Modeling Working Memory Tasks on the Item Level

Dasen Luo, Ph. D.,
Indiana University of Pennsylvania

Guopeng Chen, Ph. D.,
East China Normal University, China

Fanlin Zen, and
East China Normal University, China

Bronwyn Murray
Indiana University of Pennsylvania

Abstract

Item responses to Digit Span and Letter-Number Sequencing were analyzed to develop a better-refined model of the two working memory tasks using the finite mixture (FM) modeling method. Models with ordinal latent traits were found to better account for the independent sources of the variability in the tasks than those with continuous traits, and the discretely distributed factors appeared to represent short-term storage (STS), general attention control (GAC), and the specific control mechanisms initiated by the interfering operations of mental sorting (MS) and backward ordering (BO). When related to the general ability factor (G) defined by the WISC-R verbal and performance scores and the total achievement score, the general working memory factors STS and GAC both seemed to share substantial variances with G, but the roles of specific factors MS and BO were less definitive. These WM factors accounted for the majority of the variability in G, with the multiple correlation between the factor mean scores of the WM factors and that of G above 0.80. Moreover, there seemed to be a discontinuity in the distribution of the ordinal GAC factor, as the two lowest subcategories of GAC were separated from the rest of the overall sample by the virtually empty third lowest subcategory, and the two outlying low subcategories contained the majority (80%) of the cases with mild mental retardation. The theoretical implications of these results were discussed.

Keywords

finite mixture modeling; general attention control; intelligence; mild mental retardation; short-term storage; working memory

---

A model of the actual mechanisms of working memory, particularly when the mechanisms' validated roles in intelligence are concerned, has been a subject of some dispute. Most current working memory models are postulated as a hybrid of mechanisms, including that of a short-term storage and other more intricate mechanisms, such as attention integration, storage with processing, updating, shifting, inhibition, etc (Cowan, 1995; Engle, Tuholski, Laughlin, & Conway, 1999; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Friedman, Miyake,

---

Corley, Young, DeFries, & Hewitt, 2006; Oberauer, 2002; Baddeley & Hitch, 1974). Whereas the model for the mechanism of short-term storage is one that has stood the test of time in cognitive psychology, those for the other working memory mechanisms are not as well defined and the covariances of these mechanisms with intelligence are often difficult to disentangle from one another. The difficulty has to do in large part with the technique typically adopted in modeling working memory tasks and their relations with intelligence. Specifically, the technical hurdles that together add to the difficulty include: (1) the inability to extricate independent sources of variability in working memory tasks; (2) the over-reliance on the correlation-based, projected latent relationship between working memory and intelligence; and (3) the lack of malleability in treating latent traits of working memory and intelligence that are possibly discrete in a heterogeneous population.

Technical Hurdles to Be Resolved

The Inability to Extricate Independent Sources of Variability

With the exception of tasks of strictly short-term storage (STS), working memory tasks are invariably compounds of various mechanisms and mental operations. Because of the interplay between the mechanisms and operations in the same task, it is often difficult to determine the independent sources of variability within a task that may represent separate mechanisms. The vagueness surrounding attention control and executive control mechanisms in various proposed models of working memory to a considerable degree reflects such a difficulty. All full-fledged working memory tasks require additional attention/executive control over the information in the memory span and the additional control is always exercised to carry out specific interfering operations such as backward-ordering, mental sorting, etc. These specific interfering operations are all expected to tax the general attention resource to an extent but they may also each require specific mental executions. In other words, the attention/execution control mechanisms may comprise a multitude of mental resources, some are more general and some are more execution-specific, and these general and specific resources are often not teased apart in the modeling of working memory tasks.

The distinction between the two general mechanisms, STS and the general attention control mechanism, also gives rise to some questions. Colom, Abad, Quiroga, Shih, and Mendoza (2008), for example, found that working memory mechanisms beyond STS at best provided moderate additional predictive power, and questioned the virtue of various proposed working memory mechanisms that transcend STS. The ambiguity about this distinction is also reflected to a degree by the uncertainty about the forward digit span task and the backward digit span task. Although the backward span task often seems to correlate moderately higher with intelligence than its prototypical STS forward span sibling, it is unclear whether the moderately higher correlation truly signifies an independent source of variability in the backward span task (Engle et al., 1999; Rosen & Engle, 1997) or whether it is merely the product of possibly stronger psychometric property of the backward span task due to its wider range of item difficulty. These questions will not be satisfactorily answered without an effective distinction between the independent sources of the variability in working memory tasks.

To date the analysis of working memory mechanisms has been conducted nearly exclusively on the task-level. The task-level analysis, however, seems to be seriously limited in its capacity of effectively differentiating independent sources of variability in working memory, particularly when the to-be-differentiated sources each are to be represented by latent traits. The difficulties are interwoven with the two general approaches commonly adopted to modeling the working memory traits: the correlated-factors approach and the orthogonal-factors approach. The former treats the working memory factors that putatively reflect varied aspects of control (e.g., shifting, updating, inhibition, etc) as correlated predictors for intelligence, and focally tests the unique variances, typically modeled as partial regression

weights, that these factors share with intelligence. This approach tends to relegate the variability shared among all working memory factors to a backseat status and leave the variability unanalyzed. The shared variability is nonetheless non-unitary, comprising of at least two separate sources of variability, namely, STS and general attention control, and the shared sources may be the main contributors to the working memory-intelligence relationship.

The orthogonal-factors approach is in principle more suitable for determining the independent sources of variability in working memory tasks, but has been in practice adopted less often. The approach requires a demanding task selection paradigm similar to that of the multi-trait-multi-method design, with the general attention control mechanism pervading multiple tasks and with each specific control mechanism also to be defined by multiple tasks sharing the same executive operation. In particular, each orthogonal factor needs at least three observed indicators to be identified in traditional factor analysis, and with each observed task indicating three or more latent factors, namely, STS, general attention control, the execution-specific control, and possibly also a content factor, each factor is unlikely to be reliably defined by only three observed indicators. It may nonetheless be rather difficult, if not entirely impractical, for researchers of working memory to design sufficiently more tasks consistent with the strenuous latent structure to meet this need.

