Applicability of Australian Adaptations of Intelligence Tests in New Zealand with a Dunedin Sample of Children

Christina M. Rodriguez, Lee A. Treacy, and Paula J. Sowerby
University of Otago, Dunedin, New Zealand

Laura E. Murphy
University of Tennessee—Memphis
Boling Center for Developmental Disabilities

The current study evaluated the applicability of the Australian adaptations of the Wechsler Intelligence Scale for Children – Third Edition and the Stanford-Binet Intelligence Scale – Fourth Edition for a Dunedin sample of children. Previous research with earlier versions of the measures suggested that New Zealanders obtain scores at or above those of the U.S. normative sample. Ninety children participated, 60 between ages 5 to 8 receiving the SB-FE and 60 between ages 7 to 10 receiving the WISC-III (thus, 7- and 8-year-olds received both tests). Contrary to previous research, the current Dunedin sample obtained means comparable to the American norms. Limitations due to sample characteristics and sample size are discussed.

Historically, intelligence testing has been a polemic issue. Charges against intelligence testing are numerous and varied, including concerns regarding the biased, reductionistic, and limited predictive qualities of intelligence tests (Sattler, 1992). Doubts about the cultural equity of such tests have been repeatedly raised (e.g., see Sattler, 1992 for review). These concerns have been echoed by professionals in New Zealand (e.g., see Olssen, 1988), including concerns about cultural bias (Shuker, 1988) as well as the appropriateness of item content and norms for New Zealand children (Ballard, 1988).

The reported vices of intelligence tests have challenged the designers of standardized measures of scholastic ability and intelligence. Much of this condemnation has been met with a defense built upon improved revisions of intelligence tests and stronger psychometric evidence, although many of the issues remain far from resolved. Intelligence tests for children are intended to provide a current summary of relative strengths and weaknesses. Such measures are currently administered to New Zealand children for assessment of general cognitive and neuropsychological functioning, although the appropriateness of the instruments remains untested.

The leading measures for individual intellectual testing for children are designed overseas, with items and norms often reflecting the North American child's experience. Concerns raised about individually administered IQ tests in New Zealand include the unsuitability of the content of some test items for New Zealand children (e.g., St. George & Chapman, 1987; Tuck, Hanson, & Zimmerman, 1975). These concerns have been met by different researchers investigating IQ tests incorporating various item modifications (e.g., Chapman & St. George, 1984; St. George & Chapman, 1987; Tuck et al., 1975). Undoubtedly, clinicians administering these overseas scales have often implemented idiosyncratic item alterations, clearly violating standardized testing procedures. For instance, some have implemented a list of item modifications recommended by the "Chief Psychologist, Department of Education, which has been widely adopted..." (Silva, McGee, & Williams, 1981, p. 11.8), although these modifications were issued 5 years after the test was released and whether these recommended changes were, indeed, accessible to all examiners is unknown. Such arbitrary procedures only serve to amplify fears about the meaningfulness of IQ scores for New Zealand children, although such modifications are certainly understandable in the absence of any guidance from test authors on comparability across cultures.

Two frequently administered intelligence tests are the Wechsler Intelligence Scale for Children – Third

Edition (WISC-III; Wechsler, 1991) and the Stanford Binet Intelligence Scale – Fourth Edition (SB-FE; Thorndike, Hagan, & Sattler, 1986), both of which replace a line of earlier versions. Both IQ tests also have official Australian adaptations which presumably render them more suitable for New Zealand children. Surprisingly, neither of these adaptations has been systematically investigated despite the controversy and despite their continued application in this country. In order to address concerns about the comparability of normative data with New Zealand children, this issue must be directly studied.

To date, the only evidence pertaining to the validity of the WISC-III on New Zealand samples comes from an examination of the WISC-III's Third Factor (i.e., Freedom from Distractibility) and Processing Speed indices (Fernando, 1995). This study found both index scores approximating the American standardization means, and thus the author suggests the WISC-III may be suitable for New Zealand children aged 8 to 10 years. However, such interpretations were limited by the scope of the IQ subtests covered as well as the age range.

A supplement available on the use of the Australian adaptation of the SB-FE details results of the measure with three age groups of Australian children (de Lemos, 1989). For the 7- to 8-year-old age group, the author reports a Composite IQ score 4.5 points higher than the normative mean and a smaller standard deviation (de Lemos, 1989). In addition, two of the obtained mean SB-FE area scores were higher than the standardization sample (Verbal Reasoning and Abstract/Visual Reasoning, 109.5 and 105.8 respectively; de Lemos, 1989).

