparent then that while the relationships between Heath and M.4 are poor for the first component, the second and third components reach acceptable levels of percentage agreement when corrections are applied. However, overall, the relationships are somewhat lower than those Parnell reports, and there are apparent differences between ratings for males and females. The best agreements are for the second component in males, and for the third component in females. Undoubtedly some of the differences are due in part to use of age-scaled corrections, which are important in the Parnell and Sheldon systems. Heath uses a universal reference scale, i.e. the same criteria are applied for both sexes at all ages. On the other hand, it could be argued that Heath rates males and females differently from the M.4 chart.

Objectivity of raters. When they use the same method, how closely do independent raters agree? We looked for the answer in comparison of our ratings of the New Zealand series after both authors had used the M.4 data and Heath's method together. Table 2 indicates that among all the rating differences the lowest are the mean differences between Heath's and Carter's component ratings. Although the differences are small, Heath's mean ratings of females are significantly lower for the first component than Carter's, and higher than Carter's for the second component (table 3). The magnitudes of the differences are approximately one-sixth to one-fifth of a rating unit. These small differences are probably due in part to Carter's giving some credence to the M.4 chart corrections for fat in rating the second component.

In both male and female series the percentage agreement ± ½-unit is uniformly high (93.3–100%) for all three components. Only 3.6% of the total differences exceed one-half unit.

In view of the uniformly high overall correlations between Heath and Carter for all three components (r's = + 0.88 to + 0.92), the relatively low correlations of + 0.80 for the second and + 0.70 for the third component for the male ratings appear to be anomalous when the almost identical means and high percentage agreements are considered. Perhaps the explanation lies in the small variability of a selected sample.

Sheldon et al. ('40) reported correlations of r = + 0.92 to + 0.94 between raters for the three components. Tanner ('54) found 90% ± ½-unit overall agreement in the anthroposcopic somatotyping of three experienced raters (Dupertuis, Honeyman Heath, Tanner) and ". . . reliability coefficients of about 0.83 for the first two components and 0.92 for the third, when the ratings cover the full range of the scale. Ectomorphy appears to be the easiest component to rate and mesomorphy the hardest." Among the three raters absolute mean differences ranged from 0.00 to 0.28 and reliability coefficients from 0.82 to 0.93. For the three raters the percentage differences greater than ½-unit for the first component ranged from 5–17%, for the second from 12–17%, and for the third from 3–13%.

Hunt and Barton ('59) report correlations on the first and second components when five raters compared Hooton's and Sheldon's rating methods (N = 28 to 30). For the first component the correlations ranged from + 0.66 to + 0.89; for the second they ranged from + 0.66 to + 0.82.

Dupertuis and Emanuel ('56) cite correlations of + 0.82 to + 0.86 between expert raters who compared Hooton's and Sheldon's methods.

Damon et al. ('55, '60) report "virtually identical" ratings for comparisons of raters. Sheldon and Damon made these ratings on two sets (124 and 146 subjects respectively) of photographs of "normal whites." Their mean differences by components range from 0.02 to 0.08.

Roberts and Bainbridge ('63) establish somewhat less demanding criteria in their study. They state, "A satisfactory standard of consistency was considered to have been attained when no two ratings of the same subject differed by more than one point per component" (p. 341). On an additional selection of photographs from the same series their average difference was 0.15 points per component per individual.

Berry and Deshfukh ('64) checked their reliability against Tanner's ratings on 28 of their subjects. Tanner and Berry "agreed to within ± ½-unit in 81% of the cases. In 15% ratings differed by one point, and in 3.5% by one and one-half points." Absolute mean differences ranged from 0.03 to 0.19.

Livson and McNeill ('62) report a recent revision of somatotyping by Sheldon, somewhat different from his previous system. The authors state that the differences between "old and present values" for the three components are as follows (a) for the first component, r = + 0.70, and the mean is 1.0 higher; (b) for the second component, r = + 0.73, and the mean remains the same as before; (c) for the third component, r = + 0.79, and the mean is 0.3 higher.