Differentiation between various variability sources of working memory can be more readily achieved from the analysis of item responses. On the item-level, the latent traits for working memory mechanisms are each defined by multiple items, and item responses to different tasks putatively measuring general mechanisms (e.g., STS and general attention control) and specific mechanisms (e.g., those related to backward ordering and mental sorting) can be analyzed together to differentiate additional variance sources that can not be distinguished within the same individual tasks. The task selection demand for item-level analysis is considerably less than that for task-level modeling, as the latent traits are defined not by tasks, but by items, which are much easier to be generated within the confines of the intended latent structure. Moreover, unlike latent traits in task-level models that are identified according to the patterns of the between-task correlations, the latent traits on the item-level are determined according to the axiom of local independence for item responses. The axiom states that the latent traits account for all associations, including, but not limited to, linear correlations, among the responses and the items are probabilistically independent of each other given the participant's levels on the latent traits. The micro-level independent variance sources of working memory thus identified can then be related to traits of omnibus abilities, such as that of general intelligence (g) (Spearman, 1904) or scholastic performance, to evaluate their respective relevance. The item-level modeling of latent traits is thus more refined for identifying possible independent sources of variability inherent in working memory tasks than the task-level modeling.

Over-Reliance on Correlation-Based Projected Relationship

Current evidence supporting a very strong bearing of working memory on intelligence has come almost invariably from the modeling of observed correlations/covariances. This type of modeling is purported to identify patterns of observed correlations between working memory and intelligence measures, and the theoretical latent relations between working memory and intelligence, albeit very high at times (0.80 or higher in factorial correlations), are "projected" in that they are only theoretical accounts for the often intricate patterns of correlations among tasks. Moreover, the very high latent correlations are projected from moderate observed correlations (0.4-0.5) (Ackerman, Beier, & Boyle, 2005) through an upward linear correction of the observed correlations to counter the downward influence of error and specific variability in the manifest variables. The very strong projected latent relationship between working memory and intelligence has not been validated through classification or assessment of

individual persons, and questions naturally arise about the justifiable magnitude of the upward correction. The problem is further exacerbated by the possibility that working memory may be nonlinearly related to intelligence, casting further doubts about the accuracy of the upward linear correction. A validation of such a very strong theoretical relationship between working memory and intelligence through classification and assessment of individuals seems to be direly needed, as the claimed strong relationship cannot be taken with good faith without such a validation.

Possibly Discrete Latent Traits

The latent traits representing the working memory mechanisms may not linearly scale with the observed working memory performance. In particular, contrary to the conception that these traits function on the monotonically increasing/decreasing continuum, the functioning of these mechanisms may be better depicted as discrete, ascending/descending clusters with adjacent clusters not necessarily contiguous to each other on an interval/ratio scale. Discrete traits tend to arise when the population is heterogeneous, which is a possibility calling for serious scrutiny (Dolan & van de Maas, 1998; Lubke & Muthén, 2005). There have been indications that a discontinuity between higher and lower ability clusters may exist, especially at the lower end of ability distribution, as suggested by the widely known phenomenon of "The Law of Diminished Returns" (Detterman & Daniel, 1989; Hunt, 1995; Jensen, 2003).

The Finite Mixture Modeling Approach to the Hurdles

Some or most of these hurdles can be surmounted using the finite mixture (FM) modeling, termed also as the latent class modeling, technology (Heinen, 1996; Lazarsfeld & Henry, 1968; McLachlan & Peel, 2000; Muthén, 2002; Vermunt & Magidson, 2002; also see Muthén, 2001a, b for less technical overviews on the general principle of FM modeling). FM modeling was initially developed to model discrete latent variables (latent population heterogeneity) underlying mostly continuous observed variables whereas latent class modeling was previously used to treat discrete latent traits indicated by largely discrete observed variables, but the two modeling technologies have become one and the same in recent years. The technology can also easily accommodate models with continuous latent traits for discrete observed indicators, which are traditionally treated under the rubric of item response theory, and models with a mixture of continuous and discrete latent traits (Meij-de Mei, Kelderman, & van der Flier, 2008; Vermunt, 2001). The technology thus allows one to incorporate into models a versatile collection of latent and observed variables, be they continuous or discrete, and the modeled relations are not limited to be linear. Several features of the modeling technology are particularly pertinent to the present study: (1) the logistic link function relating discrete observed indicators to latent traits; (2) the axiom of local independence; (3) the capability of specifying and testing models with continuous and discrete latent traits; (4) the estimation of posterior probability.

Logistic Link Function

The item-level modeling of working memory tasks involves relating latent traits to discrete observed variables (task items). Because these items are measured on an ordinal scale, they are related to the hypothetical latent traits underlying them in a nonlinear manner, and the commonly adopted linear models for relations between continuous observed and latent variables are obviously inappropriate. In FM modeling, the ordinal manifest indicators can be regressed to their latent predictors through a set of logistic link functions depicted below.

A logistic link function is a function that links a linear model to an observed ordinal item response. For illustration, let the observed response be denoted as Y^t_i for participant i and Item t, and code Y^t_i as 0, 1, …m,…M^t. An example of Y^t_i is a response to a working memory item t for which the perfect recall is scored as 3, i.e., $M^t = 3$. Response Y^t_i follows a multinomial

probability distribution, and the parameters of the distribution are the probabilities of the participant giving any of the possible responses, i.e., $P(Y^t_i = m, m = 0, 1, \ldots, M^t)$. There are four response probabilities for the exemplified item (i.e., for responses 0, 1, 2, and 3), but as one of them is dependent on the other three, a proper constraint is needed for the identification of parameter estimation. Provided that a constraint on the first item-category (e.g., the 0 category) is employed, then the logistic link function that relates the observed response to its latent trait(s) for the unconstrained categories of $m = 1, 2, \ldots, M^t$ (e.g., 1, 2, 3 for the 4-category item) is as follows:

$$
\begin{aligned}
\text{logit}^t_m &= \ln\left[\frac{P(Y^t_i=m \mid X_1=x_1, X_2=x_2\cdots)}{P(Y^t_i=m-1 \mid X_1=x_1, X_2=x_2\cdots)}\right] \\
&= \beta^t_{0m} + \beta^t_1 x_1 + \beta^t_2 x_2 \cdots
\end{aligned}
\tag{1}
$$

$x_1, x_2, \ldots$ in Equation (1) are scores of the latent traits $X_1, X_2, \ldots$, and model parameters β^t_{0m}, $\beta^t_1, \beta^t_2 \ldots$, are regression weights relating the log-transformed odds ratio between two adjacent item-categories, m and $m-1$, to the latent traits. The form of logistic function (logit) is adjacent-category logistic (ref., O'Connell, 2006), for which each item t with M^t+1 (e.g., 4) categories results in M^t (e.g., $4-1=3$) such unconstrained equations and one special equation for the constrained 0 category. Each equation for the same item differs only in the model intercept, i.e., β^t_{0m}, whereas each of the slopes, e.g., $\beta^t_1, \beta^t_2 \ldots$, is invariant across item-categories.