The predecessors of both the WISC-III and SB-FE have been studied with New Zealand samples, and collectively, these studies suggest that children in this country perform at or above the U.S. standardization sample. The large-scale Dunedin Multidisciplinary Study (Silva et al., 1981) reports on IQ scores for 5-year-olds on the Stanford-Binet Intelligence Scale: Form L-M (SB-LM; Terman & Merrill, 1973) and for 7-year-olds on the Wechsler Intelligence Scale for Children—Revised (WISC-R; Wechsler, 1974), including item alterations in the latter test. This study reports mean IQ scores 4 to 7 points above the U.S. norms (Silva, 1982; Silva et al., 1981). Based on these higher scores, Silva (1982) cautions examiners on the use of these measures in this country. However, the sample has been acknowledged as not representative of New Zealand children as a whole, with limited ethnic diversity and higher socioeconomic status (Share, McGee, & Silva, 1989). Moreover, the effect of the particular item modifications adopted in the study may have facilitated test performance.

Another study found considerably higher mean SB-LM and WISC-R IQ scores for a small sample of Wellington children, although neither the sample's representativeness nor possible item changes were addressed (Cumming & Marsh, 1985). Other researchers using the WISC-R (including item modifications) have found IQ scores to be slightly higher (from 2 to 5 points) than the normative means (Chapman & St. George, 1984; St. George & Chapman, 1987). Such higher scores suggest that the measures would be less likely to identify those with significant developmental disabilities in need of intervention. In contrast, a study of Christchurch 11- to 12-year-olds conducted shortly after the WISC-R's release incorporated several item alterations and found scores comparable to the normative means (Tuck et al., 1975).

Based on previous research with the WISC-R and SB-LM, the newest editions with Australian adaptations may be expected to lead to higher mean IQ scores than the normative U.S. means. Thus, the purpose of the current study was to examine the scores of a Dunedin sample of children on the Australian adaptations of the WISC-III and SB-FE. Such an investigation provides a preliminary indication of the applicability of these measures with New Zealand children. Primary school children were targeted for study, with one group receiving the SB-FE and another group receiving the WISC-III. A subsample received both tests at the age overlap between the measures in order to examine their association, which has been relatively understudied.

Method

Participants

Ninety children participated, 30 in each of three different age groups (5- and 6-year-olds, 7- and 8-year-olds, 9- and 10-year-olds). The sample was somewhat overrepresented by males (59%). With regard to ethnic self-identification, Pakeha/European New Zealander was the major ethnic group sampled (92.1%), with 3.4% of Maori descent, and 4.5% "Other" (including Pacific Islander, Asian, and unspecified).

Eighty percent of the children sampled lived with both biological parents. The remaining subjects either lived in single-parent homes (15.5%) or in reconstituted families with step-parents (4.5%). The number of siblings ranged from 0 to 5, with most children having either one or two siblings (74.5%). Distribution of the sample based on annual family income was as follows: $14,999 or less, 11%; $15,000-29,999, 27%; $30,000-

44,999, 28%; $45,000 or more, 34%. The highest educational level attained by the majority of mothers was within Form 1 to Form 7 (49%), followed by polytechnic, vocational training, or some undergraduate university (36%). Fewer mothers graduated university or teacher's college (12%) and still fewer pursued postgraduate study (3%). In general, maternal educational attainment compared favorably to paternal educational level. The highest academic level achieved by the majority of the fathers was within Form 1 to Form 7 (64%), followed by 26% at the level of polytechnic, vocational training, or some undergraduate university. Only 8% of fathers' highest educational attainment was a degree from university or teacher's college, with 2% of fathers obtaining a postgraduate degree.

The majority of children were recruited from two primary schools, selected on the basis of their respective urban (Dunedin) and suburban (Mosgiel) locations to provide approximately equal distributions across the two geographical locations. The remainder of the children (<10) were recruited via advertisement in a community newspaper. No attempt was made to exclude any children or modify test conditions in order to enhance the representativeness of the sample.

Measures

The *Stanford-Binet Intelligence Scale—Fourth Edition* (SB-FE; Thorndike, Hagan, & Sattler, 1986; Australian adaptation, 1989) is an individually administered set of 15 tasks assessing intellectual functioning in individuals aged 2 years through 23 years. Examinees do not respond to all 15 subtests but rather to those tasks suitable to their age and ability level. Thus, for the age groups involved in the current study, all children received the following eight subtests in four areas: Vocabulary, Comprehension, and Absurdities subtests providing the Verbal Reasoning Standard Age Score (SAS); Pattern Analysis and Copying subtests yielding the Abstract/Visual Reasoning SAS; Quantitative subtest for the Quantitative Reasoning SAS; and Bead Memory and Memory for Sentences subtests for a Short-Term Memory SAS. Older or more advanced children received 4 additional subtests: Matrices, which contributes to the Abstract/Visual Reasoning SAS; Number Series for the Quantitative Reasoning SAS; and Memory for Digits and Memory for Objects for the Short-Term Memory SAS. The SB-FE Composite IQ is based on performance across the four areas; thus, the Composite IQ is based on 8 subtests for some children and on 12 subtests for others. The area SAS and Composite IQ scores are based on a deviation IQ ($M = 100$, $SD = 16$) and individual subtests are based on scaled scores ($M = 50$, $SD = 8$).

