

What Does the WISC-IV Measure? Validation of the Scoring and CHC-based Interpretative Approaches

Hsin-Yi Chen

Department of Special Education,
National Taiwan Normal University
Professor

Timothy Z. Keith

Department of Educational Psychology,
The University of Texas at Austin, TX, U.S.A
Professor

Yung-Hwa Chen

The Chinese Behavioral Science Coporation
Professor

Ben-Sheng Chang

Department of Psychology,
Soochow University
Associate Professor

Abstract

The validity of WISC-IV current four-factor scoring structure and the Cattell-Horn-Carroll (CHC) theory-based models of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) were investigated via the application of higher-order confirmatory factor analyses of scores from the Taiwan WISC-IV standardized sample ($n = 968$). Results reveal that the WISC-IV measures the same construct across ages, the resulting interpretation could be applied to children with various age levels. Both the four-factor structure and CHC-based model were supported. Variance explained was similar across models. The general factor accounted for 2/3 of common variance. First order factors, in total, contributed an additional 1/3 of common variance. The WISC-IV measures crystallized ability (Gc), visual processing (Gv), fluid reasoning (Gf), short-term and working memory (Gsm), and processing speed (Gs). In particular, either separating Gf and Gv, or combining them as the Perceptual Reasoning Index (PRI) provides meaningful explanation. Arithmetic showed significant and split loadings. For children in Taiwan, Arithmetic appears a reflection of Gsm/Gf and Gc.

Keywords: CHC theory, higher order CFA, WISC-IV

Introduction

After 10 years of research on the third edition of the Wechsler Intelligence Scale for Children (WISC-III), one of the primary goals for the recently published fourth edition of this test (WISC-IV) (Wechsler, 2003a, 2007a) was to update its theoretical foundations. New subtests were incorporated to improve measurement of fluid reasoning, working memory, and processing speed (Wechsler, 2003b).

Contemporary science in intelligence generally agrees upon a hierarchical model of cognitive abilities. General intelligence (*g*) tends to emerge whenever a sufficient number of cognitively complex variables are analyzed (Carroll, 1993). Among empirical cognitive theories, the Cattell-Horn-Carroll theory (CHC theory) (Carroll, 1993, 2005) is widely considered as a comprehensive and suitable framework for exploring the nature of cognitive instruments (Flanagan, McGrew, & Ortiz, 2000; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Keith, Kranzler, & Flanagan, 2001; Keith & Witta, 1997; Kranzler & Keith, 1999; Phelps, McGrew, Knopik, & Ford, 2005; Reynolds, Keith, Fine, Fisher, & Low, 2007; Roid, 2003; Woodcock, McGrew, & Mather, 2001). Briefly, CHC model locates cognitive abilities into three structural levels. On top of the CHC hierarchy is the *g*. In the middle level are 10 broad abilities, the broad abilities are considered too broad to be represented by any single measure, thus there are over 70 narrow abilities on the ground level. The currently identified broad abilities are crystallized intelligence (*Gc*), fluid intelligence (*Gf*), quantitative reasoning (*Gq*), short-term memory (*Gsm*), long-term retrieval (*Glr*), visual processing (*Gv*), auditory processing (*Ga*), processing speed (*Gs*), reading and writing ability (*Grw*), and decision/reaction time/speed (*Gt*). Since this model accommodated both theoretical cognitive constructs and empirical findings, no single measurement nowadays covers all CHC abilities, and extension of the construct is still an ongoing action (McGrew, 1997, 2005; McGrew & Flanagan, 1998).

In a cross-cultural analysis of the WISC-III, data from several countries, including Taiwan, demonstrated firm support for the four-factor scoring structure, namely, Verbal Comprehension (VCI), Perceptual Organization (POI), Freedom from Distractibility (FDI), and Processing Speed (PSI), as proposed by the publisher (Georgas, van de Vijver, Weiss, & Saklofske, 2003). In the WISC-IV, POI is updated by the concept of Perceptual Reasoning (PRI), and FDI is renamed to be Working Memory (WMI). The WISC-IV four-factor structure has been supported as a fitting model

(Wechsler, 2003b, 2007b). Recently, higher order confirmatory factor analyses on the American WISC-IV norming sample by Keith et al. (2006) suggested that using five CHC broad abilities (Gc, Gv, Gf, Gsm, and Gs) provides a better structure than does the four-factor scoring solution. In addition, several WISC-IV subtests, such as Arithmetic, Similarities, Picture Concepts, Matrix Reasoning, Block Design, Picture Completion, Coding, and Symbol Search, have been suggested in the literature as measuring multiple abilities and could show possible cross loadings in factor analysis. In particular, Arithmetic may provide a mixed measure of fluid and quantitative reasoning, quantitative knowledge, working and short-term memory, verbal comprehension, and speed (Flanagan et al., 2000; Flanagan & Kaufman, 2004; Keith et al., 2006; Phelps et al., 2005). Indeed, one of the WISC-IV revision goals was to increase the working memory load of this subtest (Wechsler, 2003b, p. 8). Table 1 summarizes these hypothesized cross-loadings. All of these cross-loadings need to be evaluated from a cross-culture perspective.

Table 1 Hypothesized and Actual Cattell-Horn-Carroll Broad Ability Classifications of the WISC-IV Subtests Based on a Population of Taiwanese Children

| Subtest | Primary | | Secondary | |
|--------------------------|--------------|-----------|--------------|--------|
| | Hypothesized | Actual | Hypothesized | Actual |
| Similarities | Gc | Gc | Gf ? | Gf |
| Vocabulary | Gc | Gc | | |
| Comprehension | Gc | Gc | | |
| Information | Gc | Gc | | |
| Block Design | Gv | Gv | Gf ? | |
| Picture Concepts | Gf | Gf | Gc ? | |
| Matrix Reasoning | Gf | Gf | Gv ? | |
| Picture Completion | Gv | Gv | Gc ? | |
| Digit Span | Gsm | Gsm | | |
| Letter-Number Sequencing | Gsm | Gsm | | |
| Arithmetic | Gf / Gsm | Gsm or Gf | Gc ? | Gc |
| Coding | Gs | Gs | Gsm ? | |
| Symbol Search | Gs | Gs | Gv ? | Gv |
| Cancellation | Gs | Gs | | |

Note. Gc=crystallized intelligence; Gv=visual processing; Gf=fluid intelligence; Gsm=short term memory; Gs=processing speed

The Taiwan version of the WISC-IV was recently developed (Wechsler, 2007a, 2007b). It is worth investigating the psychological structures for the Taiwan children population. A comparison of current results to the U.S. findings (Keith et al., 2006), would also improve understanding of the commonalities of cognitive processes across cultures.