In Equation (1), the slopes $\beta^t_1, \beta^t_2 \ldots$, indicate how the variation in the latent traits Xs affects the logit between the two adjacent categories m and $m - 1$. The role of the intercept, β^t_{0m}, is more complex, because it needs to be interpreted together with the previous logit. For $m = 0$, 1, 2, and 3, four equations with four different intercepts $\beta^t_{0m=0}$ (for category 0), $\beta^t_{0m=1}$, $\beta^t_{0m=2}$, and $\beta^t_{0m=3}$, will be specified, although for the same person i, the $\beta^t_1 x_1 + \beta^t_2 x_2 + \ldots$ part of the logits remains unchanged in all these equations.

To illustrate Equation (1) more explicitly, suppose the logits for the item with $M^t+1=4$ categories have intercepts $\beta^t_{0m=1} = 0.1077$, $\beta^t_{0m=2} = 0.4847$, and $\beta^t_{0m=3} = -0.1455$, and there is only one latent trait in the model with a slope $\beta^t_1 = 2.2161$. For the purpose of parameter identification, the intercept for the category of 0 is set to zero, $\beta^t_{0m=0} = 0$. If participant i has a trait score $X_1 = 0.5$, then the four logits have predicted values of

$$
\begin{aligned}
\text{logit}_1 &= (0.1077 - 0) + 2.2161\,(0.5) = 1.2158, \\
\text{logit}_2 &= (0.4847 - 0.1077) + 2.2161\,(0.5) = 1.4851, \\
\text{logit}_3 &= (-0.1455 - 0.4847) + 2.2161\,(0.5) = 0.4779,
\end{aligned}
$$

and

$$
\text{logit}_0 = (0) + 2.2161\,(0.5) = 1.1081,
$$

respectively.

Each unconstrained logit (e.g., $logit_1 = 1.2158$, $logit_2 = 1.4851$, or $logit_3 = 0.4779$) reflects the amount of change in probability from one item-category to the next. For instance, $logit_1 = 1.2158$ indicates that for those who have the trait score $X_1 = 0.5$, the probability of giving response $m = 1$ is

$$
e^{\text{logit1}} = e^{1.2158} = 3.3728
$$

times the probability of giving response 0. This ratio (the odds ratio) is also provided by the other unconstrained logits. For example, the value of $logit_2 = 1.4851$ reflects the odds ratio between the two response probabilities of the adjacent categories $m = 2$ and $m = 1$, i.e.,

$$e^{logit2} = e^{1.4851} = 4.4152.$$

Moreover, successive logits can also be used to derive odds ratios between non-adjacent categories. For example, to obtain the odds ratio contrasting $m = 2$ to $m = 0$, one simply calculates $e^{logit1} * e^{logit2}$, or $e^{logit1+logit2}$, as these odds ratios are governed by the following relation:

$$\frac{\frac{P(Y_i^t=2 \mid X_1=x_1)}{P(Y_i^t=0 \mid X_1=x_1)}}{} = \frac{P(Y_i^t=1 \mid X_1=x_1)}{P(Y_i^t=0 \mid X_1=x_1)} * \frac{P(Y_i^t=2 \mid X_1=x_1)}{P(Y_i^t=1 \mid X_1=x_1)}.$$

In other words, the odds ratio for categories $m = 2$ and $m = 0$ is

$$e^{1.21575+1.48505} = 14.8916.$$

By the same token, the odds ratio between categories $m = 3$ and $m = 0$ is

$$e^{1.21575+1.48505+0.4779} = 24.0143.$$

The odds ratios related to the constrained category (i.e., category 0) can be used together to back-transform the logits into separate item-category response probabilities for participant i through the following equation:

$$P\left(Y_i^t=m, m=0, 1, 2, 3 \mid X_1=0.5\right)$$
$$= \frac{e^{\Sigma_0^m \log it^t m}}{e^{\log it^t m=0} + e^{\Sigma_0^1 \log it^t m} + e^{\Sigma_0^2 \log it^t m} + e^{\Sigma_0^3 \log it^t m}}$$
$$= \frac{e^{\Sigma_1^m \log it^t m}}{1 + e^{\log it^t m=1} + e^{\Sigma_1^2 \log it^t m} + e^{\Sigma_1^3 \log it^t m}}. \tag{2}$$

For example, the conditional probability for response $m = 2$ given the X_1 score 0.5 is

$$P\left(Y_i^t=2 \mid X_1=0.5\right)$$
$$= \frac{e^{1.1081+1.2158+1.4851}}{e^{1.1081} + e^{1.1081+1.2158} + e^{1.1081+1.2158+1.4851} + e^{1.1081+1.2158+1.4851+0.4779}}$$
$$= \frac{(3.0284)(14.8916)}{3.0284 + (3.0284)(3.3728) + (3.0284)(14.8916) + (3.0284)(24.0143)}$$
$$= \frac{14.8916}{1 + 3.3728 + 14.8916 + 24.0143}$$
$$= 0.3441$$

The conditional probabilities for response $m = 1$ and 3 given the X_1 score 0.5 according Equation (2) are 0.0779 and 0.5449, respectively, and that for $m = 0$ is 0.0231. In the above example, if the latent trait X_1 is continuous, the adjacent-category logistic link function yields the model known as the partial credit item response model (Masters, 1982). Notice that in Equation (2) the conditional probability can also be expressed exclusively in terms of the odds ratios that are referenced to category 0 (the self-referenced odds ratio for category 0 is 1). In this sense, the adjacent-category logistic model is a special case of the more general

multinomial logistic model that treats an arbitrary chosen item-category (e.g., the first one or the last one) as the baseline reference category to determine the probabilities of item responses.

The logistic link function described in Equations (1) and (2) can also be applied to models with latent endogenous variables that are discrete. For example, an ordinal latent variable can be treated as a dependent measure for one or more latent explanatory variables. The same adjacent-category logistic function described above can be used to relate the latent as well as observed exogenous variables to the logits of the dependent measure, and the conditional probabilities of the latent endogenous variable given the latent exogenous variables, e.g., $P(X_1 = x_1 \mid X_2 = x_2, X_3 = x_3, \ldots Z_1 = z_1, \ldots)$, where $Z_1 = z_1, \ldots$, are observed exogenous variables also known as covariates (e.g., age, ethnicity, etc.), can in turn be derived from the logits through the aforementioned back-transformation.