Standardization of the SB-FE involved a sample of 5013 individuals in 17 age groups selected to be representative of the U.S. population, stratified based on the 1980 U.S. Census on such variables as geographic region, community size, ethnic group, age, gender, and socioeconomic status (Sattler, 1992). With regard to reliability, internal consistency is excellent, with a median coefficient across ages for the Composite IQ of .97 (Sattler, 1992). The manual (Thorndike et al., 1986) reports strong retest stability coefficients for an intertest interval from 2 to 8 months for two groups of children on the Composite IQ (.90 and .91), but lower coefficients for the area SAS (ranging from .51 to .88, with Quantitative Reasoning SAS lowest and Verbal Reasoning SAS highest for both groups of children). For concurrent validity, the manual (Thorndike et al., 1986) reports the SB-FE's Composite IQ correlation is .81 with its predecessor, the SB-LM, .83 with the WISC-R, .80 with the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989), and .91 with the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981).

The *Wechsler Intelligence Scale for Children—Third Edition* (WISC-III; Wechsler, 1991; Australian adaptation, 1992) is an individually administered battery of 13 subtests designed to measure the intellectual ability of children aged 6 years through 16 years, 11 months. Six subtests assess verbal skills on a Verbal Scale and 7 subtests assess perceptual-motor skills on a Performance Scale. Five subtests in each of the two scales are designated standard tests: Information, Similarities, Arithmetic, Vocabulary, and Comprehension subtests contribute to the Verbal IQ (VIQ); Picture Completion, Coding, Picture Arrangement, Block Design, and Object Assembly determine the Performance IQ (PIQ). The Verbal and Performance Scales combine to provide the Full Scale IQ (FSIQ). The three supplementary subtests (Digit Span on the Verbal Scale; Symbol Search and Mazes on the Performance Scale) are excluded from calculation of the IQ summary scores unless a standard subtest has been invalidated or omitted. However, Digit Span and Symbol Search are required for the calculation of four factor-based index scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, Processing Speed). The WISC-III uses the deviation IQ ($M = 100$, $SD = 15$) for the VIQ, PIQ, FSIQ and index scores and scaled scores ($M = 10$, $SD = 3$) for the individual subtests.

The WISC-III standardization sample included 2200 children, 200 at each of 11 age groups stratified on such variables as age, gender, ethnicity, geographic region, and parental education based on the 1988 U.S. Census (Wechsler, 1991). The WISC-III purports to have excellent reliability, reporting internal consistency coefficients of .95 for VIQ, .91 for the PIQ, and .96 for FSIQ in the standardization sample (Wechsler, 1991). The WISC-III also reportedly provides a stable measure of intelligence, with reliabilities for an intertest interval ranging from 12 to 63 days of .87 for the PIQ, and .94 for both the VIQ and FSIQ (Wechsler, 1991). The WISC-III also reports adequate concurrent and predictive validity across age groups for children with and without disabilities (Wechsler, 1991). The manual (Wechsler, 1991) indicates that the WISC-III VIQ, PIQ, and FSIQ scores correlate concurrently with the WISC-R (coefficients range from .81 to .90), with the WAIS-R (range .80 to .90), and with the WPPSI-R (range .73 to .85). Moderate to strong correlations are also reported with tests of school achievement and school grades (Sattler, 1992).

Procedure

Children in the 5 to 6 and 7 to 8 year age groups were administered the SB-FE and those in the 7 to 8 and 9 to 10 year age groups were administered the WISC-III. (Six-year-olds did not receive the WISC-III because of concerns regarding "floor" effects [see Sattler, 1992].) Thus, the 7- to 8-year-old group received both IQ tests, in a counterbalanced order, administered in two separate test sessions by the same examiner, with an intertest interval of 1 to 10 weeks (*Mdn* = 3 weeks). Each IQ test was individually administered within a single session, allowing for breaks. Testing was completed from April to October of 1996. The majority of children were tested in a quiet room at their respective schools. Those parents preferring their child be tested outside the school, as well as those children recruited via advertisement, were tested in a quiet room in the university's clinical psychology training clinic.