The purpose of this study was threefold. First, we investigated the constructs underlying the

Taiwan WISC-IV by comparing a current four-factor model to a CHC theory-based model. Since our goal was to confirm existing models, we decided to take a confirmatory approach for factor analysis, instead of an exploratory manner. Using confirmatory factor analysis for investigating construct validity is widely recognized and applied in the academic field (e.g., Hou, 2009; Wang, 1998). Second, we tested and verified abilities measured by subtests and possible cross-loadings. Finally, we used a Schmid Leiman-type orthogonalization procedure (cf. Schmid & Leiman, 1957; Watkins, 2006) to investigate the factor structure via higher-order CFA. We compared across models the influence of the higher order general factor and residualized effects of lower order factors.

Method

Participants

We analyzed the Taiwan WISC-IV standardization responses from 968 children (males $n = 485$; females $n = 483$). This nationally representative sample was divided into 11 age groups from ages 6 to 16, with 88 children in each age group. This sample was carefully selected to match the 2007 Taiwan Census on region, gender, and parent educational level. The mean age was 11.49, with a standard deviation of 3.18; the average Full-Scaled IQ (FSIQ) was 100 ($SD = 15$). A detailed description of this sample is provided in the Taiwan WISC-IV technical manual (Wechsler, 2007b).

Instrumentation

The Taiwan version of the WISC-IV (Wechsler, 2007a, 2007b) has 10 core subtests and 4 supplemental subtests. The 10 core subtests are: Similarities (SIM), Vocabulary (VOC), Comprehension (COM), Block Design (BLD), Picture Concepts (PCn), Matrix Reasoning (MR), Digit Span (DS), Letter-Number Sequencing (LNS), Coding (CD), and Symbol Search (SYS). The four supplemental subtests are Information (INF), Picture Completion (PIC), Arithmetic (ARI), and Cancellation (CA). Contents of most test items are identical to those on the American WISC-IV. Revisions were made on most verbal subtests to accommodate cultural differences (Wechsler, 2007b, p. 50). All composites and subtests demonstrated good reliabilities, with average internal reliability estimates ranging from .85 to .96 for composites, and .74 to .91 for core subtests.

Analysis of the data

Tests for the higher order confirmatory factor structure were based on analysis of covariance

structure models using LISREL 8.8 (Jöreskog & Sörbom, 2006). Following the main procedures of Keith and his colleagues (Keith, 2005; Keith et al., 2006; Keith & Witta, 1997), equivalence of covariance matrices across age bands first tested whether the WISC-IV measured the same constructs across ages.

Both the current four-factor scoring model and the CHC theory-based model with hypothesized cross loadings were then tested individually. The initial four-factor structure is the one reported in the WISC-IV manual (Wechsler, 2007a). For the 14-subtest version, this model specifies four verbal comprehension subtests (Similarities, Vocabulary, Comprehension, Information) on the first factor, four perceptual reasoning subtests (Block Design, Picture Concepts, Matrix Reasoning, Picture Completion) on the second factor, three working memory subtests (Digit Span, Letter-Number Sequencing, Arithmetic) on the third factor, and three processing speed subtests (Coding, Symbol Search, Cancellation) on the fourth factor. This model was defined as the initial four-factor model (model A1) in our analyses.

We chose Keith's (Keith et al., 2006, Figure 3) initial CHC model as the starting CHC model (model B1). It specified four subtests (Similarities, Vocabulary, Comprehension, Information) on the Gc factor, two subtests (Block Design, Picture Completion) on the Gv factor, three subtests (Picture Concepts, Matrix Reasoning, Arithmetic) on the Gf factor, two subtests (Digit Span, Letter-Number Sequencing) on the Gsm factor, and three processing speed subtests (Coding, Symbol Search, Cancellation) on the Gs factor. In comparison with the WISC-IV four-factor construct, this CHC model split the four perceptual reasoning subtests onto separate tests of visual processing and fluid reasoning factors and placed the Arithmetic subtest on the fluid reasoning factor (Flanagan & Kaufman, 2004; Keith et al., 2006).

In the testing process, hypothetical split loadings of the following subtests were examined and verified separately: Arithmetic, Similarities, Block Design, Picture Concepts, Matrix Reasoning, Picture Completion, Coding, and Symbol Search. To detect underlying structure and possible cross-loadings as precisely as possible, we only deleted statistically non-significant factor loadings.

We also used a calibration-validation approach. Roughly 70% of the standardization sample ($n = 668$) was randomly selected as the calibration sample for hypotheses testing. The remaining 30% of the cases ($n = 300$) validated the results through cross-validation. Once a best-fitting solution from each of the WISC-IV four-factor models and the CHC-based models was calibrated and validated, the final parameter estimates and g-loadings were reported based on the entire sample ($n = 968$). We then applied a Schmid Leiman-type orthogonalization procedure to investigate the sources of variance explained and g loadings in each model.

For all the higher order confirmatory factor analyses, maximum likelihood was the estimation method because of its robustness and sensitivity to incorrectly specified models (Hu & Bentler, 1998). Criteria were evaluated jointly to assess overall model fit (Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988). These included weighted least squares χ^2 , the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the normed fit index (NFI), the non-normed fit index (NNFI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the Akaike Information Criterion (AIC). A value of .90 served as the rule-of thumb lower limit of acceptable fit for all indices ranging from zero to 1, with 1 indicating a perfect fit (Hoyle & Panter, 1995; Kline, 2005). An RMSEA of less than .06 and a SRMR of less than .08 corresponded to a “good” fit (Hu & Bentler, 1999; McDonald & Ho, 2002). For comparisons of competing models, the chi square difference ($\Delta\chi^2$) tested nested models (Loehlin, 2004). The AIC was used to test non-nested rival models (Kaplan, 2000), with smaller AIC values indicating a better fit.

Results

Does the Taiwan WISC-IV measure the same construct across ages?

In a multi-sample CFA model, we first constrained the variance/covariance matrices to be equal across four age groups (ages 6-7, 8-10, 11-13, and 14-16). This constrained model fit the data well (RMSEA = .046; SRMR = .077; NNFI = .99). These findings suggested that regardless of age level, children showed fairly invariant WISC-IV subtest correlation patterns. Because any factor structure is derived from these variance/covariance matrices, these findings suggest that any factor structure we may derive should be considered applicable to children of all ages. Subsequent analyses thus combined data across ages.

