Some psychometric properties of the Queensland Test of Cognitive Abilities with New Zealand European and Maori children[1]

Ross St. George

Department of Education, Massey University

The Queensland Test of Cognitive Abilities was administered to a sample of 397 New Zealand European and 303 Maori school children aged between eight years six months and fourteen years five months. Mean score performances were compared. More importantly, internal test properties were studied for each ethnic group to address the questions of psychometric and conceptual equivalence.

A small statistically significant difference was evident between the two ethnic groups in terms of mean score performance. A regression analysis suggests however that this difference is of limited practical importance and this is supported by its real magnitude and the ethnic group score distribution overlap. Test reliability, validity and item difficulty were found to be very smilar for both ethnic groups supporting a conclusion of psychological equivalence, although reliability coefficients were only modest in nature and some items extremely easy.

It is argued that in general the comparative data concurs with trends evident in more recent studies of both quantitative and qualitative aspects of European and Maori cognitive abilities and that the study of internal psychometric aspects of test characteristics should routinely be undertaken on a more extensive basis with psychological and educational measures used with ethnically and culturally different groups.

The comparative performances of European and Maori subjects on measures of cognitive ability has been a topic of considerable research and much debate. Thomson (1859) probably undertook the first scientific study when he compared the quantity of millet seed Maori skulls could hold with that of Englishmen and with the aid of tapes and compasses concluded that Maori ". . . heads are smaller than the heads of Englishmen . . . (and thus) . . . inferior to the English in mental capacity" (p. 81). Since that time some twenty four studies spanning forty years have appeared employing quite a wide range of cognitive ability measures with samples from these two populations.

In a review of these studies St. George (1977) has noted how the pattern of research on Maori and European performances on cognitive ability measures has been a microcosm of events in this field of Anglo-American psychology. New tests have been successively employed in comparative studies and the 'explanations' advanced for performance differences have ranged across linguistic factors, social and educational opportunity reasons, cultural deficit and cultural difference reasons, to even a hint of the perennial race-intelligence debate.

Four features evident from the review of empirical investigations of Maori and European performances on measures of cognitive ability which have contributed to this study are summarized below.

1. The weight of empirical evidence did not consistently favour the proposition that Maori subjects performed less well than Europeans on all cognitive measures. This

view has arisen from a partial reading of the empirical evidence, some of which rests on small and methodologically suspect studies.

2. Performance differences between these two ethnic groups were not necessarily reduced by the use of non-verbal scales. It had been assumed that English language content increased test difficulty for Maori subjects, however, the removal of overt linguistic content does not necessarily assure scale comparability, a point repeatedly emphasized in the literature (see Anastasi, 1976).

3. When viewed chronologically, comparative studies have in more recent years evidenced fewer statistically significant differences in the performances of Maori and European subjects on both quantitive and qualitative measures of cognitive ability.

4. Although many comparative studies of Maori and European performances on ability measures have been undertaken, psychometric data addressing the issue of actual test equivalence or comparability has been extremely meagre.

Almost without exception the principal concern of the many studies in this field has been the magnitude and direction of any Maori-European difference in mean score performance. With the finding of a mean score performance difference, statistical or otherwise, *post facto* supposition as to its meaning has been considerable. This style of research was, as Price-Williams (1975) has noted, representative of a cross-cultural psychology focused upon the cataloguing of similarities and differences in aspects of human behaviour. It is Price-William's contention that a more relevant cross-cultural psychology is one that seeks to identify and understand the impact of cultural variables on behaviour. The methodological development that has accompanied this conceptual shift in the approach to studying aspects of behaviour cross-culturally has been the emphasis now given to researching the equivalence of comparability or behavioural science scales. The theoretical rationale and methodological approaches for the researching of psychological test equivalence across culture group samples can be found in Brislin, Lonner and Thorndike (1973) Cleary, Hum-

phreys, Kendrick and Wesman (1975) and Drenth and van der Flier (1976).

This paper reports the summary statistics from the administration of the Queensland Test of Cognitive Abilities (QT) (McElwain and Kearney, 1970) to samples of New Zealand European and Maori children. It looks first at comparative performances but gives particular attention to the within-group psychometric properties of the QT for these samples in order to assess scale comparability.

Method

Sample. A sample of 700 subjects comprising 397 European and 303 Maori school children, aged between 8 years 6 months and 14 years 5 months in six 12 month age groups with approximately equal numbers of males and females were tested with the QT. Subjects were from primary and secondary schools in both urban and rural areas and from a range of socioeconomic status backgrounds as assessed on the Elley and Irving (1972) scale.

Ethnic group identification (Maori and European) was solely on the basis of the subject's personal choice. This method of subject ethnic group identification is supported by McDonald's (1976) analysis of the categories 'Maori' and 'Pakeha' (European) and Vaughan (1972) has shown that Maori and European subjects of this age range make clear and consistent ethnic group affiliation choices.