The Australian adaptations of the WISC-III and SB-FE were administered by one of three examiners with postgraduate training in child assessment techniques. As described above, 8 or 12 subtests of the SB-FE were presented depending on the child's age and ability level, allowing for computation of the area SAS and Composite IQ scores. For the WISC-III, the 10 standard subtests in addition to the supplementary Digit Span and Symbol Search subtests were administered, thus allowing for determination of all IQ scores and index scores. Standardized testing conditions and scoring were strictly observed, with no modifications made to any items on the Australian adaptations.

All 120 test protocols were rechecked for scoring errors or inconsistency. In addition, to assess interscorer reliability, 20% of the protocols (12 SB-FE and 12 WISC-III protocols) were randomly selected to be rescored "blind" by an experienced examiner. To avoid artificially inflating interscorer reliability by including subtests with objective scoring criteria (i.e., with minimal variability in scoring), only those subtests considered to potentially entail interscorer disagreement were independently rescored. All interscorer reliability coefficients were acceptable, with no significant differences between mean raw scores for any of the subtests (all $p > .01$). For the WISC-III, interscorer reliability coefficients were Information, $r = .98$; Similarities, $r = .98$; Vocabulary, $r = .92$; and Comprehension, $r = .93$. For the SB-FE, reliability coefficients were Vocabulary, $r = .87$; Comprehension, $r = .99$; Absurdities, $r = .88$; and Copying, $r = .80$.

Results

All analyses were conducted using the SPSS for Windows statistical package. In light of the number of comparisons and analyses performed, a significance level of alpha = .01 was adopted. The effects of sample demographics (with the exception of ethnicity, for which there was insufficient variance) are presented first for the summary scores (WISC-III FSIQ, VIQ, PIQ and index scores; SB-FE Composite IQ and area SAS), followed by an examination of the descriptive statistics for the IQ tests' summary scores, and ending with a review of intercorrelations.

Sample Demographics

Based on *t*-tests, no significant sex differences in mean WISC-III test performance across the seven

Table 1. Descriptive Statistics for the SB-FE

	N	Mean	SD
SB-FE Composite IQ	**60**	**101.33**	**12.15**
Verbal Reasoning SAS	**60**	**103.03**	**10.31**
Vocabulary	60	51.70	5.71
Comprehension	60	50.77	5.10
Absurdities	60	51.45	6.77
Abstract/Visual Reasoning SAS	**60**	**97.75**	**12.51**
Pattern Analysis	60	51.63	9.09
Copying	60	45.38	5.08
Matrices	20	52.65	5.89
Quantitative Reasoning SAS	**60**	**106.12**	**14.52**
Quantitative	60	52.40	7.77
Number Series	20	54.25	5.77
Short-Term Memory SAS	**60**	**97.48**	**14.57**
Bead Memory	60	48.35	8.67
Memory for Sentences	60	48.92	6.84
Memory for Digits	20	52.05	7.34
Memory for Objects	20	50.55	6.14

summary scores were found (all $p > .01$). Similarly, no significant sex differences emerged for mean SB-FE scores for area SAS or Composite IQ scores (all $p > .01$). However, some sex comparisons for the SB-FE summary scores approached significance. Specifically, girls attained marginally higher scores on the Short-Term Memory SAS, $t(58) = 2.61, p = .012$, Abstract/Visual Reasoning SAS, $t(58) = 1.95, p = .056$, and Composite IQ, $t(58) = 1.97, p = .054$.

The effect of family composition was examined with t-test comparisons between children in single- versus two-parent homes. For both the WISC-III and SB-FE summary scores, no significant mean differences were detected between groups (all $p > .01$). One-way analyses of variance were conducted to examine the effect of annual family income on test performance. Across all summary scores for both IQ tests, no significant differences emerged across the four income levels (all $p > .01$).

Differences in child's IQ test performance based on parental educational attainment were examined via t-tests, grouped into those with highest educational level at Form 7 or below and those with highest educational level above Form 7. For paternal educational level, no significant group differences were found in mean test performance for either the WISC-III or SB-FE summary scores (all $p > .01$). However, three comparisons for paternal educational level approached significance. In particular, children whose fathers had attained above Form 7 education obtained marginally higher scores on the SB-FE Verbal Reasoning SAS, $t(56) = 2.13$, $p = .038$. Similarly, children with fathers above Form 7 educational levels obtained marginally higher scores on the WISC-III VIQ, $t(57) = 2.03, p = .047$, and on the WISC-III Verbal Comprehension index, $t(57) = 2.12$, $p = .038$. For maternal educational attainment, no significant group differences were found across the WISC-III and SB-FE summary scores (all $p > .01$). However, two comparisons approached significance. Children whose mothers had over Form 7 education obtained marginally higher scores on the SB-FE Abstract/Visual Reasoning SAS, $t(58) = 2.25$, p = .028, and on the WISC-III Perceptual Organization index, $t(58) = 2.07$, p = .043.

