

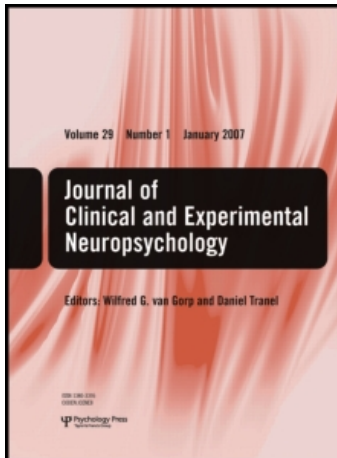
This article was downloaded by: [Chinese University of Hong Kong]

On: 23 September 2009

Access details: Access Details: [subscription number 789751637]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Clinical and Experimental Neuropsychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713657736>

### Measuring vocabulary by free expression and recognition tasks: Implications for assessing children, adolescents, and young adults

Agnes S. Chan <sup>a</sup>; Mei-chun Cheung <sup>b</sup>; Sophia L. Sze <sup>a</sup>; Winnie W. Leung <sup>a</sup>; Rebecca W. Y. Cheung <sup>a</sup>

<sup>a</sup> Neuropsychology Laboratory, Department of Psychology, The Chinese University of Hong Kong, Hong Kong

<sup>b</sup> Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong

First Published on: 14 March 2008

**To cite this Article** Chan, Agnes S., Cheung, Mei-chun, Sze, Sophia L., Leung, Winnie W. and Cheung, Rebecca W. Y. (2008) 'Measuring vocabulary by free expression and recognition tasks: Implications for assessing children, adolescents, and young adults', *Journal of Clinical and Experimental Neuropsychology*, 30:8, 892 — 902

**To link to this Article:** DOI: 10.1080/13803390701861384

**URL:** <http://dx.doi.org/10.1080/13803390701861384>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Measuring vocabulary by free expression and recognition tasks: Implications for assessing children, adolescents, and young adults

Agnes S. Chan,<sup>1</sup> Mei-chun Cheung,<sup>2</sup> Sophia L. Sze,<sup>1</sup> Winnie W. Leung,<sup>1</sup>  
and Rebecca W. Y. Cheung<sup>1</sup>

<sup>1</sup>Neuropsychology Laboratory, Department of Psychology, The Chinese University of Hong Kong, Hong Kong

<sup>2</sup>Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong

Vocabulary tests are commonly used for assessing verbal ability. Most vocabulary tests employ the free expressive format that taps both verbal knowledge and expressive ability. The recognition format, which requires less expressive ability, has been suggested to be more sensitive in assessing the fund of verbal knowledge. We assessed vocabulary performance of 285 normal individuals (aged 6 to 23 years) using both free expressive and recognition tasks. Results showed that participants aged 6 to 15 performed significantly better on the recognition than on the free expressive task. While the recognition task significantly correlated with the Test of Nonverbal Intelligence–Third Edition (TONI-III) IQ for individuals aged 9 and above, the free expressive task correlated with TONI-III IQ only for young adults. Shortened 18-item versions yielded high reliability and correlation with the full version and reliable association with TONI-III IQ. These results highlight the utility of recognition tasks in measuring the fund of knowledge and suggest the possibility of developing shorter vocabulary tests for more cost-effective assessment.

**Keywords:** Vocabulary test; Chinese; Free expressive; Recognition; Multiple choice.

## INTRODUCTION

Vocabulary tests have been included in major tests of intellectual functioning, such as the Wechsler (Wechsler, 1997, 2003), Stanford–Binet (SB; Roid, 2003), Kaufman (Kaufman & Kaufman, 1990), and Shipley (Shipley, 1986) intelligence scales. Vocabulary tests have also been developed as stand-alone tests commonly used for estimating IQ, such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981, 1997), the Expressive One-Word Picture Vocabulary Test–Revised (Gardner, 1990), and the Mill Hill Vocabulary Test (Raven, 1982). These vocabulary tests employ different testing formats and may be categorized into “expressive” vocabulary tests—where the participant is required either to give free expressive

responses to explain the meaning of vocabulary items (e.g., the Vocabulary subtests of the Wechsler and Stanford–Binet scales) or to hear a definition of a word or see a picture, then provide a one- or two-word response consistent with the definition or picture (e.g., Kaufman Assessment Battery for Children–Second Edition, K-ABC-II; Kaufman & Kaufman, 2004)—and “receptive” vocabulary tests—where the participant is required to identify the meaning of a word out of a few verbal (e.g., the Shipley and Mill Hill vocabulary tests) or pictorial (e.g., the Peabody Picture Vocabulary Test) options.

Many of these well-accepted vocabulary tests were found to have good psychometric properties. For example, Vocabulary is the most reliable subtest in the Wechsler Intelligence Scale for

Address correspondence to Agnes S. Chan, Department of Psychology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR (E-mail: aschan@psy.cuhk.edu.hk).