Does the Taiwan WISC-IV measure the intended four-factor structure?

As indicated by all goodness-of-fit indexes reported for the initial four-factor model (model A1) in Table 2, all fit values were within the ideal range. It appears that the scoring model fit Taiwan data well.

Since the Arithmetic subtest is known for measuring multiple abilities, tapping both verbal and memory domains, we tested a second model in which Arithmetic was allowed to load on both the verbal comprehension and working memory factors (model A2). Model A2 actually fit the data

Table 2 Comparison of Fit of Models Testing Hypotheses about the WISC-IV Structure

| model | χ^2 | df | χ^2/df | $\Delta\chi^2$ | Δdf | GFI | AGFI | NFI | NNFI | CFI | RMSEA | SRMR | AIC |
|--|----------|----|-------------|----------------|----------------|-----|------|-----|------|-----|-------|------|--------|
| WISC-IV four-factor models (n = 668) | | | | | | | | | | | | | |
| A1. Initial four-factor model | 165.71 | 73 | 2.270 | --- | --- | .97 | .95 | .98 | .99 | .99 | .044 | .030 | 229.71 |
| A2. Arithmetic on WMI + VCI | 143.35 | 72 | 1.991 | 22.36** | 1 ^a | .97 | .96 | .99 | .99 | .99 | .039 | .028 | 209.35 |
| CHC models (n = 668) | | | | | | | | | | | | | |
| B1. Arithmetic(ARI) on Gf | 161.36 | 72 | 2.241 | --- | --- | .97 | .95 | .98 | .99 | .99 | .043 | .031 | 227.36 |
| B2. Arithmetic on Gsm | 180.12 | 72 | 2.502 | --- | --- | .96 | .95 | .98 | .99 | .99 | .047 | .032 | 246.12 |
| B3. Arithmetic on Gc | 245.15 | 72 | 3.405 | --- | --- | .95 | .93 | .98 | .98 | .98 | .060 | .041 | 311.15 |
| B4. Arithmetic on a separate factor | 164.40 | 71 | 2.315 | --- | --- | .97 | .95 | .98 | .99 | .99 | .044 | .031 | 232.40 |
| B5. Arithmetic on Gf + Gsm + Gc + Gs | 141.07 | 69 | 2.044 | 20.29** | 3 ^b | .97 | .96 | .99 | .99 | .99 | .040 | .029 | 213.07 |
| B6. Arithmetic on Gsm+Gc | 150.53 | 71 | 2.120 | --- | --- | .97 | .95 | .98 | .99 | .99 | .043 | .030 | 218.53 |
| C1. Similarities(SIM) on Gc + Gf | 137.97 | 70 | 1.971 | 12.56** | 1 ^c | .97 | .96 | .99 | .99 | .99 | .038 | .029 | 207.97 |
| C2. Block Design on Gv + Gf | 148.98 | 70 | 2.128 | 1.55 | 1 ^c | .97 | .95 | .98 | .99 | .99 | .041 | .029 | 218.98 |
| C3. Picture Concepts on Gf + Gc | 150.55 | 70 | 2.151 | --- | --- | .97 | .95 | .98 | .99 | .99 | .042 | .029 | 220.55 |
| C4. Matrix Reasoning on Gf + Gv | 145.35 | 70 | 2.076 | 5.18* | 1 ^c | .97 | .95 | .99 | .99 | .99 | .040 | .029 | 215.35 |
| C5. Picture Completion on Gv + Gc | 149.62 | 70 | 2.137 | .91 | 1 ^c | .97 | .95 | .98 | .99 | .99 | .041 | .029 | 219.62 |
| C6. Coding on Gs + Gsm | 150.39 | 70 | 2.148 | 0.14 | 1 ^c | .97 | .95 | .98 | .99 | .99 | .041 | .029 | 220.39 |
| C7. Symbol Search(SYS) on Gs + Gv | 142.84 | 70 | 2.041 | 7.69** | 1 ^c | .97 | .96 | .99 | .99 | .99 | .039 | .029 | 212.84 |
| Cross-Validation in combination (n = 300) | | | | | | | | | | | | | |
| D1. CHC: ARI/SIM/SYS with cross loadings | 108.85 | 69 | 1.578 | --- | --- | .95 | .92 | .97 | .99 | .99 | .044 | .036 | 180.85 |
| D2. Combining Gf Gv (based on D1) | 110.17 | 70 | 1.574 | --- | --- | .95 | .92 | .97 | .99 | .99 | .044 | .037 | 180.17 |
| Final model (N = 968) | | | | | | | | | | | | | |
| A2.WISC-IV: Arithmetic on WMI+VCI | 189.32 | 72 | 2.629 | --- | --- | .97 | .96 | .99 | .99 | .99 | .041 | .027 | 255.32 |
| D1. CHC: ARI/SIM/SYS with cross loadings | 166.21 | 69 | 2.409 | --- | --- | .98 | .96 | .99 | .99 | .99 | .038 | .026 | 238.21 |
| D1a. Combining Gf and Gv (based on D1) | 158.91 | 70 | 2.270 | --- | --- | .98 | .97 | .99 | .99 | .99 | .036 | .024 | 228.91 |
| D1b. Deleting small loadings (based on D1a) | 189.32 | 72 | 2.629 | --- | --- | .97 | .96 | .99 | .99 | .99 | .041 | .027 | 255.32 |

^a Compared to the previous model; ^b Compared to Model B1; ^c Compared to Model B6

* $p < .05$. ** $p < .01$.

better than did model A1. The factor loadings of Arithmetic on verbal comprehension and working memory factors were .31 and .52 respectively. The χ^2 difference between A2 and A1 was significant, suggesting that, after allowing for cross-loadings, A2 had a better fit.

Does the Taiwan WISC-IV measure CHC abilities?

As indicated by the starting CHC model B1 in Table 2, this CHC model provided a good fit to the data. Both models A1 and B1 had similar goodness-of-fit, suggesting that both provided reasonable frameworks. We used model B1 as the initial CHC model and proceeded with the following subtest examinations.

1. What does the Arithmetic subtest measure?

Even though results revealed that model B1 (with Arithmetic loaded only on the Gf factor) had a good fit, to better understand the mixed nature of this subtest, several alternative cross-loadings were explored.