Full details on subject characteristics and sampling methods are reported in St. George (1977).

Instruments. The QT is an individually administered performance measure of general cognitive ability. It consists of 60 items grouped into five subtests;—Knox Cube Imitation, Beads, Passalong, Form Assembly and Pattern Matching. These are all test types of a traditional nature in the history of performance tests but each has been substantially developed for use in culturally diverse contexts on the basis of studies in the Southwest Pacific and Australia. The development and full description of the QT, related scales, the results of research and commentary on application are covered in Kearney (1966), McElwain and Kearney (1970), Ord (1970) and St. George (1977).

McElwain and Kearney, the test authors, refer to the QT as 'culture-reduced' rather than 'culture-free' or culture-fair'. While there is no overt linguistic content and considerable steps were taken to remove covert cultural content (e.g., use of number skills in determining

and remembering item components, the elimination of fine manual dexterity skills in manipulating materials and the use of mime to administer subtests), the intention was still to assess a general cognitive ability construct as elaborated in Vernon's (1969) hierarchical model of intelligence and the Australian QT construct validation studies are interpreted as supporting this construct.

The QT was designed for use with subjects from approximately 7 years of age and older. It is essentially an unspeeded scale and is to be used as a total test. Inferences on the basis of subtest performances and differences are cautioned against. Administration time is approximately 45 to 60 minutes per subject.

Data from the following tests were also collected in the course of QT validation studies:— (a) Pacific Reasoning Series Test (Ord, 1968), (b) Otis Intermediate Test (N.Z.C.E.R., 1969) and (c) Progressive Achievement Tests of Reading Vocabulary and Reading Comprehension (Elley and Reid, 1969). Class performance data from teacher-based assessments of Language Skills, Mathematics, Nature and Social Studies plus a teacher ranking of class groups on the criterion of 'General Ability' were also obtained.

Procedures. All children were administered the QT individually during normal school hours by one of three trained testers (one male, two female). The Pacific Reasoning Series Test (P.R.S.T.) and Otis Intermediate Test (Otis Int) were administered by the same male tester to all four class groups participating in the reliability and validity studies. These studies involved 127 children, 80 European and 47 Maori, in four class groups at the class levels of Standard 3, Standard 4, Form 1 and Form 2. The progressive Achievement Tests of Reading Vocabulary and Reading Comprehensive (Pat-RV and PAT-RC) were administered by the class teachers.

The three teacher-based assessments were derived from the unweighted summation of class test results in the respective subject areas as recorded in the Register of Progress and Achievement (Form E 19/16). The teacher ranking of 'General Ability' was obtained at the end of the first month of the school year, prior to the administration of the above tests.

All children in the six-month interval QT test-retest reliability study were tested by the same male tester.

Further procedural notes are reported in St. George (1977) along with QT subtest analyses. The focus in this paper is primarily upon QT total score performance and all data has been treated in raw score form.

Results

European and Maori QT performances

A 3-way ANOVA (gender x age x ethnic group) revealed no statistically significant gender main effect on QT scores, $F(1,676) = 1.83$, N.S., nor interaction effects involving gender. Consequently, the data were collapsed across gender. Table 1 reports sample sizes and summary statistics for the six age groups of European and Maori children.

The 3-way ANOVA referred to above showed however, statistically significant differences for age, $F(5,676) = 36.03$, $p<.01$, and ethnic group, $(F(2,676) = 20.71$, $p<.01$. Inspection of the data in Table 1 shows that scores in general evidenced the expected increase with age. The ethnic group difference favors European children at each age level although varying in magnitude. This latter result was further investigated through regression techniques.

First, the regression equations for QT scores with age were determined for the European and Maori samples. These are reported in Table 2.

Table 1: *QT mean scores and standard deviations for European and Maori samples by age group.*

Age Group	European			Maori		
	N	Mean	S.D.	N	Mean	S.D.
8:6- 9:5	36	34.83	7.16	37	31.76	7.38
9:6-10:5	57	37.93	6.22	33	34.39	5.13
10:6-11:5	125	40.48	5.38	69	39.32	5.52
11:6-12:5	63	40.90	5.41	59	39.03	6.02
12:6-13:5	78	44.62	5.28	60	43.48	4.77
13:6-14:5	38	43.79	5.83	45	41.69	6.17

Table 2: *Linear regression equations and correlation coefficients for QT scores on age for European and Maori samples.*

	European	Maori
y	0.16 x + 19.19	0.18 x + 13.79
r	0.425, p<.01	0.496, p<.01
$\bar{x}$	137.74	139.92
$\bar{y}$	40.80	38.98
N	397	303

Where y = linear regression equation, r = correlation between QT score and age, $\bar{x}$ = mean age, $\bar{y}$ = mean QT score and N = sample size.