Children–Revised (WISC-R; Sattler, 1992, Chapter 7), has one of the highest test–retest correlations (Iverson, 2001), is the subtest that best predicts verbal (Wechsler, 1991) and full-scale (Chang, Tang, & Chan-ho, 1995; Wechsler, 1958) IQ, and has minimal practice effects (McCaffrey, Duff, & Westervelt, 2000). Sex differences are negligible (Kaufman, McLean, & Reynolds, 1991; Snow & Weinstock, 1990), and they are not likely to be affected by academic motivation or achievement (Anastasi, 1988; Vernon, 1979). In practice, vocabulary tests were also easy to be constructed and administered (Ullstadius, Gustafsson, & Carlstedt, 2002). In addition, it resists the regressive effect of aging and remains relatively stable throughout the life span of normal individuals (Bayles, Tomoeda, & Boone, 1985; Obler & Albert, 1985; Schum & Sivan, 1997). It has been found that vocabulary is resistant to neurological and psychological disturbances including early dementia (Melvold, Au, Obler, & Albert, 1994; Sullivan, Sagar, Gabrieli, Corkin, & Growdon, 1989), traumatic brain injury (Brooks & Aughten, 1979), and diffuse brain damage (Irving, 1971; McFie, 1975; Zillmer et al., 1992). All these advantages have made vocabulary tests one of the most widely used mental ability tests, whether alone or as part of test batteries (Lezak, Howieson, & Loring, 2004, Chapters 4, 8, & 13).

While the vocabulary test has been very useful in assessing level of verbal knowledge and estimation of general intelligence, it is worth noting that free expressive format has been the mainstream in vocabulary testing for decades, as can be seen in the Wechsler Intelligence Scales for Children (WISC to WISC-IV; Wechsler, 1949, 1974, 1991, 2003), Wechsler Adult Intelligence Scales (WAIS to WAIS-III; Wechsler, 1958, 1981, 1997), Stanford–Binet Intelligence Scales (SB to SB-IV; Roid, 2003; Terman & Merrill, 1960; Thorndike, Hagen, & Sattler, 1986), and the Kaufman Adolescent and Adult Intelligence Test (KAIT) and Kaufman Assessment Battery for Children (K-ABC, K-ABC-II; Kaufman & Kaufman, 1983, 2004). These tests require the test-taker to have good expressive language ability in order to effectively express their vocabulary knowledge. Thus, they are measuring both the semantic knowledge of words and the expressive verbal ability of individuals (Nation, 1990). This raises some concerns as the actual fund of knowledge of some individuals with relatively poorer expressive ability may be underestimated by the free expressive format of vocabulary testing. For instance, some brain dysfunctions such as damage of the frontal region (Banich, 1998; Damasio, 1991; Novoa &

Ardila, 1987; Rey, Levin, Rodas, Bowen, & Nedd, 1994) or subcortical region (Butters, 1984; Butters, Wolfe, Martone, Granholm, & Cermak, 1985; Butters, Wolfe, Granholm, & Martone, 1986; Nagaratnam & Gilhortra, 1998; Rey et al., 1994) adversely affect the individuals' expressive ability. These patients tend to be impaired on free-recall tasks but could perform relatively normally on recognition tasks (Butters, 1984; Butters et al., 1986; Butters et al., 1985; Massman, Delis, Butters, Levin, & Salmon, 1990). Thus, Kaplan and her colleagues have highlighted the importance of using tests that are less dependent on expressive ability (e.g., multiple choice) to test the individuals' fund of knowledge (Joy, Fein, Kaplan, & Freedman, 1999), and the recognition format has been incorporated into some of the intelligence tests (Wechsler Adult Intelligence Scales–Revised as a Neuropsychological Instrument, WAIS–R NI, Kaplan, Fein, Morris, & Delis, 1991; WISC-IV Integrated, Wechsler et al., 2004).

The wide acceptance of the intelligence scales (Roid, 2003; Wechsler, 1997, 2003) and their Vocabulary subtests has turned the free expressive response format into the mainstream of vocabulary testing. How the free expressive and recognition formats differ in their estimation of verbal knowledge and/or intellectual functioning remains unknown. Thus, the present study aimed to evaluate the performance of different age cohorts (from 6 to 23 years old) on free expressive (i.e., spontaneous) and recognition (i.e., multiple choice) vocabulary tasks. The goal was to examine whether test format has an effect on test scores, an issue that has important implications as this would lead to differential estimation of vocabulary knowledge and intellectual functioning of individuals.

As a subanalysis of this study, we also examined the sensitivity of shorter 18-item versions of the vocabulary test against the 36-item full version. The more commonly used vocabulary tests such as those in the WISC-IV and WAIS-III consisted of 36 and 33 items, respectively (Wechsler, 1997, 2003). Administration is time consuming especially as they are administered using the free expressive format. Given the usually large battery of tests to be administered at a single clinical session, it is of practical and clinical value to have a short yet sensitive test, so as to reduce fatigue of the individual assessed, and to make clinical assessments more cost-effective.

It is noted that there is a lack of well-established vocabulary tests for the Chinese-speaking population (Chan, Shum, & Cheung, 2003). Some attempts have been made to develop Chinese vocabulary tests in Taiwan (Chen & Chen, 2002)

and China (Dai, Gong, & Zhong, 1990; Gong, 1989). However, since vocabulary is known to be sensitive to language and culture (Tamayo, 1987), these tests may not be ecologically valid for the Chinese-speaking population in Hong Kong as both the language (Cantonese in Hong Kong, Putonghua in Taiwan and China) and culture in Hong Kong are very different from those in Taiwan and China. Given that the vocabulary test is an important assessment for assessing general intelligence and the level of verbal ability, the results from this study would on one hand provide useful information on vocabulary testing in general and on the other produce an ecologically valid assessment instrument developed specifically for the local Chinese-speaking population in Hong Kong.