The Axiom of Local Independence

The response to an item t, Y^t_i, is typically inter-dependent on responses to other items in the same test. Because of the inter-dependency, the joint unconditional probability for a particular set of responses, say, $P(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1)$, is generally not equal to the product of the separate unconditional probabilities for the responses, i.e.,

$$P\left(Y^{t=1}_i=2, Y^{t=2}_i=2, Y^{t=3}_i=1\right) \neq P\left(Y^{t=1}_i=2\right) * P\left(Y^{t=2}_i=2\right) * P\left(Y^{t=3}_i=1\right).$$

It is hypothesized in FM models that the inter-dependency between observed responses is entirely rooted in the latent traits shared among the responses, and with the latent traits statistically controlled, the conditional probabilities of the responses as exemplified in the section of *Logistic Link Function,* are independent (the axiom of local independence). Because of this local independence, the joint conditional probability for a particular response pattern can be obtained as the product of the conditional probabilities of the separate responses.

For example, the conditional probability for participant i with a latent trait score of $X=0.5$ to give a response 2, i.e.,

$$P\left(Y^{t=1}_i=2 | X_1=0.5\right)=0.3441,$$

has been obtained for item $t = 1$ through Equations (1) and (2) in the subsection of *Logistic Link Function*. Through similar link functions conditional probabilities of responses to other items can also be estimated. Suppose the conditional probabilities for $Y^{t=2}_i = 2$ and $Y^{t=3}_i = 1$ (scores 2 and 1 for Item 2 and Item 3) are estimated to be

$$P\left(Y^{t=2}_i=2 | X_1=0.5\right)=0.2900$$

and

$$P\left(Y^{t=3}_i=1 | X_1=0.5\right)=0.3114,$$

then the local independence axiom stipulates that the joint conditional probability for the response pattern, $P(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 0.5)$, is the product of 0.3441, 0.2900, and 0.3114, i.e.,

$$P\left(Y^{t=1}{}_i=2, Y^{t=2}{}_i=2, Y^{t=3}{}_i=1|X_1=0.5\right)=0.3441*0.2900*0.3114=0.0311.$$ (3)

In other words, the latent trait X_1 is hypothesized to account for all associations among $Y^{t=1}{}_i$, $Y^{t=2}{}_i$, and $Y^{t=3}{}_i$. In FM modeling, model fit is evaluated on the basis of the hypothetical probability structure estimated according to the axiom of local independence. Unlike the latent structure undergirding traditional modeling technologies that is limited to mostly linear correlations/covariances, the probability structure of FM modeling is open to a larger variety of associations among manifest variables, and is thus more inclusive.

Discrete and Continuous Latent Traits

FM models can include continuous as well as discrete latent traits. A discrete latent variable may represent unordered clusters in the population (nominal), but it may also reflect subgroups that are ordered on certain latent dimensions (ordinal). An ordinal latent variable may effectively approximate a continuous trait, particularly when the level of the ordinal variable is increased (Aitkin, 1999; Heinen, 1996; Vermunt, 2001). A "discretized" or non-parametric latent trait has an important advantage over a parametric continuous trait in that the former is not dependent on the often overstated assumptions (linearity, normality, etc.) for continuous traits. When their inherent assumptions are violated, the continuous traits are likely to result in biased estimates (Heinen, 1996; Vermunt, 2001; Vermunt & Magidson, 2005a). "Discretized" latent traits are also better-equipped for identifying possible population heterogeneity, including a possible discontinuity in the distribution of latent traits. Unlike a continuous trait whose trait scores are forced to follow the normal distribution, an ordinal discrete trait is nonparametric in the sense that no predetermined distribution is imposed on its trait scores, and the scores can take on various forms of distribution (e.g., a bimodal distribution). It is therefore possible that on the metric of response probability scores of a discrete trait from one or more subgroups are distributed not (or almost not) contiguously with those from the other higher or lower subgroups, manifesting a pronounced population heterogeneity.

The Estimation of Posterior Probability

The joint conditional probability for a particular response pattern described in Equation (3), e.g.,

$$P\left(Y^{t=1}{}_i=2, \quad Y^{t=2}{}_i=2, Y^{t=3}{}_i=1|X_1=0.5\right)=0.0311,$$

can be used in conjunction with the estimated marginal probabilities of the latent trait(s) (e.g., $P(X_1 = 0.5)$) to estimate the posterior probability of the latent trait(s) given the participant's response pattern. The posterior estimation follows the well-known Bayes' theorem,

$$\begin{aligned}
&P\left(X_1=x_1 \mid Y_i^{t=1}=m^{t=1}, Y_i^{t=2}=m^{t=2}, \ldots\right)\\
&=\frac{P(X_1=x_1)P\left(Y_i^{t=1}=m^{t=1}, Y_i^{t=2}=m^{t=2}, \ldots \mid X_1=x_1\right)}{P\left(Y_i^{t=1}=m^{t=1}, Y_i^{t=2}=m^{t=2}, \ldots\right)}\\
&=\frac{P\left(X_1=x_1, Y_i^{t=1}=m^{t=1}, Y_i^{t=2}=m^{t=2}, \ldots\right)}{P\left(Y_i^{t=1}=m^{t=1}, Y_i^{t=2}=m^{t=2}, \ldots\right)}
\end{aligned}$$ (4)

where the denominator, $P(Y^{t=1}{}_i = m^{t=1}, Y^{t=2}{}_i = m^{t=2}, \ldots)$, is the marginal probability for the response pattern $Y^{t=1}{}_i = m^{t=1}, Y^{t=2}{}_i = m^{t=2}, \ldots$, summed over all levels of the latent trait X_1.