When the same observers repeat ratings as later intervals they usually achieve better agreement than in comparisons of ratings with different observers. When Hunt ('59) repeated the first component ratings of 30 photographs of army personnel after a two-year interval, he reports a correlation of + 0.958. Damon ('62) re-rated the photographs of 199 white and 65 Negro soldiers after an interval of six months. His percentage agreement ± ½-unit for the white soldiers was 92.4%; his correlations for the components ranged from +0.83 (second component for whites) to + 0.93 (third component for both whites and Negroes). For the whites only, his mean differences were – 0.01 for all three components. Damon ('60) also re-rated 174 females, ages 22–75 years, after a 6–8 week interval. His percentage agreement ± ½-unit averaged 93.7%; his absolute mean differences for the components ranged from 0.05 to 0.14; and his correlations ranged from + 0.79 for the second, to + 0.90 for the first and third components.

The present study indicates that when Heath and Carter use the same methods.

the overall agreement on somatotype component ratings is as good as or better than previous comparisons between trained observers. Heath's and Carter's differences are somewhat smaller, their overall correlations are similar, and their percentage agreements to one-half unit (96%) are higher. In addition, they compare favorably with intraobserver results. They did not encounter the difficulty in rating the second component, which was reported by Tanner ('54), by Parnell ('58), and by Hunt and Barton ('59).

Influence of M.4 data on ratings. The percentage changes in ratings were calculated for individual components in order to determine the extent of Heath's changes in her original ratings when she used the M.4 data, and the extent of Carter's changes from his M.4 ratings when he applied Heath's method. For the male series, Carter's changes from M.4 ratings to Sheldonian ratings were also calculated. These data are presented in table 4.

The data in table 4 indicate that both raters made relatively numerous and similar changes in first component ratings (32.5–44.9%). They made fewer second and third component changes — except for Carter's changes (57.6%) in his male third component ratings. These changes appear to be due to the difference in age distributions of the male and female portions of the series. Almost 90% of the females were age 18 ± 1 year, while approximately one-half of the males were age 18, and the other half were age 23. Figure 1 shows that there are different height/weight ratio criteria for ages 18 and 23. Comparison of M.4 height/weight ratio criteria for the third component with the criteria implied in table 1 illustrated the influence of age-scaled criteria on somatotype component ratings.

For the first component, Carter's changes from M.4 to Sheldonian ratings were similar to the other changes; they were low for the second (10.2%) and third (13.6%) components. The similarities in criteria for height/weight ratios and for different ages in the M.4 chart and according to Sheldon, probably account for the relatively small number of changes.

Modification and adaptation of M.4 data. Let us examine briefly the construction of the M.4 chart. Parnell ('58) discusses its development and "advantages" at some length in his first chapter, *Technique*. For anyone seriously interested in somatotyping this is an important discussion. The M.4 chart was designed to conform to Sheldonian somatotype ratings based upon a seven-point scale for each component. Figure 1 shows that the M.4 chart provides for recording of seven anthropometric measurements in addition to height and weight. Individual ratings of the first component are interpreted from totals of three skinfold measurements (subscapular, suprailiac, triceps). The skinfold totals are presented in a logarithmic Fat scale, which Parnell ('58) constructed to "correct for the skewness in the male distribution of the sum of the three skinfold measurements. Their mean "Total Fat" measurement is given a rating of three and one-half to correspond roughly with Sheldon's mean of 3.2 in endomorphy among students ('40) and of average man, 3.34 ('54)" (p. 20). These totals are tabulated to account for age changes. (Parnell's M.4 chart is reproduced in figure 1, as mentioned above.)

Second component ratings are derived from interpretations of measurements of bone diameters (humerus and femur), from measurements of muscle circumferences (biceps and calf), and from "corrections for fat." The criteria for second component ratings are tabulated on the assumption that for a given rating, bone diameters and muscle circumferences are proportionate to height. The table lists measurements purportedly appropriate to a second component rating of four, for statures at intervals of one and one-half inches.

The third component is rated directly from the height/weight ratio, so that it corresponds closely to the median of the range for each component as suggested in Sheldon's ('54) tables. Height/weight ratios are distributed for half-unit ratings from one through seven, corrected for age, with the ages from 18 to 43 at five-year intervals.