## METHOD

### Scale construction and item selection

The Chinese vocabulary test used in the present study consisted of 36 items. The items were selected from local textbooks recommended by the Hong Kong Education Department (1999a, 1999b) for Chinese language, the main subject from which students learn the language as well as most vocabulary. Textbooks from five publishers were selected for each grade level from Primary One to Form Seven (Grades 1 to 13). For primary level (Grades 1 through 6), the five publishers were: Modern Educational Research Society Ltd; Keys Press; New Asia Publishing House Ltd; Educational Publishing House Ltd; and Everyman's Book Co. Ltd. This list was confirmed with advice from 10 experienced primary school teachers. For secondary level (Grades 7 through 13), the five publishers were Educational Publishing House Ltd, Everyman's Book Co. Ltd, Ling Kee Publishing Co. Ltd, Macmillan Publishers (China) Ltd, and Keys Press, also confirmed after consultation with 10 experienced secondary school teachers.

For each grade level, the frequency of words that appeared on the vocabulary list of the five textbooks was calculated. Frequency was defined as the co-occurrence of a particular word in the five textbooks for that grade level. A frequency of 100% meant that the word appeared in all five textbooks, while a frequency of 80% meant that the word appeared in four textbooks, and so on. Using words with frequency of at least 40%, a preliminary list of 51 items that included vocabulary items from the 13 grade levels was constructed.

### Development of scoring criteria

A scoring system consisting of three score values of 2, 1, and 0 points for each item was developed. The score for each item was determined according to the degree of accuracy in explaining the vocabulary item. The criteria for a score of 2 required an accurate and complete explanation of the item. Scores of 1 and 0 were given, respectively, to partially correct and incorrect explanations of the item. The explanations of the items were based on the textbooks' glossaries and the reputable Chinese dictionary *Cihai* (Cihai Editorial Board, 1979).

### Pilot study

A pilot study was conducted to select and order the items and to refine the scoring criteria. The item list was administered to 241 (133 males and 108 females) Cantonese-speaking children and adults, with educational background ranging from Primary One to doctoral level. In order to develop a list with reasonable difficulty level, the items that were too difficult or too easy were eliminated from the original list based on the inclusion criteria that at least 5% but no more than 85% of the total participants could provide the correct answers. Finally, 36 items were selected from the original 51 items, a number comparable to that of some popular vocabulary tests such as the WISC-IV and WAIS-III that consist of 36 and 33 items, respectively. The sequence of the items was also re-ordered beginning with the item with the lowest difficulty level. The scoring criteria were further refined through eliminating confusion and adding more sample responses to each score category.

### Development of the multiple-choice version

A multiple-choice (MC) version of the 36-item vocabulary test was developed. The MC version consisted of the same 36 items as those in the original spontaneous (SPN) version. There were five possible answer options to choose from for each item; a score of 2 was assigned to the target option, a score of 1 was assigned to the partially correct option, and scores of 0 were assigned to the remaining three incorrect options. The development of the scoring system for MC version was based on the results collected in the pilot test. The target option in the MC was equivalent to the 2-point answer in the free expression task. The 1-point option partially explained the semantic meaning of the vocabulary whereas the remaining three options were incorrect answers commonly

provided by participants in the pilot test. For example, for the item "Winter," the five options were: (a) wearing warm clothing; (b) snowing; (c) the cold season; (d) the coldest days; and (e) dry weather. A score of 2 was assigned to the target option (c), a score of 1 was assigned to the partially correct option (d), and scores of 0 was assigned to the other options.

## Participants

A total of 285 normal individuals, aged 6 to 23 years, were recruited from primary and secondary schools and a university in Hong Kong (Table 1). Primary and secondary participants were recruited through invitations to the schools, and university participants were undergraduates recruited from the Psychology Department who participated to fulfill course requirement. Written consents were obtained from all secondary and college students, and parental consents were also obtained for all primary students prior to test administration. The protocol was approved by the Joint Chinese University of Hong Kong–New Territories Eastern Cluster Clinical Research Ethics Committee, and the study was conducted according to the Declaration of Helsinki. All participants participated voluntarily and completed both the SPN and the MC versions of the vocabulary test.

## Materials

### Vocabulary test

The test consisted of 36 items, and each item was a two-character Chinese word. Both the SPN and MC versions were administered individually to the participant. In the SPN version, participants were asked to give free expressive responses to explain

the meaning of the 36 items. Each item was printed on a stimulus card. The examiner presented the card and read the item to the participant. The participant then gave a verbal explanation of the item. Each item was scored 0, 1, or 2 for totally incorrect, partially correct, or fully correct responses. An SPN score was computed by summing the scores for the 36 items and ranged from 0 to 72. The MC version consisted of the same 36 items. Instead of giving free responses, participants were given five answer options to choose from. Each item and the five answer options were printed on a stimulus card, which was shown to the participant. The examiner read out the vocabulary item, followed by the five options in sequence, and then asked the participant to choose the option among the five that best explained the vocabulary item. Scores for each item also ranged from 0 to 2, giving a total MC score that ranged from 0 to 72.

### Test of nonverbal intelligence

The Test of Nonverbal Intelligence–Third Edition (TONI-III; Brown, Sherbenou, & Johnsen, 1992) was administered to assess the IQ of each participant. The test consisted of 45 matrix reasoning questions, with raw scores ranging from 0 to 45. The raw scores were then converted to deviation quotients based on the norms provided in the test manual. The TONI-III was used because the test was relatively language free so it would less be affected by language ability, reducing the covariance on similar cognitive ability as measured by the vocabulary test.