The methods described by Ferguson (1970) for testing homogeneity of regression coefficients were then applied. The proposition that the slopes of the regression lines were the same was tested and could not be rejected, $F(1,696) = 0.14$, N.S. Thus, by this test the slopes of the regression lines may be deemed the same. Next, the proposition that the regression lines fitted a single common slope was tested and rejected, $F(1.696) = 12.39$, $p<.01$. The regression lines while the same were found to be separated by a small but constant difference, estimated by this procedure to be 2.15 QT score points. These results support the significant ethnic group main effect and non-significant age-ethnic group interaction effect in ANOVA.

Second, in order to assess the practical importance of the ethnic group difference and the extent of its contribution to QT performance, relative to a set of other variables, multiple regression methods were employed.

The multiple regression methods described by Kerlinger and Pedhazur (1973) enabled an analysis to be made of the combined and separate contributions of a set of five variables on performance on the criterion variable of QT score. In addition the relative importance of the variables as 'explanations' of variance in QT performance were able to be ascertained.

The variables of age, socioeconomic status, gender, urban/rural location and ethnic group were entered in the multiple regression analysis of QT scores. The matrix of zero-order correlation coefficients between QT score and the five variables are reported in Table 3

and the multiple regression analysis is reported in Table 4.

In evaluating the contribution and relative importance of each of the variables attention is directed to the columns of Table 4 titles 'Zero Order Variance Accounted For' and 'Variance Accounted For When nEtered Last'. The first reports the squared zero order correlation coefficient which indicates the proportion of variance accounted for in the criterion variable by each variable without consideration of their relationships with other variables. The second reports the variance accounted for by each variable after the joint effects of the other variables have been semipartialled out. Semipartial correlations are also known as part correlations and may be used, as done here, to estimate the unique contribution of each variable entered

Table 3: *Matrix of zero-order correlations between QT score and the five variables entered in the multiple regression analysis (N=673)*[1]

	1	2	3	4	5	6
1. QT Total Score						
2. Age	.45					
3. SES	-.15	-.03				
4. Gender	-.03	.02	.05			
5. Urban/Rural	.14	.35	.00	.03		
6. Ethnicity	.15	-.05	-.35	.02	-.08	

All values of .12 and above significant: $p<.01$, two-tailed test.

[1] All correlations are based on a sample of 673. Twenty-seven subjects were removed from the analysis because of missing socio-economic status data.

Table 4: *Multiple regression analysis—QT score.*
Criterion—Queensland Test Score
Multiple Correlation Coefficient (R) = .49
$R^2 = .24$
F Ratio = 41.36 ($p<.01$)

Variable	Beta	F to delete	Zero Order Variance Accounted For	Variance Accounted For When Entered Last	P
Age	.45	156.11	.20	.18	.01
SES	-.09	6.21	.02	.01	N.S.
Gender	-.02	0.54	.00	.00	N.S.
Urban/Rural	.00	0.00	.02	.00	N.S.
Ethnicity	.14	15.05	.02	.02	.01

into the analysis after controlling for the indirect effects of covariation amongst the other variables.

Inspection of Table 4 shows that the variables of socio-economic status, gender and urban/rural location failed to account for any unique variance in QT score at the specified significance level. Thus the small but statistically significant zero order correlations for QT score and socio-economic status and QT score and urban/rural location were misleading. When account was taken of covariance, the direct contribution of these variables to QT score was reduced to statistically non-significant proportions.

The variables of age and ethnicity were found however to make statistically significant direct contributions to variance in QT scores. Age was found to directly account for 18 percent of the variance and ethnic group membership 2 percent. As a result 80 percent of the variance can be attributed to differences in subject ability on the QT if variable specification error has not occurred. This analysis confirmed that the variables age and ethnic group both had a significant effect on QT performance but that while age of subject is of considerable importance the influence of ethnic group membership is comparatively minor.

In real terms the limited practical importance of the ethnic group mean score difference is also pointed to by the degree of score distribution overlap indicated by the standard deviation range of 4.77 to 7.38 points with an average of 5.88 points when the mean score difference range is 1.14 to 3.54 points and averages 2.15 points. Equally telling is the standard error of measurement range of 4.37 to 6.79 points (average 5.22). Thus the average ethnic mean score difference is a half to one-third of one standard error of measurement.

QT Reliability

The reliability of the QT with European and Maori children was investigated by both internal consistency and test-retest methods. Standard error of measurement estimates were also derived.

1. Internal consistency estimates of QT reliability. Internal consistency estimates were obtained by the Kuder-Richardson Formula 20 (K-R 20) procedure. K-R 20 estimates were obtained for the European, Maori and combined samples of each age group level. These are reported in Table 5.