Summary Score Descriptive Statistics

Based on the absence of statistically significant differences in test scores across demographic characteristics, all subsequent analyses are based on the full sample of participants. Table 1 presents the descriptive statistics for the SB-FE summary and subtest scores. Examination of these data reveals that the SB-FE Composite IQ is within the standard error of measurement (SEm) of the normative mean of 100. In addition, all of the obtained area SAS are within the SEm of the normative means of 100, with the exception of the Quantitative Reasoning SAS (M = 106.12), which is somewhat higher in the current sample. All summary scores for the SB-FE yielded smaller standard deviations (SD range 10.31 to 14.57) than the normative SD of 16, indicating a restricted range of scores in the current group of children.

Examining the individual subtests, obtained means for all of the subtests in the Verbal Reasoning and Short-Term Memory areas were comparable to the mean of 50 in the normative sample (i.e., all within one SEm). In the Abstract/Visual Reasoning area, although Pattern Analysis is comparable to the normative mean, Copying was somewhat lower (M = 45.38) and Matrices was somewhat higher (M = 52.65). In the Quantitative area, the Quantitative subtest was comparable to the normative mean although Number Series was higher (M = 54.25). Both subtests that were higher are those that were administered to older or advanced children, which may explain the higher mean scores. Nearly all of the obtained subtest standard deviations were lower than the normative standard deviation of 8.

Table 2 presents the descriptive statistics for the WISC-III summary and subtest scores. These data reveal that the WISC-III FSIQ is within the SEm of the normative mean of 100. Moreover, all of the obtained summary scores are within the SEm of the normative means of 100, although the Processing Speed index mean (M = 105.69) is almost greater than the normative mean (SEm = 5.83). The standard

Table 2. Descriptive Statistics for the WISC-III Summary and Subtest Scores (N = 60)

	Mean	SD
WISC-III FSIQ	**99.95**	**11.75**
Verbal IQ	**98.22**	**11.76**
Information	9.82	2.85
Similarities	9.90	2.89
Arithmetic	9.03	2.77
Vocabulary	9.12	2.44
Comprehension	10.33	2.90
Digit Span	9.60	2.89
Performance IQ	**102.25**	**13.23**
Picture Completion	9.12	2.44
Coding	11.07	2.87
Picture Arrangement	9.65	3.41
Block Design	10.58	3.66
Object Assembly	10.97	2.61
Symbol Search	10.69	3.14
Verbal Comprehension	**99.20**	**12.61**
Perceptual Organization	**101.22**	**13.21**
Freedom from Distractibility	**97.35**	**12.86**
Processing Speed	**105.69**	**13.34**

Table 3. Intercorrelations for SB-FE and WISC-III Summary Scores [a]

	SB-FE Comp IQ	VR	A/VR	Quan	STM	WISC-III FSIQ	VIQ
SB-FE Composite IQ							
Verbal Reasoning SAS	.80**						
Abstract/Visual Reasoning SAS	.79**	.49**					
Quantitative Reasoning SAS	.77**	.45**	.58**				
Short-Term Memory SAS	.78**	.67**	.42**	.34*			
WISC-III FSIQ	**.86****	**.73****	**.70****	**.68****	**.63****		
Verbal IQ	.81**	.82**	.55*	.62**	.61**	.87**	
Performance IQ	.66**	.42	.65**	.54*	.46*	.86**	.50**

[a] Intercorrelations *within* each IQ test are based on $N = 60$; Correlation *between* IQ tests are based on $N = 30$.
* $p \leq .01$
** $p \leq .001$

deviations for all WISC-III summary scores were smaller (*SD* range 11.75 to 13.34) than the normative SD of 15, indicating a restricted range of scores in the current sample of participants.

Clinicians often report on VIQ-PIQ differences and an examination of this information indicates that the children in the present sample obtained a similar frequency of significant VIQ-PIQ differences (35% greater than 11 point difference compared to 36% in the normative sample). With regard to the individual WISC-III subtests, all obtained scores approximated the normative means of 10 and standard deviations of 3.

Intercorrelations

Table 3 presents the Pearson correlations between the SB-FE summary scores and the major WISC-III summary scores (index scores omitted to reduce the number of analyses in order to minimize Type I error). This table reveals a strong correlation between the WISC-III FSIQ and the SB-FE Composite IQ ($r = .86$, $p \leq .001$). Similarly, the SB-FE Verbal Reasoning SAS significantly correlated with the WISC-III VIQ ($r = .82$, $p \leq .001$), and the SB-FE Abstract/Visual Reasoning SAS correlated with the WISC-III PIQ ($r = .65$, $p \leq .001$). This association between the two IQ tests is supported by similar IQ scores on both tests obtained for the 7- to 8-year-old age group. Comparing the tests on ability classifications (i.e., Very Superior to Mental Retardation; see Sattler, 1992 for details on both tests), the two IQ tests placed 30% of the children in different classification ranges. An examination of these discrepancies reveals no systematic differences (i.e., neither IQ test was consistently easier or harder).