### Procedure

Participants were assessed individually using a test battery that included the TONI-III and the

**TABLE 1**  
Demographic characteristics of the participants

Age group <sup>a</sup>	Gender		Age <sup>a</sup>		Education <sup>a</sup>		TONI-III IQ	
	Male	Female	Mean	SD	Mean	SD	Mean	SD
6	11	12	6.00	—	0.04	0.21	114.78	13.25
7	10	15	7.00	—	1.00	0.29	110.08	16.01
8	9	13	8.00	—	2.00	0.31	108.95	18.44
9	11	11	9.00	—	3.14	0.47	110.68	19.26
10–11	17	23	10.45	0.53	4.53	0.51	108.03	17.64
12–13	12	4	12.50	0.52	6.13	0.50	104.75	16.57
14–15	15	18	14.70	0.47	7.73	1.01	104.76	16.82
16–23	40	64	18.84	1.59	11.89	1.61	110.89	15.23

Note. TONI-III = Test of Nonverbal Intelligence–Third Edition.

<sup>a</sup>In years.

vocabulary test. The SPN version of the vocabulary test was administered first followed by the MC version. This ordering was to avoid carryover of practice effects from the MC version (where one correct option and one partially correct option were presented to the participant for each vocabulary item) to the SPN version. TONI-III was administered after the vocabulary test. At the end of the assessment, participants were briefly reported on their performance on the TONI-III.

## Data analyses

To examine the item characteristics of the test, item analyses including item difficulty level (i.e., the percentage of participants scoring 2 on each item) and item-total correlation were performed on the 36 items. Internal consistency of the scale was also calculated to examine the reliability of the scale. To examine the effect of age on test performance, planned comparisons using post hoc contrasts in analysis of variance (ANOVA) were performed on the SPN score to delineate the age groups. To examine the discrepancy in performance on the SPN and MC versions, the two scores were compared across the age groups using an 8 (age group)  $\times$  2 (version) ANOVA with repeated measures and post hoc paired *t* tests. The effect of gender was tested using a 2 (gender)  $\times$  2 (version) ANOVA with repeated measures. Correlation with IQ was examined with bivariate correlation between TONI-III IQ and the SPN and MC scores. In testing the validity of the shorter versions of the test, correlations were computed between the shorter and the full versions and between the shorter versions and TONI-III IQ. Internal consistency, item-total correlation, and paired *t* tests were conducted to examine the psychometric properties of the two shorter versions.

## RESULTS

### Reliability and item analysis of the vocabulary test

Item analyses included item difficulty level and item-total correlation of the 36 items (Table 2). The difficulty level of each item was the percentage of participants getting a score of 2. Results indicated that the items had a good range of difficulty levels that ranged from 84.6 to 6.3% for the SPN version and 89.8 to 15.8% for the MC version. Reliability was examined using Cronbach's alpha on internal consistency. Reliability of the SPN and

the MC versions were .96 and .87, respectively, indicating that both versions had adequate reliability.

### Cohort and gender effect on test performance

#### Cohort effect

The performance of normal participants on the SPN and the MC versions is shown in Figure 1. Visual examination showed that scores on both SPN and MC versions increased sharply from age cohorts 6 through 9, became attenuated from age 10 through 15, and leveled off from age 16 through 23.

ANOVA results indicated that there was significant age cohort differences in both SPN,  $F(17, 267) = 97.54, p < .001$ , and MC scores,  $F(17, 267) = 56.12, p < .001$ . Results from the planned comparisons (alpha level adjusted for family-wise errors) were generally consistent with the visual examination. Specifically, age cohorts 6 to 9 were significantly different from the subsequent cohorts (age 6 vs. 7, estimate =  $-4.32, p = .033$ ; age 7 vs. 8, estimate =  $-5.79, p = .005$ ; age 8 vs. 9, estimate =  $-7.09, p = .001$ ; age 9 vs. 10, estimate =  $-7.59, p < .001$ ). Differences between age cohorts 10 and 11 (estimate =  $-1.06, p = .632$ ), 12 and 13 (estimate =  $2.75, p = .431$ ), and 14 and 15 (estimate =  $-1.49, p = .573$ ) were not statistically significant. All differences between successive cohorts from age 16 through 23 were not statistically significant (all  $ps > .05$ ). The means and standard deviations are thus presented with age cohorts with no significant difference grouped together (Table 3).

#### Gender effect

We examined the effect of gender on the performance of the vocabulary test using ANOVA with repeated measures, with gender (male vs. female) as between-subject factor and version (SPN vs. MC) as within-subject factor. The Gender  $\times$  Version interaction was nonsignificant, Wilks's lambda,  $F(1, 283) = 0.01, p = .92$ , indicating that the discrepancy between SPN and MC versions was similar for males and females. In addition, no main effect of gender on test performance,  $F(1, 283) = 0.36, p = .55$ , on both SPN (male,  $M = 35.19, SD = 17.53, n = 125$ ; female,  $M = 36.21, SD = 18.65, n = 160$ ) and MC (male,  $M = 42.64, SD = 12.81, n = 125$ ; female,  $M = 43.77, SD = 12.22; n = 160$ ) versions. These results suggested that males and females performed similarly on both the SPN and MC versions of the vocabulary test. Hence we combined males and females in all subsequent analyses.