Several investigators have reported significant relationships among skinfold measurements, soft tissue radiographs, total body fat and the first component. In fact Parnell adopted the total of the three skinfold measurements for obtaining provisional estimates of the first component because his finding seemed to reflect well the total body subcutaneous fat. He used Edwards' ('50) data, which showed that "the coefficient of correlation between the sum of measurements at 53 sites and the sum at the three sites used in this survey showed very close agreement — + 0.99." (Parnell, '54, p. 218).

Hunt and Barton ('59), discussing the relationships between skinfolds, radiographs and first component ratings, report that "observers generally do not agree as well in assessing the first component as do workers who measure the subcutaneous fat layer from skinfolds or radiographs" (p. 34). They emphasize that it is probably valuable to use skinfold measurements to represent the first component.

In an Army series of Negro and white soldiers, Damon et al. ('62) reported r's of + 0.72 to + 0.75 between skinfolds and first component ratings.

When Heath examined the M.4 chart data for the subjects in this study, she realized at once that her interpretations of the data could not be in accord with Parnell's criteria. Likewise when Carter applied Heath's method in re-rating the series, he too departed from Parnell's criteria. Actually, the criteria of the M.4 chart are automatically modified if Heath's method is applied. However, the authors recognize the potential value of the raw anthropometric data. But they found they were compelled to interpret these raw data empirically, because according to Heath's method the rating scale is open at both

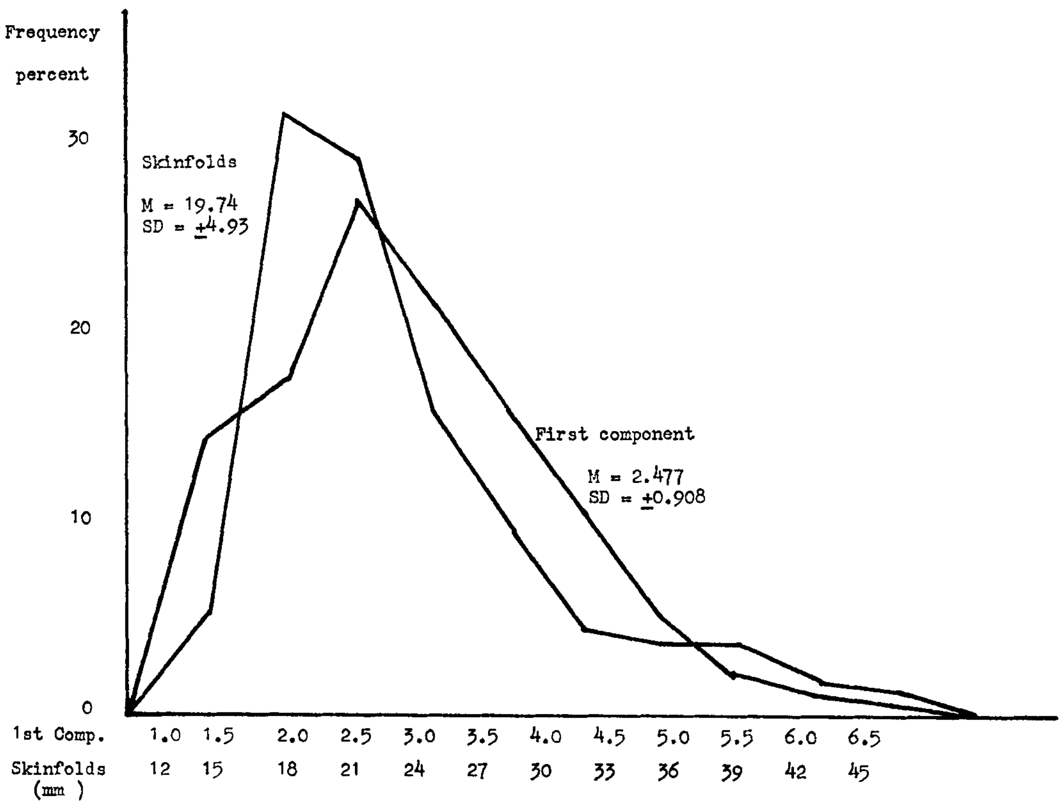

Fig. 2 Frequency distribution for first component ratings and total skinfold measurements for Olympic Games and British Empire Games athletes (N = 162).

ends, the height/weight ratio criteria are the same for both sexes at all ages, and a linear relationship is maintained between height/weight ratios and somatotype component ratings.