We first tested models in which Arithmetic was allowed to load on only one factor. Although all had acceptable fit, model B1, with Arithmetic loading on Gf, fit the data better than models allowing a single loading of Arithmetic on Gsm (model B2), Gc (model B3), or a separate first order factor (model B4). Fit values of model B4 were excellent, but the g loading of this factor was 1.38. This was larger than the g loading of Gf (.92). Given this improbable value, we did not consider load Arithmetic on a separate factor in future models.

We tested other models with cross-loadings for Arithmetic. In the literature, Arithmetic is said to relate to Gf, Gsm, Gc, and Gs factors. When Arithmetic was allowed to cross load on all of these four factors (model B5), this model explained the data quite well. The χ^2 difference between models B5 and B1 was significant, suggesting that the Arithmetic subtest truly is a mixed measure. The loadings of Arithmetic on the Gf, Gsm, Gc, and Gs factors were .11, .37, .27, and .11, respectively. Interestingly, the Gf factor loading was small and statistically not significant. Cross-validation analysis showed that loadings of Arithmetic on Gf and Gs were also not significant. Consequently, in model B6, Arithmetic was loaded only on Gsm (.51) and Gc (.33). The g loadings for Gc, Gv, Gf, Gsm, and Gs were .80, .94, .96, .85, and .65 respectively. All values were considered reasonable theoretically and were solidly cross-validated.

Taken together, these results suggest that when constrained to load on only a single factor, Arithmetic loaded best on the Gf factor. However, when allowed to load on multiple factors, Arithmetic primarily loaded on Gsm, with a salient secondary loading on Gc. Fit values for models B1 and B6 were excellent, providing evidence that both were reasonable frameworks. The AIC value

for model B6, however, was slightly lower, suggesting that cross-loading of Arithmetic on Gsm and Gc better explained the underlying constructs measured by this test and provided a comparatively better approach for interpretation. Therefore, we chose model B6 as the base model for cross-loading examination of other selected subtests.

2. What do the other subtests measure?

(1) *Similarities*. The Similarities subtest requires children to describe how two objects or concepts are similar. Compared to other verbal comprehension subtests such as Vocabulary and Information, this subtest appears to require relatively more inductive reasoning. It thus was a reasonable assumption that this subtest should have a minor cross-loading on Gf. As shown in model C1, allowing this subtest to load on both the Gc and Gf factors resulted in a statistically significantly improved fit. The factor loadings on Gc and Gf were .67 and .20 respectively. Both were statistically significant. This pattern was verified in a cross-validation sample which showed loadings on Gc and Gf as significant, at .59 and .22, respectively. Although primarily a measure of crystallized intelligence, Similarities also appears to tap into fluid intelligence.

(2) *Block Design*. In this subtest, children are asked to re-create a design when viewing a model or a picture. It mainly requires visual processing ability (Gv), including perceptions of spatial relations and mental manipulations of visual patterns. It may also involve some Gf abilities (Kaufman, 1994; McGrew & Flanagan, 1996). Cross-loading (model C2) resulted in no statistically significant loading of this subtest on Gf, and no improvement in model fit. Moreover, the construct of Gv seemed to collapse as g loading on Gv became non-significant. Therefore, Block Design appears to be an important visual processing measure.

(3) *Picture Concepts*. In this subtest, children are presented with rows of pictures and must choose one from each row to form a group with a common characteristic. Although mainly requiring inductive ability, it also involves general information or verbal mediation (Flanagan & Kaufman, 2004). Cross-loading of this subtest on Gf and Gc (model C3) did not improve model fit. The Gc loading was not significant. Therefore, Picture Concepts appears to measure fluid reasoning only, with no cross loading on crystallized intelligence.

(4) *Matrix Reasoning*. This subtest is generally considered to be a good Gf measure, requiring manipulating abstractions, rules, generalizations, and logical relationships. Since this task requires visual processing and mental rotation, Matrix Reasoning may have cross loading on Gv in addition to Gf (model C4). Model C4 showed a small but significant improvement in fit. When a cross-loading was allowed, Matrix Reasoning loaded higher on Gv (.48) than it did on Gf (.27). Both the g loading of Gf and the loading of Matrix Reasoning on Gf were not significant, however.

In this model, constructs for Gf and Gv appeared mixed, making identification of a pure fluid intelligence factor and a pure visual processing factor questionable.

To further investigate the relation between Gf and Gv, we examined a first-order model for the correlation between Gf and Gv. Without cross-loadings for any Gf and Gv subtest, the correlation between these two factors was .96 and was cross-validated with a value of .92. Such strong associations suggested that Gf and Gv may not be separable. We further tested their separability by constraining the correlation among Gf and Gv to 1, while constraining their correlations with other factors to be the same. The resulting model fit the data well ($\chi^2 = 143.04$, $df = 70$, AGFI = .96, RMSEA = .04); the fit of this model was not significantly different from a model without the constraints ($\chi^2 = 135.68$, $df = 66$, AGFI = .96, RMSEA = .04, $\Delta\chi^2 = 7.36$, $df = 4$, $p = .12$). Results revealed that either combining Gf and Gv into one factor or separating them could be acceptable options, but combining the two factors might be more supported from a parsimony standpoint.

(5) *Picture Completion*. For this subtest, children are asked to name or point to the important part missing from a picture within a specified time limit. This task may involve visual processing (Gv) and general information (Gc) abilities (Flanagan & Kaufman, 2004). Cross-loading of these abilities in model C5 resulted in no statistically significant model improvement, and the loading on Gc was not significant. Therefore, Picture Completion measures Gv, with no extra loading on Gc.

(6) *Coding*. This subtest requires children to copy symbols paired with simple geometric shapes or numbers within a specified time limit. Besides measuring the processing speed, successful performance on this task may also indicate recall ability (Wechsler et al., 2004). In model C6, we let this subtest load on both Gs and Gsm. The loading on Gsm (-.03) was not significant, and no model improvement was found. Successful performance in Coding therefore mainly required processing speed.

(7) *Symbol Search*. In this subtest, children are required to scan a search group and rapidly indicate whether the target symbol(s) matches any of the symbols in the search group. This task may involve both processing speed (Gs) and the ability to perceive and think with visual stimuli (Gv) (Keith et al., 2006). Model C7 resulted to be a better fitting model. Loadings of this subtest on Gs and Gv were statistically significant, at .64 and .18, respectively. This pattern was cross-validated. Although primarily measuring processing speed, Symbol Search also required visual processing.