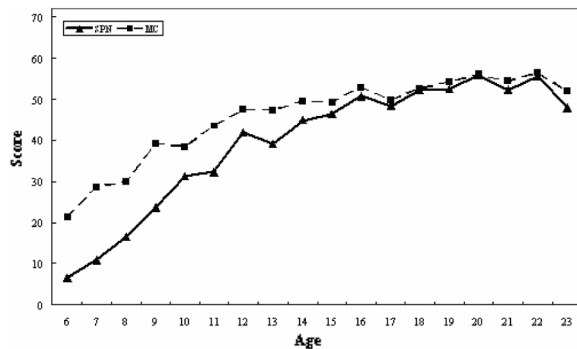
**TABLE 2**  
Item ordering, item difficulty, and corrected item-total correlation of the spontaneous and multiple-choice scores of the vocabulary test

New order	Items	English meaning	Spontaneous		Multiple choice	
			Item difficulty	Item-total correlation	Item difficulty	Item-total correlation
1	學校	school	84.6	.52	81.1	.55
2	寬恕	forgive	71.6	.68	73.0	.42
3	著名	famous	71.6	.63	51.2	.31
4	注意	attention	61.4	.57	61.1	.36
5	特別	special	60.4	.67	57.2	.65
6	干擾	disturb	58.9	.62	26.7	.27
7	冬天	winter	56.8	.49	89.8	.30
8	綽號	nickname	54.4	.50	40.4	.34
9	觀察	observe	52.6	.62	71.9	.58
10	突然	sudden	50.5	.39	65.6	.22
11	打掃	cleanse	47.4	.51	24.9	.02
12	本領	strengths	47.0	.68	54.4	.42
13	幻想	imagine	46.7	.68	53.3	.43
14	樸素	down-to-earth	43.9	.73	48.4	.49
15	點綴	decorate	43.2	.76	54.7	.61
16	蛻變	evolve	42.1	.67	51.6	.42
17	踴躍	keen	37.5	.60	31.2	.57
18	奉承	please	37.2	.74	51.2	.56
19	奢侈	extravagant	36.8	.63	46.0	.32
20	倒影	reflection	36.5	.48	58.6	.36
21	俘虜	hostage	34.7	.62	53.7	.43
22	孝順	filial piety	31.9	.64	84.2	.54
23	歡迎	welcome	27.7	.57	63.5	.05
24	豐富	abundant	27.4	.63	36.1	.41
25	羨慕	envy	19.6	.52	20.0	.42
26	魁梧	hefty	18.9	.74	57.9	.65
27	風俗	custom	18.2	.69	37.5	.64
28	演繹	rendition	17.9	.51	41.4	.53
29	醞釀	brew	17.2	.53	36.5	.34
30	遺憾	regret	16.8	.60	38.2	.06
31	恭敬	revere	15.8	.63	73.3	.51
32	瀏覽	browse	15.4	.46	19.6	.24
33	驅遣	command	15.1	.60	18.6	-.06
34	啟蒙	initiate	14.7	.69	15.8	.01
35	氣餒	give up	8.4	.51	20.4	.41
36	渾然	entire	6.3	.71	67.0	.13

### Discrepancy between SPN and MC scores

Visual examination of the distribution of SPN and MC scores across different age cohorts in Figure 1 suggested a discrepancy in the SPN and MC scores in all age cohorts, with the greatest discrepancy in the younger cohorts, especially from age 6 to 11, which gradually diminished until the two scores became very close from age 17 to 23. To test whether this discrepancy is statis-

tically significant, we performed an ANOVA with repeated measures, with cohort (8 age cohorts) as between-subject factor and version (SPN vs. MC) as within-subject factor. Results indicated that there was a Version  $\times$  Cohort interaction,  $F(7, 277) = 34.70, p < .001$ , suggesting that the SPN–MC discrepancies were different for the eight age cohorts. Post hoc paired  $t$  tests were done to examine which cohort had significant SPN–MC discrepancies. Since eight comparisons



**Figure 1.** Performance of individuals on the spontaneous (SPN) and multiple-choice (MC) versions of the vocabulary test by age cohorts.

**TABLE 3**

Performance of normal children on the spontaneous and multiple-choice versions of the vocabulary test and paired *t* tests on the SPN–MC discrepancies

Age cohort	SPN		MC		Paired <i>t</i> tests		
	Mean	SD	Mean	SD	<i>t</i>	<i>df</i>	<i>p</i>
6	6.48	4.25	21.30	6.17	-12.56	22	.000
7	10.80	6.13	27.82	5.82	-21.84	24	.000
8	16.59	7.97	29.91	7.93	-7.19	21	.000
9	23.68	6.82	39.09	4.45	-11.30	21	.000
10–11	31.75	7.01	40.73	7.18	-8.28	39	.000
12–13	40.50	10.77	47.44	8.06	-4.63	15	.000
14–15	45.94	7.65	49.42	5.30	-2.90	32	.007
16–23	52.44	6.53	53.73	5.42	-1.93	103	.056

Note. SPN = spontaneous. MC = multiple choice.

were done, the Bonferroni adjustment gave an adjusted alpha level of .006. This alpha level was used in all subsequent comparisons involving the eight age cohorts.

Post hoc paired *t* tests indicated that the SPN–MC discrepancy was statistically significant for all but the 14–15 and 16–23 cohorts (Table 3), with greater discrepancies for the younger cohorts. This suggested that younger children, compared with older children and adolescents, performed significantly less well in spontaneously explaining the vocabulary than in recognizing their meaning. These results suggest that the MC version may be a more sensitive assessment for measuring the fund of knowledge of younger individuals.

### Correlation with general intellectual functioning

Since vocabulary tests have been used as an estimation of intellectual functioning, analyses were performed to examine the relationship between

SPN and MC scores and TONI-III IQ. As the cohort effect has been demonstrated on vocabulary test scores, we stratified the analysis by the eight age cohorts in order to control for the cohort effect. Bivariate correlations were computed between SPN and TONI-III IQ and between MC and TONI-III IQ.