A model with a 3-level latent trait X_1 for three observed items, such as the items described in the *The Axiom of Local Independence* section, can be used to illustrate the Bayes' posterior estimation in Equation (4). The three levels of X_1 are valued as $X_1 = 0$, $X_1 = 0.5$, and $X_1 = 1.0$, and their marginal probability estimates can be obtained as a result of model parameter estimation. For purposes of illustration, assume these estimates are, respectively,

$$P(X_1 = 0.0) = 0.1219,$$
$$P(X_1 = 0.5) = 0.1571,$$

and

$$P(X_1 = 1.0) = 0.7210.$$

In the Bayes' formula (Equation (4)), the numerator is the joint probability of a specific response pattern and a given latent trait level, for instance, of response pattern $Y^{t=1}_i = 2$, $Y^{t=2}_i = 2$, $Y^{t=3}_i = 1$ and latent trait level $X_1 = 0.5$, and is determined as

$$P(X_1 = 0.5) * P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 3 | X_1 = 0.5\right) = 0.1571 * 0.0311 = 0.0049 \tag{5}$$

for the example. The denominator of the Beyes' formula is the marginal probability for the same response pattern in question. For example, the marginal probability, $P(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1)$, is the total probability of the response pattern $Y^{t=1}_i = 2$, $Y^{t=2}_i = 2$, and $Y^{t=3}_i = 1$ occurring at all possible levels of X_1 (i.e., 0, 0.5, and 1.0). It is obtained as the sum of the related response pattern and trait level joint probabilities shown below,

$$
\begin{aligned}
&P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1\right) \\
&= P(X_1 = 0) P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 0\right) \\
&+ P(X_1 = 0.5) P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 0.5\right) \\
&+ P(X_1 = 1) P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 1\right)
\end{aligned}
\tag{6}
$$

Notice that one of the terms to be summed on the right hand side of Equation (6) (the second term) is the joint probability for the response pattern at $X_1 = 0.5$, shown to be 0.0049 in Equation (5) above, and those for the other two latent levels (the first and the third terms) can be attained in ways similar to Equation (5) using the relevant probability estimates. Without getting into specifics, assuming that

$$P(X_1 = 0) P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 0\right) = 0.0000$$

and

$$P(X_1 = 1.0) P\left(Y^{t=1}_i = 2, Y^{t=2}_i = 2, Y^{t=3}_i = 1 | X_1 = 1.0\right) = 0.0118,$$

the marginal probability of the response pattern for the denominator of the Bayes' formula depicted in Equation (6) is then,

$$P\left(Y^{t=1}{}_i=2,\ Y^{t=2}{}_i=2,\ Y^{t=3}{}_i=1\right)=0.0000+0.0049+0.0118=0.0167 \tag{7}$$

Substituting the values of Equations (5) and (7) into the Bayes' formula in Equation (4), one obtains the estimated posterior probability for participant i with the response pattern of $Y^{t=1}{}_i = 2$, $Y^{t=2}{}_i = 2$, and $Y^{t=3}{}_i = 1$, to have the latent trait level of $X_1 = 0.5$,

$$P\left(X_1=0.5|Y^{t=1}{}_i=2,\ Y^{t=2}{}_i=2,\ Y^{t=3}{}_i=1\right)=0.0049/0.0167=0.2934.$$

The posterior probability for the same response pattern to have the trait score $X_1 = 1$, on the other hand, is

$$P\left(X_1=1.0|Y^{t=1}{}_i=2,\ Y^{t=2}{}_i=2,\ Y^{t=3}{}_i=1\right)=0.0118/0.0167=0.7066,$$

and that for $X_1 = 0$ is near 0. As 0.7066 is the highest among the three posterior probabilities, participant i, and for that matter, all participants with the same response pattern, should be classified onto the 1.0 level of X_1.

For an ordinal latent trait, the information from the posterior estimation can also be summarized in a weighted sum known as the factor mean score. In the present example, the factor mean score for participant i is determined by first multiplying the posterior probability of each X_1 level by its value (0, 0.5, or 1.0) and then summing the three products. The posterior probabilities of the three X_1 levels for participant i according to the Bayes' theorem are 0.0000, 0.2934, and 0.7066, and the factor mean score for the participant is therefore

$$0.000*0+0.2934*0.5+0.7066*1.0=0.8533.$$

The posterior probability estimation and its resultant posterior classification of the participants provide an additional means of model validation, particularly when the posterior classification results are related to some previously known subgroups. For example, the classification results based on the modeled working memory traits can be referred to the known subgroup with mild mental retardation (MMR) in contrast to that without MMR to evaluate the validity of the modeled working memory traits. The posterior classification results can also be used to unveil possible etiologies of certain exceptional groups, such as the possible etiology of working memory deficiency for the subgroup with MMR. Moreover, although the ordinal trait scores are spaced evenly on the 0-1 scale (e.g., 0, 0.5, and 1 for three levels), the frequency distribution of these scores can take on any shape, including a zero or near-zero frequency count at one or more intermediate levels and thus displaying on the metric of probability a discontinuity in the latent distribution.

The versatility of the FM modeling technology is clearly well-suited for the modeling of working memory tasks. For working memory tasks with discrete item-level responses, FM models with discrete or/and continuous latent traits can be specified and fitted to the observed item responses. Competing models that differ in number, the discrete/continuous property, and the meaning of latent variables can be compared for the selection of the best-fitting models. The defined latent traits of working memory can then be related to the latent trait of intelligence to evaluate the linear/nonlinear relations between these traits. The selected models can generate participant-classification results as well as latent trait scores analogous to factor scores, and these tangible classification results and trait scores of individuals provide an avenue to possible

validation of the latent relationship theoretically derived from these models, alleviating the concern with the over-reliance on the possibly biased upward projection from lower observed correlations to higher latent relations in traditional analyses.

Objectives of the Study

The present study was intended to develop a better-refined model of two verbal-numerical working memory tasks, Digit Span (DS) and Letter-Number Sequencing (LNS) on the basis of the item-level analysis of the tasks. DS comprises of the subtasks of forward span and backward span, with the former as a prototype of STS task and the latter as a working memory task with uncertain characteristics. The backward part of the task has been portrayed by some as still mostly a memory span task (Engle et al., 1999), although it apparently involves additional attention control over the information in the STS. It also apparently engages the interfering operation of backward ordering (BO), although whether the BO operation contributes to individual differences in intelligence in its own right is unclear. A more definitive analysis is needed to determine the separate sources of variability in the backward span subtask.

The LNS task requires the participant to mentally sort the digits from the letters in the interpolated list of numbers and letters in STS. The mental sorting (MS) process is apparently executed by an attention control mechanism, although it is quite possible that the specific mental sorting operation may also be an independent source of variability. In other words, the task is likely to be a mix of three or more working memory subsystems, although the subsystems may not be easily identified within LNS alone.