The coefficients indicate a satisfactory level of internal consistency at the six age levels for the two ethnic group samples and support the associated K-R 20 proposition that the items are sufficiently well selected from the universe of such items as to be representative of them.

The coefficient range of .70 to .86 is similar to that of .73 to .92 reported by McElwain and Kearney (1970) with Australian European and Aboriginal children aged 7 to 14 years.

2. Test-retest estimates of QT reliability. Test-retest reliability (r_{tt}) estimates were obtained from the four class groups samples with a test-retest interval of six months. The pooled class group estimates for European, Maori and all children are reported in Table 6.

The test-retest coefficients, while statistically significant, do not compare favorably with the average stability coefficient of .875 reported by McElwain and Kearney (1970) for Australian Aboriginal subjects over a one year interval. While the New Zealand

Table 5: *K-R 20 reliability coefficients for QT by age group and separate and combined ehtnic group samples.*

Age Group	European		Maori		Combined Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6- 9:5	36	.86	37	.85	73	.86
9:6-10:5	57	.82	33	.72	90	.80
10:6-11:5	125	.76	69	.76	194	.76
11:6-12:5	63	.77	59	.77	122	.78
12:6-13:5	78	.78	60	.70	138	.75
13:6-14:5	38	.82	45	.82	83	.82

QT stability estimates are of the same order for the European and Maori samples, they are also below that considered desirable for individual decision making. Mehrens and Lehmann (1975, p. 108) suggest a value of at least .85.

3. QT Standard Error of Measurement estimates. The standard error of measurement (SEm) is a very practical index of test reliability with low SEm's indicating that any single score is a close approximation to the theoretical true score. The SEm's for the QT scores at each age level and for European, Maori and combined samples are reported in Table 7.

The SEm estimates for the QT are of a comparable order for the European and Maori samples at each age group level although evidence a relatively wide band width.

QT Validity

QT validity for Maori and European children was investigated by consideration of content issues and the empirical investigation of criterion-related and construct validity.

1. Content validity. St. George (1977), following Cronbach's (1970) remarks that content validity is essentially concerned with the way a measure represents a universe of observations of interest and that the test author and test user must therefore define the universe of observations a test intends to sample, reviewed extensively QT content in relation to its intended use as an ability measure with European and Maori children. In content terms it was judged suitable for the proposed purpose.

2. QT criterion-related validity. Some QT criterion-related validity data for European and Maori children were obtained from studies with four class groups. The criterion measures were the PRST, Otis Int, PAT-RV, PAT-RC, teacher assessments of language skills, mathematics, nature and social studies and a teacher ranking of 'general ability'.

Table 8 reports the QT criterion-related validity coefficients for the European, Maori and combined class group samples.

Inspection of Table 8 shows no distinctive and systematic ethnic group difference in the pattern of coefficients. The correlations between the QT and PRST, Otis Int., Mathematics and teacher ranking of 'General Ability' for the combined samples are of a moderate positive nature. The correlations with criteria more dependent upon English language skills are not as strong. However, while not statistically significant they are still largely of a moderate positive nature.

In St. George (1977) the full correlation matrix tables are reported and these did show however that for both European and Maori samples the Otis Intermediate Test correlated extremely well with the class based assessments. The coefficients exceeded those for the QT and PRST.

3. QT construct validity. Evidence bearing upon construct validity, that is the extent to which QT test results can be accounted for by explanatory constructs in psychological theory, was sought in three ways.

First, it can be argued that the types of items employed in the QT have been found

Table 6: *QT scores test-retest correlation co-efficients for European, Maori and combined samples from four class groups (interval 6 months).*

	European	Maori	Combined Sample
N	80	47	127
r_{tt}	.64	.66	.66
p	<.01	<.01	<.01

Table 7: *SEm estimates for QT scores by age group for separate and combined ethnic group samples.*

Age Group	European		Maori		Combined Sample	
	N	SEm	N	SEm	N	SEm
8:6- 9:5	36	6.52	37	6.73	73	6.79
9:6-10.5	57	5.58	33	4.25	90	5.40
10:6-11:5	125	4.66	69	4.76	194	4.72
11:6-12:5	63	4.73	59	4.98	122	4.93
12:6-13:5	78	4.63	60	3.95	138	4.37
13:6-14:5	38	5.19	45	5.51	83	5.46

Table 8: *QT criterion-related validity coefficients by class group for European, Maori and Combined samples.*