Examining the intercorrelations within the WISC-III indicates that both the VIQ and PIQ are strongly correlated with the FSIQ ($r = .87$ and $r = .86$, $p \leq .001$, respectively), similar to the values reported in the manual ($r = .92$ and $r = .90$, respectively). The correlation between VIQ and PIQ is somewhat weaker ($r = .50$, $p \leq .001$) than reported in the manual for the comparable age group (range .57 to .70), although the value is similar to that reported by Silva et al. (1981) for the WISC-R. Clinicians interested in short-forms of the WISC-III often use a combination of the best predictors of VIQ and PIQ. In the current sample, VIQ was most strongly intercorrelated with Similarities ($r = .83$, $p \leq .001$) and Information ($r = .82$, $p \leq .001$); PIQ was most strongly intercorrelated with Block Design ($r = .76$, $p \leq .001$) and Picture Completion ($r = .71$, $p \leq .001$).

For intercorrelations within the SB-FE, the Composite IQ was significantly correlated (all $p \leq .001$) with the area SAS as follows: with the Verbal Reasoning SAS ($r = .80$) similar in magnitude to the comparable age group in the manual (range .80 to .87); with the Abstract/Visual Reasoning SAS ($r = .79$) similar to age group intercorrelations in the manual (range .79 to .86); with the Quantitative Reasoning SAS ($r = .77$) similar to age group coefficients in the manual (range .80 to .86); and with the Short-Term Memory SAS ($r = .78$) slightly lower but similar to the age group correlations reported in the manual (range .86 to .88). The SB-FE Composite IQ was most strongly intercorrelated with the Quantitative ($r = .73$, $p \leq .001$), Absurdities ($r = .70$, $p \leq .001$), and Vocabulary ($r = .67$, $p \leq .001$) subtests.

Discussion

The current study evaluated the applicability of the Australian adaptations of the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991) and the Stanford-Binet Intelligence Scale—Fourth Edition (SB-FE; Thorndike et al., 1986) with Dunedin primary school children. The study investigated the possible existence of differences across countries in WISC-III and SB-FE test performance (reflecting potential differences in demographics between New Zealand and the United States), in a preliminary assessment of the suitability of American norms with New Zealanders. Sixty children aged 5 to 8 received the SB-FE and sixty children aged 7 to 10 received the WISC-III (with 7- and 8-year-olds receiving both tests).

Previous research with earlier versions of both IQ tests (e.g., Silva, 1982) suggested that scores obtained by the Dunedin sample would be higher than the norms for American children. However, no notable differences across countries in WISC-III or SB-FE test performance emerged. Indeed, the Dunedin children obtained a normative distribution comparable to their American counterparts, with means closely approximating those obtained in the U.S. standardization sample for the IQ scores.

For the SB-FE, the Composite IQ and area scores were within the standard error of measurement with the exception of the area assessing quantitative reasoning, which appeared easier for this Dunedin sample. Nearly all of the obtained subtest means were comparable to the norms. Because children of different ages and abilities receive a different number of SB:FE subtests, only a subset of young brighter children or older children were eligible to receive 12 subtests. The two SB:FE subtests with higher mean scores were administered to this subset of children. Thus, the obtained higher mean scores potentially reflect that the younger children receive high scores for responding correctly to few items because, relative to their peers, the ability to respond to any portion of that subtest is superior. For the WISC-III, the summary and subtest scores were within the standard error of measurement of the normative means. These results for the Australian adaptations of the IQ tests provide some support for the applicability of the measures with New Zealand children.

For both the SB-FE and WISC-III summary scores, standard deviations were lower than those reported for the standardization samples of each test. This finding suggests a restricted range of scores in the current group of children, with insufficient numbers at the extreme ends of the distribution. Such a result may be attributable to sample size and characteristics (see discussion below), and a similar reduction in standard deviations for summary scores was observed in the sample involved in the development of the Australian adaptation of the SB-FE (de Lemos, 1989).

With regard to the subsample of children who received both IQ tests, a strong positive correlation between the summary scores emerged. The magnitude of these correlations is comparable to WISC-III and SB-FE correlation coefficients reported overseas for a sample of children referred for evaluation (Prewitt & Matavich, 1993). Although similar means were obtained on both tests, 30% of the children in the current sample were placed in different classification ranges. This classification difference suggests that, despite a strong association, the tests may tap somewhat different, albeit related, constructs, underscoring that an IQ test should never be used in isolation to gain a full profile of an individual's strengths and weaknesses.