Tasks with prototypical STS items, such as DS, may be analyzed together with LNS to disentangle the STS mechanism from the others. The combined set of DS and LNS responses may also enable the distinction of the general control mechanism, as the backward span subtask and LNS are conceived to share this component. The collective analysis may also help distinguish the specific attention control mechanisms apparently involved in backward span and LNS, namely, the mechanisms initiated, respectively, by the interfering operations of BO and MS, as latent traits can be specified to underlie either set of responses in addition to that shared in the joint set.

The present study was also set to investigate the discrete or continuous property of the working memory traits. The commonly held belief that these traits are continuous may misrepresent the actual distributions of these traits, particularly for certain exceptional subgroups, such as those with mental retardation or those who are gifted. Aside from the implication that such a discontinuity between the exceptional subgroups and the rest of the population in working memory may shed light on the etiology of these subgroups, the discontinuity in working memory as an important underpinning of intelligence may also suggest a possible discontinuity in the distributions of omnibus abilities, including that of g. Moreover, the discreteness of the ability factors may in part explain why relations between cognitive abilities tend to be nonlinear, as discrete factors are unlikely to relate to one another in a strictly linear form.

Method

Participants

Participants were 1197 Chinese primary school grade-3 and -4 children from Yanchen City and Shanghai, China. Of these children, about 140 had been diagnosed to be cases with mild mental retardation (MMR) and had been specially recruited for a broader project to investigate the cognitive determinants of MMR. These children were previously diagnosed using either the Chinese Wechsler Intelligence Scale for Children- Revised (C-WISC-R, Gong & Cai, 1994) or the Chinese Stanford-Binet Intelligence Test, and the diagnoses were also made using

the Chinese Adaptive Behavior Scale for the Children (Yao & Gong, 1993), which is largely an adapted version of its US counterparts for problems with life adaptations. These children were recruited from several municipal districts in Shanghai, with the main body of the sample including nearly all children (over 95%) meeting the criteria in two of the districts. Exclusory criteria had been adopted for the children with MMR so that those with behavioral and health problems other than the subnormal level of intelligence were not included in the subgroup. Among those with MMR, 17 participants had missing values on one or more measures adopted in the study. A listwise deletion treatment of missing values was used in the present study because of two considerations. First, among those with MMR, scores could be reasonably assumed as missing at random, and with the downward weights assigned to the cases with MMR, the 17 participants with MMR who did not have complete records would have very little, if at all, impact on the model estimation. The second, more substantive consideration was that displaying the classification results with reference to the MMR status was purported to demonstrate the validity of the modeled working memory traits in the present study, and the validation would be more transparent with actual rather than imputed scores. The listwise deletion of missing values resulted in 123 cases with MMR and 990 in the total sample who had complete records for the study. The other participants in the sample were regular school children included in school-based clusters from three local schools in Yanchen, China. Because the subgroup with MMR in the sample was disproportional to what is expected in the population (about 2%), the subgroup was weighted down to be comparable to the population percentage for all analyses. The children's age in month was treated as an active covariate in all analyzed models to control the possible age influence on the task performance.

Measures

An extensive battery of cognitive tasks, omnibus ability scales, and achievement tests were administered to the sample. The present study was focused on two working memory tasks, DS and LNS, and the summary scores of WISC-R Verbal and Performance subscales and the total achievement test score.

Digit Span—DS is one of the subscales of the Chinese-WISC-R, which was adapted from its US counterpart. Similar to the original, it consists of a forward subtask and a backward subtask. The forward part includes eight digit lengths ranging from two to nine, and each length has two trials. The backward part comprises of seven digit lengths (2 to 8), with two trials for each length. Several score categories (i.e., the 0 category for forward digit lengths-2 and -3 and that for the backward digit length-2) had too few counts (e.g., less than 10), and would have resulted in unreliable model parameter estimates for the categories. These categories were merged with the adjacent categories, and the merging generated the recoded lengths-2 and -3 scores in the forward subtask and the recoded length-2 score in the backward subtask.

Letter Number Sequencing—LNS is a working memory task with mental sorting as its interfering operation. In each trial of LNS, a list of numerical digits and letters (i.e., A, B, …, Z) were read in a mixed order to the participant. After each trial list was presented, the participant was asked to recall the digits first and then the letters in their respective sequential orders in the trial. The length of the list ranges from two to eight, and each length has three trials. The scores were obtained by aggregating the number of correct responses (0-3) for each length. The shortest length (length 2) generated two score categories (0 and 1) with too few counts, and they were thus aggregated to ensure reliable category-level parameter estimates.

The WISC-R Verbal and Performance Summary Scores—Four verbal subscale scores (Information, Similarity, Vocabulary, and Comprehension) were summed to for a Verbal summary score and four performance scale scores (Picture Completion, Picture Arrangement, Block Design, and Object Assembly) were added to obtain a Performance summary score.

The Total Achievement Test Score—The achievement test consisted of a Chinese subtest and a Mathematics subtest, with 30 items per subtest. The Chinese subtest included questions in three categories, vocabulary, sentence structures, and reading comprehension, and the Mathematics subtest consisted of questions in the three categories of mathematical concepts, the use of mathematical formulas, and application problems. All items were multiple choice questions. The subtest items were found to have suitable psychometric properties in a pre-test sample of 3rd- and 4th-grade children from different primary schools in the two Chinese cities. The total achievement (Achieve) test score was the sum of the grade-standardized Chinese and Mathematics subtest scores.

Analysis

The Finite Mixture Modeling Method—The FM modeling method was used to test and compare competing models of DS and LNS separately as well as those of DS and LNS combined. As the item responses of DS and LNS were ordinal variables with limited numbers of categories, the latent traits postulated to underlie these responses were modeled to relate to the observed responses through the adjacent-category logistic link function.

Model selections in the present study followed two main guidelines. First, as latent traits of working memory and omnibus intellectual abilities are conventionally treated as continuous variables, albeit subsumed under the often faulty assumption of normality, models that varied in the number of continuous traits were tested to determine the proper number of the latent traits undergirding the observed indicators. These continuous latent traits, kept orthogonal in the present study, would represent the independent sources of variability as conventionally construed. Once the best-fitting continuous-trait model was selected, the possibility of the trait being more suitably represented by a non-parametric ordinal-trait counterpart would be entertained. Technically, the baseline model for a discrete latent trait would be a 1-level-trait model, and models with increasingly more levels for the trait would be tested until additional levels of the trait would not lead to a substantially improved fit. In practice, the 1-level-trait model is obviously false for measures of cognitive abilities, so the focus of model selection in the present study was between the best-fitting ordinal-trait models and their continuous-trait prototypes. The best fitting ordinal-trait models, if found to be better fitting, would be chosen over the continuous-trait counterparts not only for their closer fit to the data but also for the theoretically important considerations listed in the The Finite Mixture Modeling Approach to the Hurdles/Discrete and Continuous Latent Traits subsection of the introduction section.