Results indicated that while the SPN score was significantly correlated with TONI-III IQ for the 16–23 young adolescents cohort ( $r = .21, p < .05$ ), the MC score was significantly correlated with TONI-III IQ for younger cohorts aged 9 ( $r = .49, p < .05$ ), 10–11 ( $r = .35, p < .05$ ), 12–13 ( $r = .52, p < .05$ ), and 16–23 ( $r = .41, p < .001$ ), with 14–15 ( $r = .30, p = .09$ ) approaching significance. Results also showed that while the correlation was small for the SPN score, correlations were more substantial for the MC score. It should be noted that only the MC score for the 16–23 cohort remained significant after Bonferroni adjustment ( $p < .003$ ), which might possibly be an effect of the small sample size. Taken together, these results seem to provide further support for the notion that the SPN and MC formats might be assessing slightly different things, especially for individuals under 16.

### Comparison between the full and shorter versions

To examine the possibility of developing a shorter vocabulary test, we split the 36-item list into Form A and Form B, composed of the odd and even number items, respectively. Three pairs of items (2 and 3, 10 and 11, and 35 and 36) were swapped to give a higher correlation between the two forms.

Internal consistency coefficients were computed for Forms A and B on both the SPN and MC versions. Results indicated that internal consistency coefficients were higher for the SPN version (Form A = .91; Form B = .92) than for the MC version (Form A = .77; Form B = .78), but both forms have adequate reliability. Correlation with the full version was very high for both Form A (SPN = .98, MC = .94) and Form B (SPN = .98, MC = .94) and so was the correlation between themselves (SPN = .92, MC = .76).

To scrutinize the two forms on their comparability, paired *t* tests were done for each cohort (Table 4). Results indicated that for both the SPN and MC versions, scores between Forms A and B were comparable, with nonsignificant score differences at the Bonferroni corrected alpha level at  $p < .003$ .

**TABLE 4**  
Performance of the different age cohorts on Forms A and B of the vocabulary test

Age cohort	SPN							MC						
	Form A		Form B		Paired <i>t</i> tests			Form A		Form B		Paired <i>t</i> tests		
	Mean	SD	Mean	SD	<i>t</i>	<i>df</i>	<i>p</i>	Mean	SD	Mean	SD	<i>t</i>	<i>df</i>	<i>p</i>
6	3.48	2.23	3.00	2.34	1.37	22	.185	10.39	3.29	10.91	4.55	-0.50	22	.622
7	5.56	3.14	5.24	3.33	0.77	24	.448	13.28	4.57	15.44	3.48	-1.91	24	.068
8	8.18	4.16	8.41	4.55	-0.30	21	.767	15.18	3.92	14.73	4.76	0.59	21	.564
9	11.77	4.23	11.91	3.53	-0.17	21	.867	20.45	3.33	18.63	3.67	1.57	21	.131
10-11	15.78	4.30	15.98	3.63	-0.34	39	.740	20.73	3.53	20.00	4.95	0.97	39	.338
12-13	19.44	5.37	21.06	5.84	-2.09	15	.055	23.44	4.07	24.00	4.90	-0.56	15	.584
14-15	23.09	4.06	22.85	4.74	0.32	32	.754	24.73	3.13	24.70	3.62	0.04	32	.967
16-23	25.91	3.98	26.53	3.76	-1.51	103	.133	26.40	3.50	27.33	3.31	-2.28	103	.025

Note. SPN = spontaneous score. MC = multiple-choice score.

Bivariate correlations were computed between TONI-III IQ and both SPN and MC versions for both Forms A and B to examine the relationship between the shorter forms with intellectual functioning as measured by TONI-III. Results were less consistent than those on the full 36-item version. For the shorter SPN versions, SPN-Form B correlated significantly with TONI-III IQ for the cohorts aged 12-13 ( $r = .51, p < .05$ ) and 16-23 ( $r = .23, p < .05$ ). For the shorter MC versions, MC-Form A correlated significantly with TONI-III IQ in the cohort aged 9 ( $r = .34, p < .05$ ) and 16-23 ( $r = .32, p = .001$ ). MC-Form B correlated significantly with TONI-III IQ in the cohort aged 10-11 ( $r = .47, p < .05$ ) and 16-23 ( $r = .34, p < .001$ ). While these results did not exactly match those of the full 36-item version, the general pattern was similar in that MC-Form A and MC-Form B remained significant for the cohort aged 16-23 after Bonferroni adjustment ( $p < .002$ ), suggesting that there is reliable association between the two shorter MC forms and intellectual functioning in young adults.

## DISCUSSION

Results indicated that there were significant cohort differences in performance on the vocabulary test, in both the free expressive and the recognition test formats. Based on planned comparisons, eight age cohorts were formed that reflected a rising trend in both scores from cohorts aged 6 to 15, which corresponded to a transition between childhood and adolescence, and a more stable trend among the older age 16-23 cohort. No significant gender effect was found on test scores in our sample, which was in line with findings in other studies (Kaufman et al., 1991; Snow & Weinstock, 1990).

The major purpose of the study was to examine whether there was significant difference in test performance using free expressive (SPN) versus recognition (MC) format in vocabulary testing. Based on the eight age cohorts, significant discrepancies in SPN and MC scores were found for children and younger adolescents from age 6 to 13 cohorts, but not for older adolescents and young adults (age 14-23 cohort). Moreover, the discrepancy showed a clear increasing trend where it was largest for the youngest cohorts and became more attenuated with increasing age. These results suggested that not only young children, but even adolescents up to the age of 13, may be affected by the extra load imposed by the expressive language component of the free expressive format. Individuals in this age range appeared to have the knowledge of the vocabulary items (as shown by their ability to recognize the answers in the MC version for the items that they failed to score full marks on in the SPN version), but failed to freely express the meanings in the SPN version, leading to a discrepancy between the two scores. These results provided evidence to support the idea that while the conventional SPN format in vocabulary testing taps on both the vocabulary store and expressive language performance of individuals (Nation, 1990), the MC format provides an alternative in which vocabulary store might be tested relatively independently from expressive language performance. Given that expressive language ability develops with age, our data seemed to suggest that the MC format might be useful for young children and adolescents whose expressive language ability is not yet fully developed. In this regard, individuals with expressive language problems might also benefit from this form of testing.