The findings from the current sample indicate the absence of significant effects of several demographic factors (gender, family composition, family income, and parental education) on IQ summary scores. Some differences of interest approached statistical significance. Specifically, girls obtained somewhat higher SB-FE Composite IQ scores as well as in the area scores measuring perceptual skills and short-term memory. In addition, higher paternal educational attainment was marginally associated with higher scores in verbal skills on the WISC-III and SB-FE. Higher maternal educational attainment was associated with marginally higher scores in perceptual-organizational skills on the WISC-III and SB-FE. A larger sample of New Zealanders would be needed to replicate this finding in order to clarify whether parental educational level significantly contributes differentially to IQ test performance (see Sattler, 1992 for a review of the association between parental education and intelligence test scores).

Several limitations need to be considered in generalizing interpretations of the current study's results to other New Zealanders. The current sample did not provide an adequate representation of several sociodemographic attributes characterizing Dunedin or New Zealand as a whole. Most importantly, some ethnic groups were not sufficiently represented in the current sample of children. Specifically, only 3.4% were identified by their parents to be of Maori descent, although the 1991 New Zealand Census (Department of Statistics, 1992) identified that this ethnic group constitutes 20% of children aged 5 to 14 in New Zealand and 6.8% of such children in Dunedin. The 1991 N.Z. Census also reports that 11% of New Zealand children and 6.6% of Dunedin children belong to other ethnic groups (including those of Pacific Island, Asian, or Indian origin) compared to the 4.5% of such ethnic groups obtained in the current sample. Consequently, a greater percentage of the current sample (92.1%) identified with the Pakeha/European New Zealander group than reported in the 1991 N.Z. Census (69% of children in New Zealand, 86.6% in Dunedin). Post-hoc statistical analyses excluding the non-Pakeha children found no difference in the current study's results. The relative ethnic homogeneity of the current sample may have also contributed to reduced variability in scores and augmented the comparability of obtained scores with the U.S. norms. Most importantly, the generalizability of the current findings to New Zealand Maori and Polynesian ethnic groups is precluded, and thus any of the obtained findings may not apply to non-Pakeha children without further investigation.

With regard to parental educational attainment, the current group of children tended to have slightly *more* educated mothers (51% with above Form 7 education) compared to either New Zealand women (42%) or Dunedin women (47%) as reported by the 1991 N.Z. Census. In contrast, the current group of children had fathers with representatively *less* education (36% with above Form 7 education) than in either New Zealand as a whole (50%) or Dunedin (54%). With respect to family composition, a smaller percentage (15.5%) of the current sample of children came from single parent homes than is typical in New Zealand (24.5%) or Dunedin (22.4%). However, the sample's annual family income approximates the distribution for both New Zealand and Dunedin at all income levels (Department of Statistics, 1992).

A further limitation of the current study pertains to the age range selected for the current study, and interpretations should not be generalized to children outside the 5- to 8-year-old range for the SB-FE or outside the 7- to 10-year-old range for the WISC-III. Additionally, only sixty children were selected for each measure, and larger sample sizes are required to increase confidence in the obtained results. The lowered standard deviations may have been influenced by the restricted age range and sample size.

The absence of significant differences in New Zealander's SB-FE and WISC-III test performance in the current study is not consistent with previous findings on their predecessors (e.g., Cumming & March, 1985; Silva et al., 1981). This discrepancy suggests the WISC-R and SB-LM's relative unsuitability with a New Zealand sample may reflect that these versions of the measures did not have official Australian adaptations. The current findings provide preliminary support for the WISC-III and SB-FE's status as superior instruments relative to the previous versions in terms of their applicability with New Zealand children. These Australian adaptations may have addressed some of the difficulties regarding item content.

However, item content comprehension alone does not eliminate concerns about the effect of item content on sequencing. Although the specific content may be understood, items may still be inappropriately placed in terms of their difficulty level. Inappropriately sequenced items can lead to the establishment of premature ceilings, thus underestimating the performance of individuals. Such item sequencing difficulties were reported in the sample involved in the development of the Australian adaptation of the SB-FE, leading to "double ceilings" (de Lemos, 1989). Such instances of premature and double ceilings were observed in the current sample as well (see Treacy, 1996 for details on item sequencing difficulty), and a more flexible approach to establishing ceiling levels might be appropriate for New Zealand children as recommended by the SB-FE Australian adaptation researchers (de Lemos, 1989).