Model fit in the present study was indicated by a set of model fit indexes, including the Likelihood Ratio Chi-Square statistic (L^2), Log Likelihood (LL) index and the LL-based Bayesian Information Criterion (BIC) that takes into consideration both model-data discrepancy and model parsimony for the model assessment (Schwarz, 1978). The L^2 index is an asymptotic chi-square index, and is only available when all observed indicators are discrete (e.g., working memory task items). It was adopted to evaluate the omnibus model fit of working memory models whenever applicable in the present study primarily because it can be used to generate bootstrapping significance tests of model fit. When some of the observed indicators are continuous (e.g., the Verbal, Performance, and Achievement summaries scores) the L^2 index is unavailable and the LL and LL-based indexes are suitable (Vermunt & Magidson, 2005b). The BIC index was mostly adopted for the selection of competing models, as the index is an effective indicator when used to systematically search for potential sources of model-data discrepancy (Gelman & Rubin, 1999). In the present study, such potential sources of discrepancy included the number and the definition of the latent traits and the possible discrepancy between models with improperly prescribed normal, continuous traits and the data with non-normal latent distribution. A better model fit is indicated by a lower magnitude of all these fit indexes.

As explained in the introduction, the FM modeling method is equipped to produce classification results on the basis of posterior probability estimates. For discrete latent variables, the posterior membership probability of the participant for each discrete factor level, often termed as the modal, is estimated and is then compared to those for the other factor levels to classify the participant into the most likely factor level subcategory and these classification results provide an additional means of model verification when judged against certain known categories of the participants in the actual sample. In the present study, the subgroups with and without MMR served as two known participant categories against which the validity of the model-based classification results were evaluated.

Modeling Digit Span and Letter-Number Sequencing Separately—Models that differed in the number of continuous latent traits (e.g., one, two, and three) were first tested for DS and LNS, respectively, to provide guidance for the further modeling of DS and LNS. Multiple traits in the same model were specified to be orthogonal so that independent variance sources could be determined. More refined models were then specified on the basis of the best-fitting continuous trait models (a two-factor model for DS and a one-factor model for LNS) with the necessary zero factor loading constraints and with the continuous traits replaced by ordinal discrete traits that varied in category levels to attain better-fitting models.

Modeling Digit Span and Letter-Number Sequencing Jointly—Based on the respective best fitting models of DS and LNS, models for both DS and LNS combined were tested. For both tasks, a latent trait of STS loading on all 22 items of both tasks was treated as the default for any additional latent traits defined. The backward span items of DS (except for the length-2 item) and all LNS items were postulated to be underpinned by a GAC trait, and the LNS items and the DS backward span items each might reflect two additional traits of MS and BO, respectively. The plausibleness of these traits was tested, and the appropriate category levels of the discrete traits determined. All latent traits for the two working memory tasks remained orthogonal in the models to better capture the independent sources of variance.

Modeling Verbal and Performance Summary Scores and the Achievement Score —Just as the relatively basic cognitive capacity of working memory may be a hybrid of independent sources of variability, the variability shared by the complex intellectual measures of intelligence and achievement is likely to be a blend of various variance sources. "Breaking down" the general variability of omnibus abilities, or *g*, into more refined pieces, however, would be considerably more difficult (Brody, 1992; Carroll, 1993; Deary, 2000; Jensen, 1998), and was not the objective of the present study. In the present study, the variability shared among such ability measures was treated as a unitary endogenous factor, which is how it is typically treated in conventional analyses.

Using the structural equation modeling method, a factor analysis of the subtests constituting the summary ability scores (i.e., the eight WISC Verbal and Performance subtests and the Chinese and Mathematics achievement subtests) led to a dominant general factor and three minor Verbal, Performance, and achievement group factors. The latent trait reflecting the variability shared among the three summary scores is plausibly a close proxy of the dominant general factor found among the eight subtests. In the present study, the continuity/discreteness property of the trait was subjected to a test. The best-fitting continuous or discrete latent trait as a proxy of the general factor (G) was then treated as the criterion for the working memory traits to evaluate the relative relevance of these traits to intelligence in three extended models encompassing some (DS or LNS) or all working memory and intelligence variables. In the extended models, the orthogonal working memory traits each were specified to predict the intelligence trait, and the strength of the prediction was evaluated.

Statistical Program for Finite Mixture Modeling—The FM modeling of the study was accomplished using Latent Gold 4.5 (Vermunt & Magidson, 2005b, 2008). The program is designed to specify and test models incorporating various kinds of observed and latent variables (nominal, ordinal, and continuous) and is capable of accommodating linear and nonlinear relations that arise from the diverse composition of variables. For ordinal item-responses such as those to the working memory tasks in the present study, the Latent Gold 4.5 default link function is adjacent-category logistic. The program also provides the option to generate classification results and latent trait scores for individual participants on the basis of the estimated posterior probabilities. In the case of continuous latent traits these scores are tantamount to conventional factor scores or, in the parlance of item response theory, the person parameter estimates. For ordinal discrete traits, the program both classifies participants onto the most probable trait levels (modals) and assigns factor mean scores to the participants. These factor mean scores are comparable to the factor scores of continuous traits, and can be used to evaluate inter-trait correlations. Because outcomes of the estimation include not merely yields of linearly projected inter-trait correlations but also group membership classifications, they are subjected to additional means of validation.

The model parameter estimation for each specified model was implemented by Latent Gold 4.5 using its posterior mode (PM), which is an adjusted (penalized) form of the maximum likelihood method.

Results

Table 1 lists the results from the preliminary statistical analyses, including the Pearson correlations among the task-level scores of DS and LNS, the Verbal and Performance summary scores and the total achievement score, and the WISC-R Full IQ, with the reliability estimates in the diagonal and the descriptive statistics at the bottom. The correlations between the working memory tasks and the WISC-R summary scores are comparable to those reported in previous studies where very strong estimated factorial relations between working memory traits and intelligence traits were obtained.