Heath found that the skinfold measurements in particular influence fine discrimination (± ½-unit) of relative expression of the first and second components. Table 3 shows that she made + one-unit changes in 1.7% of male first component ratings, and minus one-unit changes in 1.6% of female second component ratings. However, as mentioned above, both raters made fairly numerous one-half unit changes. Of course changes in third component ratings are inversely related to changes in the first and second components, because of the linear relationship between height/weight ratios and somatotype ratings in Heath's method.

In addition to the 120 subjects in this study, Heath studied the skinfold measurements and the somatotypes of 162 Olympic Games and British Empire Games athletes from two of Tanner's ('64) series. Heath's ratings of these series indicate that they are uniquely low in the first and third components and high in the second, compared with large American series she has rated. Therefore, it is not surprising that the range of skinfold measurements is small and heavily concentrated at the low end of the distribution. Figures 2 and 3 show the frequency distributions of first component ratings and skinfold measurement totals (in millimeters) for the Olympic and British Empire athletes and for the males in the New Zealand series. With a mean of 2.48 in the first component, the athletes are about one S.D. lower than Sheldon's ('54) mean of 3.34 for males in general. The mean of 3.40 for the New Zealand males is about the same as Sheldon's. Comparison of distributions of the first component with distributions of skinfold measurements, as shown in figures 2 and 3, suggest that the relationship between the first component and skinfold measurements may be similar to the linear relationship between height/weight ratios and somatotype component ratings. Relation-

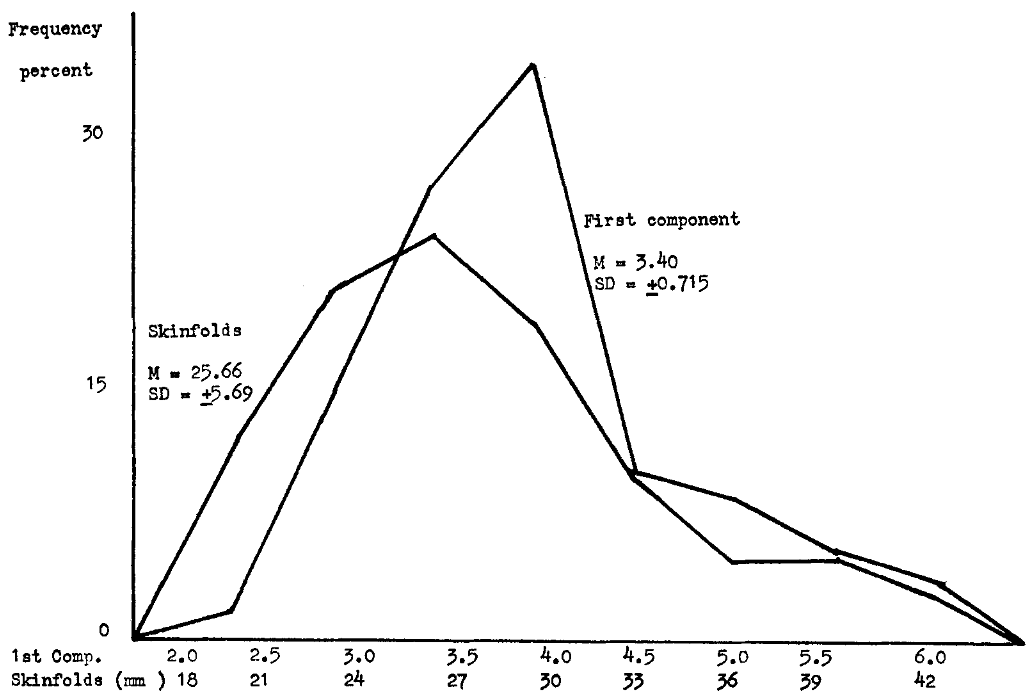

Fig. 3 Frequency distribution for first component ratings and total skinfold measurements for New Zealand men (N = 59).

ships among bone diameters, muscle girths, stature and somatotype ratings have not been studied as yet, but it will be important to investigate these relationships in the context of Heath's method.