Overall, the current study's findings that the Australian adaptations of the WISC-III and SB-FE adequately differentiate intellectual abilities of different age groups demonstrates the clinical utility of these instruments for Dunedin children. The findings attest to the improved status of the Australian adaptations in relation to their predecessors in terms of applicability for New Zealand children, providing some reassurance that the U.S. normative data may be reasonably employed in this country. In particular, the SB-FE intelligence scale extends the age range, allowing assessment of children too young for the WISC-III. The present study provides an initial glimpse of the applicability of two of the most popular intelligence tests with Dunedin children. However, large-scale norming across all of the ages of intended use is clearly still warranted, with appropriate stratification across all of the relevant demographic characteristics as well as representation of various geographic regions of New Zealand.

Clearly, many issues remain unsettled, particularly with regard to the stability of intelligence test scores. Concerns about the stability of scores reflect continuing uncertainty about whether practitioners are aiming to assess current functioning or to predict future functioning. Thus, opponents of intelligence tests often object because of concerns about the purpose of such tests (see Tuck, 1983 for discussion). Although the measures offer some promise with regard to applicability for New Zealanders, doubts and questions about the purpose, use and misuse of such measures will persist without further research and, conceivably, without further debate.

References

Ballard, K. D. (1988). Interpreting Stanford-Binet and WISC-R IQs in New Zealand: The need for more than caution. In M. Olssen (Ed.), *Mental testing in New Zealand: Critical and oppositional perspectives.* Dunedin, New Zealand: University of Otago Press.

Chapman, J. W., & St. George, R. (1984). The WISC-R: Results with a New Zealand Sample and Relationships with TOSCA and PATs. *New Zealand Journal of Educational Studies, 19,* 184-188.

Cumming, K. J., & Marsh, R. W. (1985). A Comparison of British Ability Scale IQs with those from the WISC-R and Stanford-Binet (L-M) scales. *New Zealand Journal of Educational Studies, 20,* 190-191.

De Lemos, M. (1989). *The use of the Stanford Binet Intelligence Scale in Australia.* Victoria, Australia: Australian Council for Educational Research.

Department of Statistics. (1992). *1991 New Zealand Census.* Te Tari Tatau, Wellington, New Zealand: Author.

Fernando, K. (1995). *The meaning of the WISC-III Third Factor.* Unpublished PhD thesis, University of Auckland.

Olssen, M. (1988). *Mental testing in New Zealand: Critical and oppositional perspectives.* Dunedin, New Zealand: University of Otago Press.

Prewett, P. N., & Matavich, M. A. (1994). A comparison of referred students' performance on the WISC-III and the Stanford-Binet Intelligence Scale: Fourth Edition. *Journal of Psychoeducational Assessment, 12,* 42-28.

Sattler, J. M. (1992). *Assessment of children* (Revised and Updated Third Edition). San Diego, CA: Jerome M. Sattler Publisher.

Shuker, R. (1988). The Otis Test: Its development and use. In M. Olssen (Eds.), *Mental testing in New Zealand: Critical and oppositional perspectives.* Dunedin, New Zealand: University of Otago Press.

Silva, P. A. (1982). Interpreting Stanford-Binet and WISC-R IQs in New Zealand. *New Zealand Journal of Educational Studies, 17,* 195.

Silva, P. A., McGee, R. O., & Williams, S. M. (1981). *From birth to seven: Child development in Dunedin: A multidisciplinary study.* Report available University of Otago Medical School.

Share, D. L., McGee, R. O., & Silva, P. A. (1989). IQ and reading progress: A test of the capacity notion of IQ. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 97-100.

St. George, R. & Chapman, J. W. (1987). WISC-R item characteristics: A study with 11-year-old New Zealand children. *New Zealand Journal of Psychology, 16,* 28-36.

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: Manual for the third edition Form L-M.* Boston: Houghton Mifflin.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition.* Chicago, IL: Riverside Publishing.

Treacy, L. (1996). The applicability of the Wechsler Intelligence Scale for Children (Third Edition) with a New Zealand sample. Unpublished fourth year honours thesis, University of Otago, Dunedin, New Zealand.

Tuck, B. (1983). Education and tests of scholastic ability: Is there a baby in the bath water? *New Zealand Journal of Educational Studies, 18,* 165-171.

Tuck, B. F., Hanson, A. L., & Zimmerman, M. (1975). The WISC-R: A New Zealand study of norms and validity. *New Zealand Journal of Educational Studies, 10,* 52-58.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children - Third Edition.* San Antonio, TX: Psychological Corporation.

Address for correspondence:

Christina M. Rodriguez, PhD

now at

Arizona State University West

Social and Behavioral Sciences

4701 West Thunderbird Road

PO Box 37100

Phoenix

AZ 85069-7100

E-mail: CRODRIGU@asuvm.inre.asu.edu