3. Validation analysis

Three subtests were found to have significant cross-loadings in the previous calibration analyses: Arithmetic (Gsm and Gc), Similarities (Gc and Gf), and Symbol Search (Gs and Gv). We further cross-validated these findings by incorporating them in the same model (model D1). This

model was tested in the validation subsample ($n = 300$). All cross-loadings remained statistically significant. The g loadings for G_c , G_v , G_f , G_{sm} , and G_s were .76, .85, .98, .91, and .51, respectively. All were reasonable values.

Since a model combining G_f and G_v may fit the data as well or better, we developed model D2 to test this hypothesis. Model D2 fit the data fairly well, with all fit values within the ideal range. The AIC values for D1 and D2 were very close, suggesting that the constructs of G_v and G_f were quite similar for the Chinese children population. Meaningful explanation could be reached by either separating or combining G_v and G_f abilities.

Final model establishment

The final step was to re-estimate the best solution using the total sample of 968 cases. For comparison, we selected one comparatively best approach from each of the previously tested four-factor models and the CHC-based groups. Consequently, model A2 was chosen as the best four-factor model. Model D1 was selected as the best solution for the CHC-based runs.

As shown in Figure 1, the Arithmetic subtest was cross loaded on both working memory and verbal comprehension factors in model A2. Loadings were both statistically and practically meaningful (all loadings were above .30). As indicated by all goodness-of-fit indexes, this model provided an excellent fit to the total sample. For model D1, as revealed in the Figure 2, the loading of Arithmetic on G_{sm} and G_c was .51 and .30. The loading of Similarities on G_c and G_f was .65 and .20. The loading of Symbol Search on G_s and G_v was .63 and .19, respectively. All loadings were statistically significant. Because fit indices for both models approached the ideal, models A2 and D1 both provided meaningful explanations of the data. Comparatively, model D1 had a slightly smaller AIC value, suggesting that this model might have better cross-validation in the future. However, the discrepancy was trivial, both models explain the data well.

The sources of variance explained by WISC-IV four-factor model (A2) and CHC-based model (D1) are presented in Tables 3 and 4. The tables show the variance accounted for by the general factor (g) versus the residualized, unique variance explained by the first-order factors (with g controlled). The strength and relative importance of factor loadings and proportion of variance explained were similar across models. Both models A2 and D1 explained about 52% of the total variance, leaving 48% unique and error variance. Comparatively, the g factor accounted for most of the total (35.9% to 36.1 %) and common (67.8% to 69.2 %) variance. First order factors, in total, contributed an additional 15.2% to 17.0% of total variance (30.8% to 32.2% of common variance).

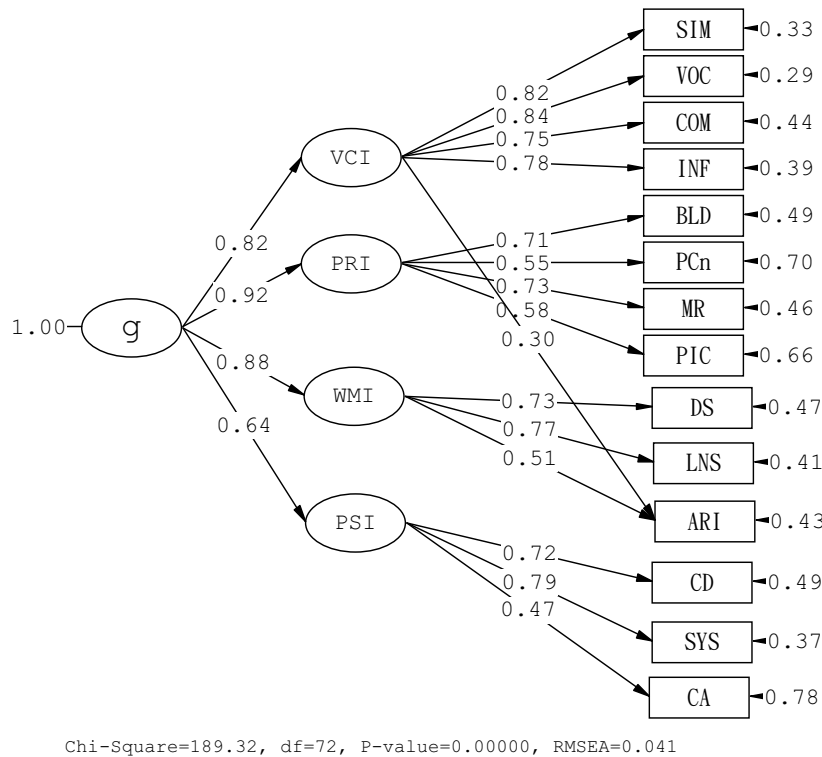


Figure 1 The Final Cross-Validated WISC-IV Four-Factor Structure (Model A2) Using All Data

Current results showed that model A2 and model D1 not only both provided meaningful interpretative frameworks, but also had quite similar accountabilities on explained variance. Interestingly, the four-factor model and CHC-based model are actually quite similar in nature. Taken from model D1, when *Gf* and *Gv* were combined as one factor (model D1a), the loading of Arithmetic on *Gsm* and *Gc* was .53 and .28. The loading of Similarities on *Gc* and a combined *Gf-Gv* was .67 and .19. The loading of Symbol Search on *Gs* and a combined *Gf-Gv* was .64 and .17, respectively. All loadings were statistically significant. When loadings less than .25 (possible with comparatively less practical meaning) were removed, the derived model (model D1b) was exactly the same as model A2 (the previously identified four-factor based solution). Thus, for Taiwanese children, both model A2 and D1 were good-fitting and reasonable options.