This study gives good indirect evidence that Heath's method is easily reproduced. Carter had considerable experience with Parnell's method and was somewhat experienced in the Sheldonian method. With very little training in Heath's method, his ratings were remarkably similar to Heath's. The evidence is indirect because both raters considered the M.4 data when they re-rated the series. However, as pointed out above, application of Heath's method automatically eliminates various aspects of Parnell's M.4 chart interpretations of anthropometric measurements. In effect, therefore, both raters applied empirical interpretations to raw data. Thus, when Heath and Carter applied the same criteria, their overall agreement on somatotype ratings is as good as or better than previous comparisons between trained observers.

Parnell ('58) recognized the potential importance of somatotyping as "an instrument . . . that would loosen the rigid joints of traditional anthropometry" (p. 2), and his modification of Sheldon's system is an important contribution. He also feels that "the photograph cannot be discarded," but adds . . . physical anthropometric estimates undoubtedly sharpen the definition of typing, both by adding greater precision and by improving agreement between observers. In practice, one comes to rely greatly on this objective check to photoscopic impressions" (p. 24). In other words, the more usable information there is available to an observer, the more likely that his ratings will be reliable.

The somatotype photograph is important in a number of aspects of somatotyping which are not central to this study. For example, dysplasias are readily recognized in the photograph. Variations in fat deposition (sites), inner and subcutaneous fat, consistency (and probably composition) of fat, are not always well represented either by skinfolds or photoscopic impressions. For example, Damon et al. ('62) observed that the connective and elastic tissues bind the skin and subcutaneous fat to the underlying fascia more tightly among Negroes than among white subjects. Thus some total skinfold measurements may be misleadingly high or low. However, experienced photoscopy helps in detection of such idiosyncrasies and influences the rater in assessment of the first component.

Day-to-day and same-day fluctuations of weight in certain subjects under certain conditions may give misleading height/weight ratios. It is known that athletes can have weight changes of 4 to 10 pounds within periods of 1 to 3 hours, due to water loss alone. Weight changes of this order alter height/weight ratios (of, for example, a subject whose stature is 70 inches) from 0.10 to 0.30. However, it is doubtful that temporary weight changes of this kind are correspondingly reflected in photoscopic impressions of the first and third components.

Sometimes under special circumstances it is not feasible to obtain somatotype photographs, although it is convenient to make anthropometric measurements. An observer skilled in somatotype rating can make reliable ratings from direct somatoscopic inspection of the subjects. Nevertheless the somatotype photograph is invaluable as part of the recorded data, and is useful and important for future detailed reference and study. It is particularly desirable for longitudinal comparisons.

The authors feel that the anthropometric data of the M.4 chart can be adapted and interpreted so as to sharpen the criteria of Heath's method, and will increase the objectivity and reliability of somatotype ratings. The authors plan further study of relationships between bone diameters, muscle circumferences and somatotype components. They plan a further study which will include subjects selected for extreme expression of somatotype components. They hope to publish a step-by-step description of somatotype procedure, simplified so that somatotyping will be a readily usable research instrument.

LITERATURE CITED

Allen, N. 1965 A factor analysis of selected college football ability test items. Unpub. Ph.D. Thesis, University of Oregon.

Behnke, A. R., O. E. Guttentag and C. Brodsky 1957 Quantification of body configuration in geometrical terms. Research and Development

Technical Report USNRDL-TR-204 NS 080-001, U. S. Naval Radiological Defense Laboratory, San Francisco.

Berry, J. N., and P. Y. Deshfukh 1964 Somatotypes of male college students in Nagpur, India. Human Biology, *36:* 157–176.

Carter, J. E. L. 1964 The physiques of male physical education teachers in training. Jour. of Phys. Ed. Assoc. of Gr. Britain and N. Ireland, *169:* 66–76.

——— 1965a The physiques of female physical education teachers in training. Jour. of Phys. Ed. Assoc. of Gr. Britain and N. Ireland, *170:* 6–16.

Carter, J. E. L., and M. L. Rendle 1965b The physiques of Royal New Zealand Air Force Men. Report to R.N.Z.A.F., Air Department, Wellington, New Zealand. (Also microcarded, University of Oregon, 1965.)

Carter, J. E. L., W. D. Ross, F. W. Kasch and W. H. Phillips 1965c Body types of middle-aged males in training. Jour. Assoc. Mental and Physical Rehab., *19:* 148–152.