	Class Group[1]	N	P.R.S.T.	Otis Int	PAT RV	PAT RC	Lang Sk	Maths	Nat-Soc Stud.	T.R. Gen.Ab.
European	1	17	.49	.66	.51	.49	—	85*	—	.69*
	2	27	.49*	.38	.36	.18	—	.36	—	.38
	3	18	.72*	.46	.38	.44	.48	.54	.48	.49
	4	18	.75*	.30	.49	.17	.28	.56	.25	.37
Maori	1	10	.77*	.44	.20	.63	—	.60	—	.40
	2	8	.74	.47	.53	.21	—	.71	—	.44
	3	16	.68*	.58	.41	.49	.66*	.46	.30	.50
	4	13	.50	.37	.06	.36	.01	.42	.22	.57
Combined Sample	1	27	.60*	.54*	.42	.48	—	.72*	—	.64*
	2	35	.50*	.37	.36	.17	—	.40	—	.37
	3	34	.69*	.53*	.37	.45	.56*	.48*	.32	.53*
	4	31	.65*	.41	.39	.31	.20	.55*	.33	.53*

[1] Class Group: 1 = Std 3, 2 = Std 4, 3 = Form 1, 4 = Form 2.
* = $p < .01$; Lang Sk = Language Skills; Maths = Mathematics; Nat-Soc Stud = Nature and Social Studies; T.R. Gen Ab = Teacher ranking of 'General Ability'.

to be good measures of the hypothetical construct of general ability that has underpinned QT development.

Second, and in accordance with this general ability construct, QT scores should correlate positively with age. The linear regression equations reported in Table 2 indicate that for both European and Maori samples the QT performance and age regression coefficients were found to be similar indicating the same underlying maturational and experientially linked increase in QT performance for both ethnic group samples.

A third source of evidence on QT construct validity comes from factor analytic studies. Principle factor analyses were undertaken upon age and QT subtest data for both ethnic groups. Three factor solutions were sought on the basis of *a priori* postulation of a general factor, a spatial-perceptual factor and a memory factor underlying QT performance. The principal factor matrices for the European and Maori samples are reported in Table 9 and 10. Summary statistics, the zero-order correlation matrices and Varimax rotated factor analyses are all reported in St. George (1977).

The principal factor solution indicates a significant first factor for both ethnic groups with loadings on all five QT subtests and age being above .30. For both groups the first factor is interpreted as being of a general

ability nature. The loading of age in this factor conforms with maturational and experiential processes associated with such a general ability construct.

Table 9: *Principal factor matrix for age and QT subtests—European sample (N=397).*

Variables	Factors 1	2	3	Communality (h²)
Age	−.519	261	.032	.338
Knox	−.574	−.469	.009	.549
Beads	−.640	.091	.321	.521
Passalong	−.596	.028	.144	.377
Form Assembly	−.478	.078	.132	.252
Pattern Matching	−.655	.033	.070	.436
Total % of variance	81.8	12.3	5.9	
Eigenvalue	2.02	0.30	0.15	

Table 10: *Principal factor matrix for age and QT subtests—Maori sample (N=303).*

Variables	Factors 1	2	3	Communality (h²)
Age	−.597	.216	−.139	.422
Knox	−.373	−.003	−.004	.139
Beads	−.601	.365	.119	.508
Passalong	−.468	−.042	−.084	.228
Form Assembly	−.508	−.171	.329	.395
Pattern Matching	−.844	−.285	−.137	.812
Total % of variance	81.6	11.7	6.7	
Eigenvalue	2.05	0.29	0.17	

The second and third factors were not significant in terms of the eigenvalue criterion of 1 and load somewhat differently on the subtests for the two ethnic groups. For the European sample the second factor is associated principally with the Knox subtest and for the Maori sample the Beads subtest. The European sample third factor loads on the Beads subtest and for the Maori sample on the Form Assembly subtest.

To investigate further QT factor structure for each ethnic group the samples were partitioned into two age groups, 8 years 6 months to 11 years 11 months and 12 years to 14 years 5 months. Principal factor analyses employing three factor solutions were performed on QT subtest data for two ethnic groups at each age level and are reported in Table 11 and 12.[2]

For both ethnic groups at the two age levels the first factor is significant (eigenvalue >1) with loadings above .30 on all subtests. Again the most parsimonious interpretation

Table 11A: *Principal factor matrix for QT subtests—European sample aged 8:6 to 11:11 (N= 262).*

Variables	Factors			Communality (h²)
	1	2	3	
Knox	.532	−.205	.103	.335
Beads	.548	−.016	−.207	.344
Passalong	.568	−.036	.006	.324
Form Assembly	.435	.240	.066	.251
Pattern Matching	.652	.052	.041	.430
Total % of variance	90.3	6.2	3.5	
Eigenvalue	1.52	0.10	0.60	

Table 11B: *Principal factor matrix for QT subtests— Maori sample aged 8:6 −11:11 (N=181).*

Variables	Factors			Communality (h²)
	1	2	3	
Knox	.383	−.091	.236	.211
Beads	.395	−.077	−.032	.163
Passalong	.459	.344	−.044	.332
Form Assembly	.530	−.261	−.151	.372
Pattern Matching	.804	−.019	.028	.647
Total % of variance	83.6	11.7	4.8	
Eigenvalue	1.44	0.20	0.08	

[2] Summary statistics and zero-order correlation matrices for these analyses are available from the author.