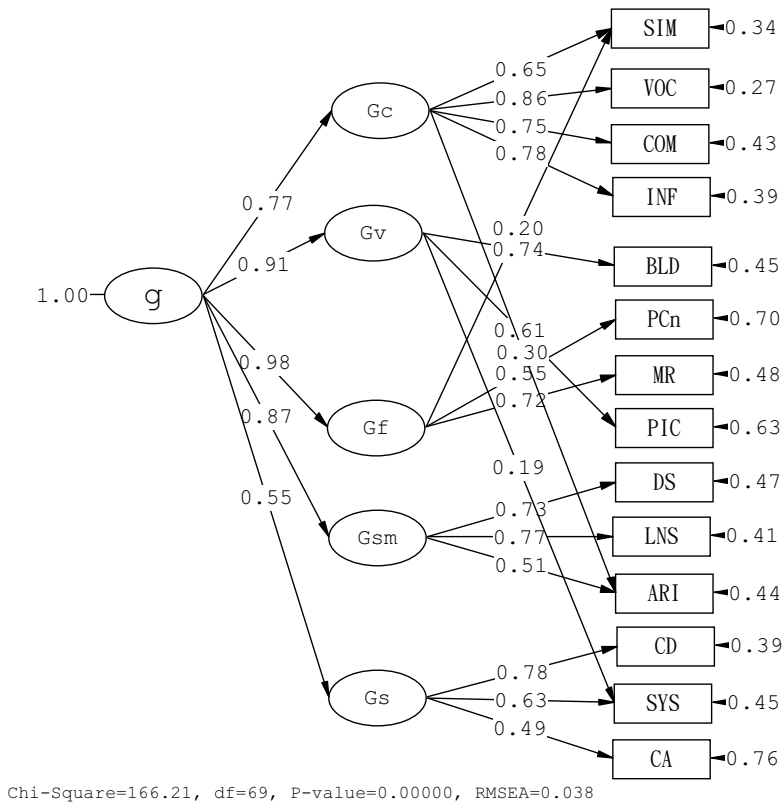


Figure 2 The Final Cross-Validated CHC-Based Structure (Model D1) Using All Data

Discussion

We found strong support for both the WISC-IV four-factor model and the CHC-based model. For Taiwanese children, both models were found similar in nature and explained the WISC-IV data equally well, thus should both be considered valid interpretative approaches. Especially, models separating or combining Gf and Gv provided relatively the same data-fit, suggesting that the Perceptual Reasoning Index or Gf-Gv interpretations are both plausible and have merit psychometrically, we suggest that clinical utility should always be evaluated for making model selection decisions.

Table 3 Sources of Variance for Each Subtest in Four-Factor Model A2 According to an Orthogonal Higher-Order CFA Approach

| Subtest | General | | | VCI | | | PRI | | | WMI | | | PSI | | | h ² | u ² |
|--------------------------|---------|------|-----|------|-----|-----|-----|-----|-----|------|-----|-----|-----|------|---|----------------|----------------|
| | b | var | | b | var | | b | var | | b | var | | b | var | | | |
| Similarities | .67 | 45 | .47 | 22 | - | - | - | - | - | - | - | - | - | - | - | .67 | .33 |
| Vocabulary | .69 | 48 | .48 | 23 | - | - | - | - | - | - | - | - | - | - | - | .71 | .29 |
| Comprehension | .61 | 37 | .43 | 19 | - | - | - | - | - | - | - | - | - | - | - | .56 | .44 |
| Information | .63 | 40 | .45 | 20 | - | - | - | - | - | - | - | - | - | - | - | .60 | .40 |
| Block Design | .66 | 44 | - | - | .28 | 8 | - | - | - | - | - | - | - | - | - | .52 | .48 |
| Picture Concepts | .50 | 25 | - | - | .22 | 5 | - | - | - | - | - | - | - | - | - | .30 | .70 |
| Matrix Reasoning | .67 | 45 | - | - | .29 | 8 | - | - | - | - | - | - | - | - | - | .53 | .47 |
| Picture Completion | .54 | 29 | - | - | .23 | 5 | - | - | - | - | - | - | - | - | - | .34 | .66 |
| Digit Span | .65 | 42 | - | - | - | - | .35 | 12 | - | - | - | - | - | - | - | .54 | .46 |
| Letter-Number Sequencing | .68 | 46 | - | - | - | - | .37 | 13 | - | - | - | - | - | - | - | .59 | .41 |
| Arithmetic | .69 | 48 | .17 | 3 | - | - | .24 | 6 | - | - | - | - | - | - | - | .57 | .43 |
| Coding | .46 | 21 | - | - | - | - | - | - | .55 | 31 | - | - | - | - | - | .52 | .48 |
| Symbol Search | .51 | 26 | - | - | - | - | - | - | .61 | 37 | - | - | - | - | - | .63 | .37 |
| Cancellation | .30 | 9 | - | - | - | - | - | - | .36 | 13 | - | - | - | - | - | .22 | .78 |
| % Total Variance | | 36.1 | | 6.2 | | 1.9 | | 2.2 | | 5.8 | | 2.2 | | 5.8 | | 52.1 | 47.9 |
| % Common Variance | | 69.2 | | 11.9 | | 3.6 | | 4.2 | | 11.1 | | 4.2 | | 11.1 | | | |

Note. b =loading of subtest on factor; var = percent variance explained in the subtest; h² = communality; u² = uniqueness

Table 4 Sources of Variance for Each Subtest in CHC-based Model D1 According to an Orthogonal Higher-Order CFA Approach

| Subtest | General | | | Gc | | | Gv | | | Gf | | | Gsm | | | Gs | | | h ² | u ² |
|--------------------|---------|------|-----|------|-----|-----|-----|-----|-----|-----|-----|------|-----|------|-----|------|-----|------|----------------|----------------|
| | b | var | | b | var | | b | var | | b | var | | b | var | | b | var | | | |
| Similarities | .70 | 49 | .41 | 17 | - | - | .04 | 0.2 | - | - | - | - | - | - | - | - | - | - | .66 | .33 |
| Vocabulary | .66 | 44 | .55 | 30 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | .74 | .26 |
| Comprehension | .58 | 34 | .48 | 23 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | .57 | .43 |
| Information | .60 | 36 | .50 | 25 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | .61 | .39 |
| Block Design | .68 | 46 | - | - | .31 | 9 | - | - | - | - | - | - | - | - | - | - | - | - | .55 | .45 |
| Picture Concepts | .54 | 29 | - | - | - | - | .11 | 1 | - | - | - | - | - | - | - | - | - | - | .30 | .70 |
| Matrix Reasoning | .71 | 50 | - | - | - | - | .14 | 2 | - | - | - | - | - | - | - | - | - | - | .52 | .48 |
| Picture Completion | .55 | 30 | - | - | .25 | 6 | - | - | - | - | - | - | - | - | - | - | - | - | .36 | .64 |
| Digit Span | .64 | 41 | - | - | - | - | - | - | .36 | 13 | - | - | - | - | - | - | - | - | .54 | .46 |
| Letter-Number Seq | .67 | 45 | - | - | - | - | - | - | .38 | 14 | - | - | - | - | - | - | - | - | .59 | .41 |
| Arithmetic | .68 | 46 | .19 | 4 | - | - | - | - | .25 | 6 | - | - | - | - | - | - | - | - | .56 | .44 |
| Coding | .43 | 18 | - | - | - | - | - | - | - | - | - | - | - | - | .65 | 42 | - | - | .60 | .40 |
| Symbol Search | .52 | 27 | - | - | .08 | 1 | - | - | - | - | - | - | - | .53 | 28 | - | - | - | .56 | .44 |
| Cancellation | .27 | 7 | - | - | - | - | - | - | - | - | - | - | - | .41 | 17 | - | - | - | .24 | .76 |
| % Total Variance | | 35.9 | | 7.1 | | 1.1 | | 0.2 | | 2.4 | | 6.2 | | 11.7 | | 52.9 | | 47.1 | | |
| % Common Variance | | 67.8 | | 13.4 | | 2.2 | | 0.4 | | 4.5 | | 11.7 | | | | | | | | |