Results from this study also bear on another common use of the vocabulary test as an estimate

of IQ. Our results showed that while the SPN format of the vocabulary test had only a weak but significant correlation with TONI-III IQ for the young adult cohort (age 16–23), the MC format showed more substantial correlations in the younger cohorts from age 9 and above. This level of association showed that our test, particularly the MC format, is comparable to results reported on the association between vocabulary test and nonverbal IQ scores ( $r = .35$  to  $r = .57$ ; Eissenthal & Harford, 1971; Hayes, 1999; Gray, Plante, Vance, & Henrichsen, 1999). It should be noted that the use of a nonverbal intelligence test may have underestimated the association between the vocabulary test and IQ as a result of the verbal versus nonverbal difference. It would be informative for future studies to further examine the predictive power of the vocabulary test on verbal and full-scale intelligence scores. Our results showed that compared with the conventional SPN format, the MC format was more strongly correlated with TONI-III IQ. This also provided further evidence to suggest the use of the MC over the SPN format as an estimate of intellectual functioning.

Results on the shorter versions of the vocabulary test also yielded encouraging results. Performance of children and adolescents on the two shorter forms (both SPN and MC versions) showed that they were highly and significantly correlated with both the full version and with each other, with comparable means and homogenous variances and satisfactory internal consistencies. The MC versions of the shorter forms were also found to be significantly correlated with TONI-III IQ, suggesting its possible use for estimating intellectual functioning particularly for young adults. These results suggested that if items were selected carefully, it is possible to construct a sensitive vocabulary test with less than 20 items. This will significantly reduce the administration time of a vocabulary test and is thus of practical clinical utility when clinicians were often under tight time pressure, and patients were often drained from lengthy assessments.

Based on the findings from this study that children from age 6 to 13 were found to have discrepancies between the SPN and MC scores on the vocabulary test, we recommend that clinicians explore the use of the multiple-choice format of vocabulary testing in assessing young children and adolescents. In spite of this potentially important finding, the present study is limited by its small sample size, the limited age range of the participants, and the fact that participants were all school-bound individuals. Further validation of the test with a larger normative sample is necessary

to establish confidence, especially in using the MC format and the shorter versions of the test. Extension of the age range to include adults and elderly individuals, and with different educational levels, is also a necessary step in order for norms to be established for the test to be used clinically on the local population. Finally, further studies to explore the use of the multiple-choice format as an estimate of IQ for individuals with expressive language problems associated with developmental disorders such as autism spectrum disorders, mental retardation, specific language impairments, or as a result of brain injuries such as temporal lobe and frontal lobe damage, should also be carried out.

Original manuscript received 23 October 2007

Revised manuscript accepted 11 December 2007

First published online 14 March 2008

## REFERENCES

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: McMillan.
- Banich, M. T. (1998). *Neuropsychology: The neural bases of mental function*. New York: Houghton Mifflin.
- Bayles, K. A., Tomoeda, C. K., & Boone, D. R. (1985). A view of age-related changes in language function. *Developmental Neuropsychology, 1*, 231–264.
- Brooks, D. N., & Aughten, M. E. (1979). Psychological consequences of blunt head injury. *International Rehabilitation Medicine, 1*, 160–165.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1992). *Test of Nonverbal Intelligence: A language-free measure of cognitive ability* (3rd ed). Austin, TX: PRO-ED.
- Butters, N. (1984). The clinical aspects of memory disorders: Contributions from experimental studies of amnesia and dementia. *Journal of Clinical Neuropsychology, 6*, 17–36.
- Butters, N., Wolfe, J., Granholm, E., & Martone, M. (1986). An assessment of verbal recall, recognition and fluency abilities in patients with Huntington's disease. *Cortex, 22*, 11–32.
- Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak, L. S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition and procedural memory. *Neuropsychologia, 23*, 729–743.
- Chan, A., Shum, D., & Cheung, R. (2003). Recent development of cognitive and neuropsychological assessment in Asian countries. *Psychological Assessment, 15*, 257–267.
- Chang, S. S. Y., Tang, C. S.-K., & Chan-ho, A. M. W. (1995). Developing a Cantonese version of the vocabulary subtest of the WAIS-R for adolescents in Hong Kong. *Bulletin of the Hong Kong Psychological Society, 34/35*, 68–82.
- Chen, Y. H., & Chen, X. Y. (2002). *Wechsler Adult Intelligence Scale-III (Chinese)*. Taipei, Taiwan: Chinese Behavioral Science Corporation. (In Chinese)
- Cihai Editorial Board. (1979). *Cihai*. Shanghai, China: Shanghai Lexicographical Publishing House. (In Chinese)