Table 12A: *Principal factor matrix for QT subtests—European sample aged 12:0 to 14:5 (N= 135).*

Variables	Factors			Communality (h²)
	1	2	3	
Knox	.587	−.065	.210	.392
Beads	.558	−.239	.018	.369
Passalong	.494	.284	.152	.348
Form Assembly	.484	.279	−.189	.348
Pattern Matching	.665	−.155	−.176	.498
Total % of variance	80.6	12.5	6.9	
Eigenvalue	1.58	0.24	0.13	

Table 12B: *Principal factor matrix for QT subtests—Maori sample aged 12:0 to 14.5 (N = 122).*

Variables	Factors			Communality (h²)
	1	2	3	
Knox	.391	.090	.320	.264
Beads	.541	.352	.090	.425
Passalong	.544	−.274	.080	.377
Form Assembly	.611	.172	−.294	.490
Pattern Matching	.755	−.241	−.050	.630
Total % of variance	77.1	13.5	9.4	
Eigenvalue	1.68	0.30	0.21	

is of a first factor of a general ability nature. Loadings on second and third factors, which do not reach the eigenvalue criterion of 1, exceed .30 in only two instances.

The principal factor analyses reported here support a general ability interpretation of European and Maori subject QT performances. In these analyses there is little clear evidence of distinct spatial-perceptual or memory factors. The variation in loadings in the second and third factors between the ethnic group samples might suggest cognitive organisation differences but the type, comparatively short nature of each subtest and restricted range of cognitive performance variables entered in these analyses cautions against such an interpretation. McElwain and Kearney (1970) did not see the QT as being a measure particularly suited to investigating cognitive organization differences on the basis of subtest performances.

Collateral support for a general ability interpretation of QT performances can be found in the similar analyses reported in McElwain and Kearney (1970) with Australian European and Aboriginal subjects. In

addition, a principal factor analysis of a subset of the New Zealand QT data incorporating other group test performances from QT criterion-related validity studies, but disregarding the ethnic group categorization because of sample size restrictions, also obtained a significant first factor of a general ability nature with loadings on the five QT subtests and the four other cognitive measures (see St. George, 1977).

QT Item Analysis

Tables 11 to 15 report the item difficulty indices in terms of the proportion of subjects passing each item for the five QT subtest for the combined and separate ethnic group samples.

Table 15: *Proportion of the combined sample and European and Maori samples passing each item for the Passalong subtest.*

Sample	N	1	2	3	4	5	6	7
				Items				
Combined	700	99	99	88	65	61	51	18
European	397	99	99	90	71	68	53	18
Maori	303	99	99	85	58	52	48	17

Decimal points omitted

Table 13: *Proportion of the combined sample and European and Maori samples passing each item of the Knox subtest.*

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Items								
Combined	700	84	83	91	91	85	97	48	68	26	40	38	14	03	05	01
European	397	83	83	93	90	82	97	47	68	26	39	38	13	04	04	01
Maori	303	84	83	89	91	87	97	50	68	27	42	38	16	03	06	02

Decimal points omitted

Table 14: *Proportion of the combined sample and European and Maori samples passing each item of the Beads subtest.*

Sample	N	1	2	3	4	5	6	7	8	9	10
						Items					
Combined	700	81	89	43	64	34	41	35	32	43	16
European	397	81	89	46	65	37	43	35	35	45	17
Maori	303	81	89	39	63	29	39	34	28	39	14

Decimal points omitted

Table 16: *Proportion of the combined sample and European and Maori samples passing each item of the Form Assembly subtest.*

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13
							Items							
Combined	700	96	96	98	98	89	97	68	81	54	51	06	36	61
European	397	97	97	99	98	93	98	72	83	52	56	07	38	60
Maori	303	94	96	97	97	84	96	64	79	56	45	04	33	63

Decimal points omitted

Table 17: *Proportion of the combined sample and European and Maori samples passing each item of the Pattern Matching subtest.*

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Items								
Combined	700	99	99	98	99	99	99	99	91	91	92	94	79	72	53	71
European	397	99	99	99	100	99	99	100	94	91	95	95	85	80	57	76
Maori	303	99	97	97	98	98	98	99	87	90	89	92	71	61	46	64

Decimal points omitted

Inspection of the item difficulty tables shows that the proportion of European and Maori subjects passing each item in the five QT subtests to be very similar. The exceptions to this would be items 4 and 5 of the Passalong subtest, possibly item 10 of the Form Assembly subtest and the last four items on the Pattern Matching subtest. These items were found to be somewhat more difficult for the sample of Maori children—although the differences could not be judged as dramatic.