Note. b =loading of subtest on factor; var = percent variance explained in the subtest; h² = communality; u² = uniqueness

Results of our work indicated that the WISC-IV measured the same constructs across ages. Besides, Arithmetic, Similarities, and Symbol Search subtests produced somewhat mixed measures of cognitive ability. Arithmetic could be considered primarily a measure of fluid intelligence or of short term/working memory. This subtest also measures some verbal comprehension and knowledge. The Similarities subtest mainly measures crystallized intelligence, plus some fluid reasoning. Symbol Search measures not only processing speed, but also some visual-spatial processing. Main and secondary abilities measured by each WISC-IV subtest are summarized in Table 1. It deserves noticing that, the fact of cross-loadings does not mean lacking content validity. First, cognitive abilities are inter-related in nature; Second, based on data from figure 1 and 2, the factor loading of each main ability is clearly higher than the value found for each secondary ability.

The results also indicate that Gf provided an excellent measure of g, as expected. In addition to being robust, the g factor accounted for the greatest amount of the total and common variance. Among the subtests, Vocabulary, Arithmetic, Letter-Number Sequencing, Similarities, Matrix Reasoning, and Block Design had the higher g loadings (larger than .65). The Cancellation subtest had the lowest g loading in both models. It should also be noticed that g accounted for 2/3 of the common variances while all the other factors explained the extra 1/3 of the common variances. They both are meaningful and important elements in explaining intelligence.

Several findings were quite different from the results reported by Keith et al. (2006) for American children. A model separating Gf and Gv fit better for American children. However, this was not the case in this current work. Taiwanese children revealed more strongly correlated fluid reasoning and visual-spatial processing factors than do American children. The correlation between visual-spatial processing and working memory for Chinese children was also higher in Taiwanese children. Since working memory and fluid reasoning are reported to be strongly associated in the literature (de Jong & Das-Smaal, 1995; Fry & Hale, 1996, 2000), it is suspected that when Chinese children are presented with visual-spatial stimuli, the visual-spatial sketchpad in the working memory system (Baddeley, 2003) might be activated somewhat semi-automatically, and thus activating the fluid reasoning engine as well. This can help explain the discrepant observation on the Picture Completion subtest. Keith et al. (2006) reported cross-loadings of Picture Completion on both visual-spatial processing (.42) and crystallized intelligence (.31). We found a loading only on the visual-spatial processing (.61). Therefore, there may be a stronger association between visual processing and non-verbal concept reasoning for Taiwanese children. We speculated that there might be an effect of cultural difference between Chinese and American on writing systems (pictographs vs. alphabet). Further research is needed for to verify such speculation.

For U.S. children, the Arithmetic subtest may be considered primarily as a measure of fluid reasoning (Keith et al., 2006) or quantitative knowledge (Phelps et al., 2005). For Taiwanese children, however, Arithmetic seems more complex. When constrained to load only on one factor, this subtest did load the highest on fluid reasoning. However, when cross loadings were allowed, this subtest showed strong cross loadings on short term/working memory (.51) and crystallized intelligence (.30), with no significant Gf loading detected. This cross-loaded model fit the data better than did the single Gf-loaded model. As Keith et al. (2006) suggested, it is quite likely that for children with different abilities, different factors may affect their performance on Arithmetic. Over the past decades, international studies repeatedly report that children from East Asian nations (i.e., Singapore, Taiwan, South Korea, Hong Kong, and Japan) show higher mathematic performance (Beaton et al., 1996; Mullis, Martin, Gonzales, & Chrostowski, 2004). Various sources of explanations for this performance include racial differences in intelligence (Lynn, 2006), differences in mathematics education, level of sophistication of strategies used in arithmetic problem solving (Fuson & Kwon, 1992), and discrepant cross-generational changes in educational outcomes (Geary et al., 1997). The underlying mechanisms for the arithmetic subtest may be different for Chinese and American children. In addition, both the current study and Keith et al. (2006) found a strong association ($r = .85$) between working memory and fluid reasoning ability. This finding is consistent with contemporary research (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, & Conway, 2005) documenting tremendous amount of shared variances between Gsm and Gf. Thus, mixed results found in this study seem reasonable.

As Prifitera, Weiss, Saklofske, and Rolfhus (2005) suggested, factor analysis is only a tool for informing how best to interpret relationships among subtests and that clinical utility should be considered when selecting factors. Phelps et al. (2005) also recognized that strong fit statistics do not prove which is the correct model, instead, they suggest which models are plausible. In this current study, we found that both models were meaningful ways for interpretation. When children's performances on perceptual reasoning related subtests are inconsistent, professionals are encouraged to check for consistency within Gf and Gv domains. Psychometric properties and linear equating formula (Tellegen & Briggs, 1967, Formula 4) for Gf and Gv composites ($M = 100$, $SD = 15$) are summarized in the Appendix A for references. Flanagan and Kaufman (2004) also provided norms for Gf and Gv composites for U.S. children.

Our results provided solid information for understanding the WISC-IV structure and related cognitive processes across culture. This work analyzed a large and nationally representative sample

of Taiwanese children, and used a cross-validation approach. Nonetheless, some inevitable limitations of the present study are also evident. This study focused primarily on the broad CHC factors, it did not tap the domains of narrow CHC abilities. Joint CFA with other measures could provide a much more precise picture of the abilities measured by the WISC-IV.

In conclusion, Both the WISC-IV four-factor structure and the CHC-based model were supported as meaningful approaches for interpreting WISC-IV performance in the Chinese culture. Our findings improve understanding of the WISC-IV constructs across cultures. Professionals are encouraged to note not only the similarities but also the discrepancies of the underlying cognitive abilities involved of each WISC-IV scores when measuring children across cultures.