Clarke, H. H., R. N. Irving and B. H. Heath 1961 Relation of maturity, structural and strength measures to the somatotypes of boys 9 through 15 years of age. Research Quarterly AAHPER, *32*:4, 449–460.

Cortes, J. B. 1961 Physique, need for achievement and delinquency. Unpub. Ph.D. thesis. Dept. of Social Relations, Harvard University.

Damon, A. 1960 Host factors in cancer of the breast and uterine cervix and corpus. Jour. nat'l Cancer Institute, *24:* 483–516.

Damon, A., E. P. Fowler, Jr. and W. H. Sheldon 1955 Constitutional factors in otosclerosis and Meniere's disease. Tr. Am. Acad. Ophthalm. and Otolaryng., *59:* 4, 444–458.

Damon, A., H. K. Bleibtreu, O. Elliott and E. Giles 1962 Predicting somatotype from body measurements. Am. Jour. Phys. Anthrop., *20:* 461–473.

Donnan, F. de S. 1959 Physique and hazardous occupations. British Medical Journal, *II:* 728–731.

Dupertuis, C. W., and I. Emanuel 1956 A statistical comparison of the body typing methods of Hooton and Sheldon. WADC Tech. Report 56–366, ASTIA Document AD 97205. Wright Air Dev. Center, USAF, Wright Patterson Air Force Base, Ohio. (As cited in Damon et al., '62.)

Edwards, D. A. W. 1950 Observations on the distribution of subcutaneous fat. Clinical Science, *9:* 259.

Heath, B. H. 1963 Need for modification of somatotype methodology. Am. Jour. Phys. Anthrop., *2:* 227–234.

Heath, B. H., C. E. Hopkins and C. D. Miller 1961 Physiques of Hawaii-born young men and women of Japanese ancestry. Am. Jour. Phys. Anthrop., *19:* 173–184.

Heath, B. H., S. Yanagisawa and S. Kondo Comparison of somatotypes of Japanese women university students in Tokyo and in Hawaii. In preparation.

Hunt, E. E. Jr., and W. H. Barton 1959 The inconstancy of physique in adolescent boys and other limitations of somatotyping. Am. Jour. Phys. Anthrop., *17:* 27–35.

Irving, R. N. 1959 Comparison of Maturity, Structural and Muscular Strength Measures for Five Somatotype Categories of Boys Nine through Fifteen Years of Age. Ed. D. Dissertation. School of Health, Phys. Ed. and Rec., University of Oregon.

Livson, N., and D. McNeill 1962 Physique and maturation rate in male adolescents. Child Development, *33:* 145–152.

Olsen, A. L. 1961 Characteristics of Fifteen Year Old Boys Classified as Outstanding Athletes, Scientists, Fine Artists, Leaders, Scholars, or as Poor Students or Delinquents. Ph.D. dissertation, School of Health, Phys. Ed. and Rec., University of Oregon.

Parnell, R. W. 1954 Somatotyping by physical anthropometry. Am. Jour. Phys. Anthrop., *12:* 209–239.

——— 1957 The Rees-Eysenck body index of individual somatotypes. Jour. Ment. Science, *103:* 209.

——— 1958 Behavior and Physique. An Introduction to Practical and Applied Somatometry. Edward Arnold Ltd., London.

Roberts, D. F., and D. R. Bainbridge 1963 Nilotic Physique. Am. Jour. Phys. Anthrop., *3:* 341–370.

Seltzer, C. C., and J. Mayer 1964 Body build and obesity — Who are the obese? Jour. Am. Med. Assoc., *189:* 677–684.

Sheldon, W. H., S. S. Stevens and W. B. Tucker 1940 The Varieties of Human Physique. Harper Bros. New York.

Sheldon, W. H., C. W. Dupertuis and E. McDermott 1954 Atlas of Men. Harper Bros. New York.

Tanner, J. M. 1954 Reliability of anthroposcopic somatotyping. Am. Jour. Phys. Anthrop., *12:* 257–265.

——— 1964 The Physique of the Olympic Athlete. George Allen and Unwin Ltd., London.

Walker, R. N. 1962 Body Build and Behavior in Young Childen: I. Body Build and Nursery School Teachers' Ratings. Monographs of Soc. for Research in Child Dev., Serial no. 84, Vol. 27, no. 3.