For both ethnic groups performance improves the first six items of the Knox subtest. While detracting from the subtest's psychometric properties, the role of these items in introducing the subject to the imitative requirements of the QT favors the retention of these very easy items.

The Beads subtest difficulty orders tend to show the second item as being easier than the first for both groups across the first six items. There appears to be some transfer of learning within these item pairs.

The short Passalong subtest shows a rapid but ordered rise in difficulty. Item 4 is a critical item in that in its solution lies the key to the subsequent items and the difference in the proportion of subjects passing Items 3 and 4 indicates this transition.

The first six items of the Form Assembly subtest are 'one-piece solution' items and were very easy for both the European and Maori children. Items 7 to 10 are all 'two-piece solution' items. Item 7 which introduces the task requirement is more difficulty than Item 8. Similarly, Item 11 which introduces the 'three-piece solution' items is very difficult. Item 13 is always administered which in part accounts for the increased proportion of correct responses.

The first eleven items of the Pattern Matching subtest were very easy for both ethnic groups. The proportions passing the remaining four items are also high indicating the need for some psychometric improvement to the subtest.

The speculations concerning item patterning and current placement are based on McElwain (1971) and McElwain and Kearney's (1970) comments. Clearly an empirical check would be interesting.

Discussion

This paper has summarized the major psychometric evidence from the application of the QT to large samples of European and Maori children. The attention given to these psychometric characteristics was because of their importance in evaluating the test's equivalence or comparability for use with children from either of these ethnic group backgrounds. Test equivalence is as important as the issue of ethnic group differences in mean score performances. Without some knowledge of test equivalence meaning cannot be attached to either score differences or similarities. In many cases however, it is mean score differences which has been given most emphasis in the literature and public discussion on test use with ethnically different groups.

With respect to the QT it was found that European children on the average obtained a slightly higher mean score than Maori children. While the difference was of statistical significance, it was found through the regression analysis that the direct and independent variance accounted for by ethnic group membership was small—two percent. Score distribution overlap for these two ethnic group samples and likely measurement error that must be taken into account in actual test use supports the conclusion that ethnic group membership is in itself of little practical importance in accounting for QT performance.[3]

The reliability studies showed the estimates of QT reliability to be comparable for

---

[3] A reviewer has helpfully pointed out that while ethnic group membership only makes a small (although statistically significant) contribution to QT performance, which in the context of the current paper has been viewed being of minor practical importance, there are situations where the small score difference could over time result in outcomes judged to be deleterious. In this regard Drosler's (1978) repeated measures proposals for improving upon classical test theory predictive validity estimates and which incorporates information on temporal decline in predictive validity and Guastello's (1982) application of catastrophe theory to the issue of differential validity in selection situations suggest the need for a more circumspect view. The author does regard the investigation of these issues as important and especially so with high useage selection and placement measures.

the two ethnic groups for K-R 20 test-retest and SEm methods. However, the test-retest data did not provide particularly strong evidence for stability in QT performance over the six month time interval. Possibly this was a restriction of range effect due to the ease of items in some subtests.

The validity studies lend some support to the proposition that the QT is a valid test for the assessment of general ability with both European and Maori subjects.

In terms of content validity it is considered that the test constructors achieved their objective of developing a general ability scale subject to minimal cultural influences and thus the QT is suitable for use with both European and Maori children.

The criterion-related validity data indicate that the QT was similarly correlated with a number of criterion measures for both European and Maori class group samples. However there are obvious sample and study limitations which means these data must be viewed cautiously. The scholastic nature of the criterion variables must also be borne in mind. More fruitful QT criterion-related validity research probably rests on the use of more applied skill-based criteria.

Data bearing upon QT construct validity suggested the utility of employing the same underlying general ability construct for both ethnic groups samples. This conclusion stems from consideration of QT content, from its relationship with subject age with both ethnic group samples and the similarity of the factor structures from the principal factor analyses.

The item difficulties for the QT subtest items were found to be virtually identical for both European and Maori children. While it would appear that the item difficulty orders for the Knox, Beads and Passalong subtests are satisfactory, the Form Assembly item order pattern is affected by the introduction of each new task requirement (pieces required for solution), and the Pattern Matching subtest probably has too many easy items thus reducing the discriminative capacity of this potentially sound adaptation of the well known Kohs block item type.

This body of psychometric data on the QT for both European and Maori children appears to be the first of its kind reported in relation to an individual cognitive ability scale for New Zealand. Indeed there is a paucity of data bearing on the technical and conceptual equivalence of the great majority of psychological and educational measures used with these populations. While some psychometric features of the QT proved to be less encouraging than might have been hoped for on the basis of the Australian studies of McElwain and Kearney (1970), the research reported provides an adequate data base for future New Zealand studies with the QT and a model for the psychometric assessment of instrument comparability across ethnic and culture group samples.