Acknowledgements

The authors like to thank all test administrators and children who participated in the Taiwan WISC-IV development. Our gratitude also goes to the Academic Paper Editing Clinic, NTNU.

References

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA'S third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 69-76). New York: Guilford.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.
- de Jong, P. F., & Das-Smaal, E. A. (1995). Attention and intelligence: The validity of the star counting test. *Journal of Educational Psychology*, 81(1), 80-92.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309-331.
- Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: Wiley.
- Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn and Bacon.
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, 7(4), 237-241.
- Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, 54, 1-34.
- Fuson, K. C., & Kwon, Y. (1992). Korean children's single-digit addition and subtraction: Numbers structured by 10. *Journal for Research in Mathematics Education*, 23, 148-165.

- Geary, D. C., Hamson, C. O., Chen, G., Liu, F., Hoard, M. K., & Salthouse, T. A. (1997). Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States. *Psychonomic Bulletin and Review*, 4, 254-263.
- Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, L. G. Weiss, F. J. R. van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 278-313). San Diego, CA: Academic Press.
- Hou, Y. (2009). The technique development of assessing preschool children's creativity in science: An embedded assessment design. *Journal of Research in Education Sciences*, 54(1), 113-141.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.8 statistical program* [Scientific Software]. Chicago, IL: Scientific Software International.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 66-71.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.
- Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581-614). New York: Guilford Press.
- Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review*, 35, 108-127.
- Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson

- Tests of Cognitive Ability (3rd ed.). *School Psychology Review*, 30, 89-119.
- Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly*, 12, 89-107.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.
- Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review*, 28, 117-144.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural analysis* (4th ed.). Hillsdale, NJ: Erlbaum.
- Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis*. Augusta, GA: Washington Summit.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410.
- McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equating analyses. *Psychological Methods*, 7(1), 64-82.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-180). New York: Guilford Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136-181). New York: Guilford Press.
- McGrew, K. S., & Flanagan, D. P. (1996). The Wechsler performance scale debate: Fluid intelligence (Gf) or visual processing (Gv). *Communique*, 24(6), 14-16.
- McGrew, K. S., & Flanagan, D. P. (1998). *The Intelligence Test Desk Reference: Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
- Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery

- investigation. *School Psychology Quarterly*, 20(1), 66-88.
- Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 3-32). San Diego, CA: Academic Press.
- Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman assessment battery for children-second edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly*, 22(4), 511-539.
- Roid, G. H. (2003). *Stanford Binet Intelligence Scale, Fifth Edition: Technical manual*. Itasca, IL: Riverside.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61.
- Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*, 31(5), 499-506.
- Wang, K. (1998). Adolescent self-efficacy scale in unintentional injury: Testing for second order factorial validity and invariance. *Journal of National Taiwan Normal University: Education*, 43(1), 33-48.
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment*, 18(1), 123-125.
- Wechsler, D. (2003a). *Manual for the Wechsler Intelligence Scale for Children-Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children-Fourth Edition technical and interpretive manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2007a). *Manual for the Wechsler Intelligence Scale for Children-Fourth Edition (Taiwan)*. Taipei, Taiwan: The Chinese Behavioral Science Corporation.
- Wechsler, D. (2007b). *Wechsler Intelligence Scale for Children-Fourth Edition technical and interpretive manual (Taiwan)*. Taipei, Taiwan: The Chinese Behavioral Science Corporation.
- Wechsler, D., Kaplan, E., Fein, D., Morris, R., Kramer, J. H., Maerlender, A. et al. (2004). *The Wechsler Intelligence Scale for Children-Fourth Edition Integrated*. San Antonio, TX: The Psychological Corporation.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside.

Appendix A

Table A1 Psychometric Properties and Linear Equating Formula^a for Gf and Gv

| Composite | Elements ^a | Reliability | SEM | Formula |
|-----------|-----------------------|-------------|------|-------------------|
| Gf | MR + PCn | .87 | 5.42 | $(2.98X) + 40.45$ |
| Gv | BLD + PIC | .87 | 5.52 | $(2.94X) + 41.28$ |

^a X= The sum of age-corrected scaled scores of corresponding two subtests

教育科學研究期刊 第五十四卷第三期
2009 年，54 (3)，85-108

魏氏兒童智力量表第四版建構內涵之探討

陳心怡

國立臺灣師範大學特殊教育學系
教授

陳榮華

中國行為科學社
教授

Timothy Z. Keith

美國德州大學奧斯汀校區教育心理學系
教授

張本聖

東吳大學心理學系
副教授

摘要

本研究目的在分析魏氏兒童智力量表第四版 (WISC-IV) 所測量之認知內涵。根據臺灣標準化樣本 ($n = 968$)，研究者以高階驗證性因素分析針對「WISC-IV 現行計分依據之四因素模式」及「Cattell-Horn-Carroll (CHC) 認知理論模式」兩種架構進行系列性效度驗證。主要研究發現為：一、不同年齡組兒童展現一致的 WISC-IV 建構內涵，亦即研究中報告的建構解釋方式適用於 6-16 歲之各年齡層兒童；二、「WISC-IV 四因素模式」及「CHC 認知理論模式」二者均為合理且具意義之詮釋方式，且二架構所解釋到之分數變異量相當一致。其中，高階 g 因素約解釋 2/3 之共同變異量，而其他初階因素則解釋到 1/3 之共同變異量；三、在 CHC 認知理論模式下，WISC-IV 測量之認知能力有結晶智力 (Gc)，視覺處理 (Gv)，流體推理 (Gf)，短期／工作記憶 (Gsm) 及處理速度 (Gs)。值得注意的是，不論將 Gf 與 Gv 依 CHC 理論分開解釋，或容許二者結合成 WISC-IV 四因素內之知覺推理因素 (PRI)，兩種建構在本研究中均獲支持為可行詮釋方式；四、不論採哪一種理論模式，算術分測驗均明顯呈現跨因素負荷量 (split loadings)。算術分測驗測量到多種內涵，其中包括流體推理能力 (Gf)，短期／工作記憶 (Gsm)，以及結晶智力 (Gc)。

關鍵字：CHC 理論、高階驗證性因素分析、魏氏兒童智力量表第四版

通訊作者：陳心怡，E-mail: hsinyi@ntnu.edu.tw

收稿日期：2009/02/03；修正日期：2009/06/04；接受日期：2009/08/04。