- Dai, X. Y., Gong, Y. X., & Zhong, L. P. (1990). Factor analysis of the mainland Chinese version of the Wechsler Adult Intelligence Scale. *Psychological Assessment, 2*, 31–34.
- Damasio, H. C. (1991). Neuroanatomical correlates of the aphasias. In M. T. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 45–70). New York: Academic Press.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—Third Edition*. Circle Pines, MN: American Guidance Service.
- Eisenthal, S., & Harford, T. (1971). Correlation between the Raven Progressive Matrices scale and the Shipley Institute of Living Scale. *Journal of Clinical Psychology, 27*, 213–214.
- Gardner, M. F. (1990). *Expressive One-Word Picture Vocabulary Test (Revised)*. Novato, CA: Academic Therapy Publications.
- Gong, Y. X. (1989). *Manual for the Wechsler Adult Intelligence Scale: Revised in China*. Changsha, Hunan, China: Hunan Medical University. (In Chinese)
- Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30*, 196–206.
- Hayes, S. C. (1999). Comparison of the Kaufman Brief Intelligence Test and the Matrix Analogies Test—Short Form in an adolescent forensic population. *Psychological Assessment, 11*, 108–110.
- Hong Kong Education Department. (1999a). *List of primary school textbooks and learning materials recommended by the Textbook Committee*. Hong Kong: Author.
- Hong Kong Education Department. (1999b). *List of textbooks recommended by the Textbooks Committee in secondary schools*. Hong Kong: Author.
- Irving, G. (1971). Psychometric assessment in a geriatric unit. In G. Stocker (Ed.), *Assessment in cerebrovascular insufficiency*. Stuttgart, Germany: George Thieme Verlag.
- Iverson, G. L. (2001). Interpreting change on the WASI-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology, 16*, 183–191.
- Joy, S., Fein, D., Kaplan, E., & Freedman, M. (1999). Information multiple choice among healthy older adults: Characteristics, correlates, and clinical implications. *Clinical Neuropsychologist, 13*, 48–53.
- Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *Wechsler Adult Intelligence Scales—Revised as a Neuropsychological Instrument (WAIS-R NI)*. San Antonio, TX: The Psychological Corporation.
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition*. Bloomington, MN: Pearson Assessment.
- Kaufman, A. S., McLean, J., & Reynolds, C. (1991). Analysis of WAIS-R factor patterns by sex and race. *Journal of Clinical Psychology, 47*, 548–557.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.
- Massman, P. D., Delis, D. C., Butters, N., Levin, B. E., & Salmon, D. P. (1990). Are all subcortical dementias alike? Verbal learning and memory in Parkinson's and Huntington's disease patients. *Journal of Clinical and Experimental Neuropsychology, 12*, 729–744.
- McCaffrey, R. H., Duff, K., & Westervelt, H. J. (2000). *Practitioner's guide to evaluating change with intellectual assessment instruments*. New York: Kluwer Academic/Plenum Press.
- McFie, J. (1975). *Assessment of organic intellectual impairment*. London: Academic Press.
- Melvold, J. L., Au, R., Obler, L. K., & Albert, M. L. (1994). Language during aging and dementia. In M. L. Albert & J. E. Knoefel (Eds.), *Clinical neurology of aging* (2nd ed.). New York: Oxford University Press.
- Nagaratnam, N., & Gilhortra, J. S. (1988). Acute mixed transcortical aphasia following an infarction in the left putamen. *Aphasiology, 12*, 489–493.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.
- Novoa, O. P., & Ardila, A. (1987). Linguistic abilities in patients with prefrontal damage. *Brain and Language, 30*, 206–225.
- Obler, L. K., & Albert, M. L. (1985). Language skills across adulthood. In J. Birren & K. W. Schaie (Eds.), *The psychology of aging*. New York: Van Nostrand Reinhold.
- Raven, J. C. (1982). *Revised manual for Raven's Progressive Matrices and Vocabulary Scales*. Windsor, UK: NFER Nelson.
- Rey, G. J., Levin, B. E., Rodas, R., Bowen, B. C., & Nedd, K. (1994). A longitudinal examination of crossed aphasia. *Archives of Neurology, 51*, 95–100.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside.
- Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego, CA: Jerome M. Sattler.
- Schum, R. L., & Sivan, A. B. (1997). Verbal abilities in healthy elderly adults. *Applied Neuropsychology, 4*, 130–134.
- Shipley, W. C. (1986). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.
- Snow, W. G., & Weinstock, J. (1990). Sex differences among non-brain-damaged adults on the Wechsler Adult Intelligence Scales: A review of the literature. *Journal of Clinical and Experimental Neuropsychology, 12*, 873–886.
- Sullivan, E. V., Sagar, H. J., Gabrieli, J. D. E., Corkin, S., & Growdon, J. H. (1989). Different cognitive profiles on standard behavioral tests in Parkinson's disease and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 11*, 799–820.
- Tamayo, J. M. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and Psychological Measurement, 47*, 893–902.
- Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M*. Boston, MA: Houghton Mifflin.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale* (4th ed.). Chicago, IL: Riverside.
- Ullstadius, E., Gustafsson, J.-E., & Carlstedt, B. (2002). Influence of general and crystallized intelligence on vocabulary test performance. *European Journal of Psychological Assessment, 18*, 78–84.

- Vernon, P. E. (1979). *Intelligence: Heredity and environment*. San Francisco: Freeman.
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore, MD: Williams & Wilkins.
- Wechsler, D. (1974). *Manual for Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corporation.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Manual of the Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D., Kaplan, E., Delis, D., Fein, D., Maerlender, A., Morris, R., et al. (2004). *Wechsler Intelligence Scale for Children—Fourth Edition Integrated (WISC-IV Integrated)*. San Antonio, TX: The Psychological Corporation.
- Zillmer, E. A., Waechter, C., Harris, B., Kahn, F., & Fowler, P. C. (1992). The effects of unilateral and multifocal lesions on the WAIS-R: A factor analytic study of stroke patients. *Archives of Clinical Neuropsychology*, 7, 29–40.