References

Anastasi, A. *Psychological testing*, 4th ed. New York: Macmillan, 1976.

Brislin, R. W., Lonner, W. J. and Thorndike, R. M. *Cross-cultural research methods*. New York: Wiley, 1973.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A. and Wesman, A. Educational uses of tests with disadvantaged students. *American Psychologist*, 1975, *30*, 15-41.

Cronbach, L. J. *Essentials of psychological testing*, 3rd ed. New York: Harper and Row, 1970.

Drenth, P. J. D. and van der Flier, H. Cultural differences and comparability of test scores. *International Review of Applied Psychology*, 1976, *25*, 137-144.

Drosler, J. Extending the temporal range of psychometric prediction by optimal linear filtering for mental test scores. *Psychometrika*, 1978, *43*, 533-549.

Elley, W. B. and Irving, J. C. A socio-economic index for New Zealand based on levels of education and income from the 1966 census. *New Zealand Journal of Educational Studies*, 1972, *7*, 153-167.

Elley, W. B. and Reid, N. A. *Progressive Achievement Tests, Teachers' manual—Reading Comprehension and Reading Vocabulary*. Wellington: N.Z.C.E.R., 1969.

Ferguson, G. A. *Statistical analysis in psychology and education*, 3rd ed. New York: McGraw-Hill, 1971.

Guastello, S. J. Moderator regression and the cusp catastrophe: Application of two-stage personnel selection, training, therapy and policy evaluation. *Behavioural Science*, 1982, *27*, 259-272.

Kearney, G. E. *Some aspects of the general cognitive ability of various groups of Australian Aboriginals as assessed by the Queensland Test*. Ph.D. thesis, University of Queensland, 1966.

Kerlinger, F. N. and Pedhazur, E. J. *Multiple regression in behavioural research*. New York: Holt, Rinehart and Winston, 1973.

McDonald, G. The categories Maori and Pakeha

as defined by research workers and by self report, *New Zealand Journal of Educational Studies*, 1976, *11*, 37-49.

McElwain, D. W. The Queensland Test. In Ritchie, J. E. and St. George, R. (Ed.), *Cross-cultural ability testing: Report of a research workshop.* Psychology Research Paper, Hamilton: University of Waikato, 1971, 5-16.

McElwain, D. W. and Kearney, G. E. *Queensland Test Handbook.* Hawthorne: A.C.E.R., 1970.

Mehrens, W. A. and Lehmann, I. J. *Measurement and evaluation in education* and psychology. 2nd ed. New York: Holt, Rinehart and Winston, 1975.

New Zealand Council for Educational Research. *Otis Tests of Mental Ability: Manual of directions.* Wellington: N.Z.C.E.R., 1969.

Ord, I. G. *Manual for the Pacific Reasoning Series Test.* Hawthorne: A.C.E.R., 1968.

Ord, I. G. *Mental tests for pre-literature.* London: Ginn, 1970.

Price-Williams, D. *Explorations in cross-cultural psychology.* San Francisco, Chandler and Sharp, 1975.

St. George, R. *The general cognitive ability of Maori and European subjects. A comparative study of Queensland Test performances and psychometric properties.* D. Phil. thesis, University of Waikato, 1977.

Thomson, A. S. *The Story of New Zealand. Past and present. Savage and civilised.* Vol. 1, London: Murray, 1859.

Vaughan, G. (ed). *Racial issues in New Zealand:* Akarana Press, 1972.

Vernon, P. E. *Intelligence and cultural environment.* London: Methuen, 1969.

Corrigendum

Kammann, R. & Campbell, K. Illusory correlation in popular beliefs about the causes of happiness. *New Zealand Psychologist*, 1982, *11(2)*, 52-63.

An error appeared in lines 3 and 5 of the abstract of this article. The abstract should have read:

Experiment I demonstrates that, contrary to objective data, most people believe that happiness is strongly associated with good health, number of friends, country or small town residence, no disability, income, intelligence and type of work. When presented with case study data in which health, friends and income were zero-correlated with happiness, subjects in Experiment II perceived positive correlations. At the same time the majority of subjects correctly detected true positive, zero, and negative correlations for other factors not usually associated with happiness. There was no evidence in recognition test data that confirming instances were better encoded than disconfirming instances in any of the relationships presented. A simple associative trace model accounts for most laboratory results. Popular beliefs about happiness could arise either from a halo effect among "good things of life" or an overgeneralization from vivid short-term to pallid long-term effects. The unobservability of inner mental states in others sets the stage for definitional confusion and illusory correlation.