Table 2 lists the model fit indexes from the competing models for DS, LNS, DS and LNS combined (DS+LNS), and for the Verbal, Performance, and Achieve summary scores. The model selections were mostly based on the BIC index, with a smaller value of BIC indicating that the model is more preferable in its balance between closeness-of-fit and model parsimony. Because the task items included in the study generated sparse data (e.g., a table with more than 3^{15} cells for DS), the chi-square statistic used to evaluate model fit is likely to be biased. The best fitting DS, LNS, and DS+LNS models were also tested statistically using the Latent Gold chi-square (L^2)-based bootstrapping procedure, each with 500 Montecarlo replications that generated the probability distribution defined by the estimates of the specified model.

The Models for Digit Span

For DS, a two-continuous-trait model, equivalent to a two-dimensional general partial credit item response theory model, fitted better than its one-trait counterpart. The BIC index from the three-continuous-trait model, however, was worse than that for the two-trait model. The two traits in the two-trait model appear to have clear connotations, as one trait loads significantly on nearly all 15 DS items and the other only has significant loadings on the backward span items. A constrained model with the loadings of the second trait in the two-trait model on all eight forward span items and the length-2 item of the backward span fixed to zero produced a better fit. The first trait can be plausibly conceived as the STS trait whereas the second trait seems to reflect certain additional processes in the backward span part.

Based on the better-fitting two-continuous-trait model of DS, the possibility of these two traits being discrete was entertained, and models with two ordinal traits varied in levels were tested. The optimal model appears to be one with seven levels for the STS trait and six levels for the second trait, and the model is also a better fit than its continuous-trait counterpart. In addition, the bootstrapped p value of 0.13 obtained for the model suggests that the model fits the data satisfactorily. Figure 1 is a graphic illustration of the model, where the loading values are linearly approximated nonlinear model estimates (Vermunt & Magidson, 2005a) and depict the relevance of the items to the latent traits in the manner of traditional factor analysis.

The Models for Letter-Number Sequencing

Models with one, two, and three continuous latent traits were tested and compared for LNS. The two-factor model is the best-fitting of all three, but the interpretation of the two factors is less clear-cut than those of DS. One of the factors has relatively weak loadings on the length-2, -3, and -4 items, and the other factor has strong loadings on these items. The two factors load about equally on the length-5, -6, -7, and -8 items. It seems whatever mechanisms of working memory are functioning in LNS, they can not be clearly unraveled within the confines of the task alone.

To further explore possibly better fitting models of LNS, models with one discrete latent trait but different levels were compared. It seems that the 9-level discrete factor model is a better candidate than the other models, including all three continuous factor models. The p value based on the maximum likelihood estimation is 1.00, and bootstrapped p value for the model is 0.04. The only marginally acceptable fit of the model indicated by the bootstrapping outcome may again be an indication of the multi-dimensionality for the LNS item responses. Whereas the meaning of the possibly additional source(s) of variance in LNS awaits further clarification, the 9-level discrete factor is nonetheless likely to represent the main source of the predictive variability in LNS, and was chosen to be the predictor of intelligence in the next part of the analysis. Figure 2 is a graphic illustration of the model.

The Models for Both Digit Span and Letter-Number Sequencing

In light of the better fit of the discrete factor models for the separate tasks of DS and LNS, the models to be tested for both DS and LNS were confined to those with exclusively discrete latent traits. The default model was that with one discrete factor with nine levels that loaded on all DS and LNS items to represent the STS trait. Additional discrete factors were included progressively into the model to test the importance of other possible working memory subsystems.

The first additional discrete trait added to the default model was a factor with zero constraints on the loadings of eight DS forward span items and the length-2 backward span item and with non-zero loadings on the rest of the DS backward span items and all seven LNS items, and this factor was presumed to reflect the additional GAC processes. The model with seven levels for the STS factor and six levels for the GAC factor bested its competitors with the same factors but different levels in model fit.

The next two ordinal factors added to the model were related to the LNS items and to the DS backward span items 3 thru 8, respectively. These factors were expected to tap the specific control mechanisms given rise by the operations of MS and BO. Of the tested 4-discrete-factor models, the model with five, six, five, and five levels, respectively, for the STS, GAC, MS, and BO factors fit the best, and the two additional factors for the LNS items and the DS backward span items appeared to improve the model fit substantially. The bootstrapped p value of 0.72 for the 4-discrete-factor model also suggests that the model fits the data well. The model is figuratively described in Figure 3.

The Models for Verbal, Performance, and Achievement Summary Scores

The Verbal, Performance, and Achieve summary scores were treated as continuous variables representing crystallized, fluid, and achievement abilities, and the latent G trait underlying them would be conceived as a proxy of *g*. Customarily this trait would be treated as a continuous one, but as the working memory traits were all found to be better characterized as discrete, it would be of interest to also examine the scale property of this trait.

Models with one discrete latent trait but with various numbers of levels to account for the three summary scores were compared to each other and to the 1-continuous-factor model. The model with one 5-level discrete factor was found to fit better than all the other competitors, including the continuous factor model. The 5-level factor for the two summary scores was used as the criterion to be related to the working memory traits in the next phase of the analysis.[1]

The Models Relating Working Memory Traits to Intelligence

The chosen model candidates for the separate DS and LNS tasks as well as for the DS+LNS combination were each broadened to include the intelligence criterion trait defined by the Verbal, Performance, and Achieve summary scores. The discrete factors of working memory were specified to be exogenous variables in the extended models for the endogenous intelligence trait (G) so that the unique contributions of the working memory factors to the G factor could be investigated.

The investigation of the contributions from the specific working memory factors to G was conducted using three approaches, (1) evaluating the regression weights related to the working memory-to-G contributions; (2) testing the nested models in which the contribution from one or more exogenous working memory factors were constrained to zero to gauge the impact of the zero constraints on model fit; and (3) obtaining posterior trait scores of the working memory factors and G to compute Pearson correlations between these trait scores.

Table 2 lists the model fit indexes from the extended models with and without zero constraints on the working memory-to-G paths (full and nested models). The specified factors of DS and LNS and the STS and GAC factors of DS+LNS are apparently important to the model fit, as dismissing any of them would lead to a serious worsening of the model fit. The relative importance of MS and BO to the fit of the extended model including both DS and LNS and the criterion measures, however, seems less definitive. Constraining the path from BO to G leads to a worse (higher) BIC index, but the path from MS to G seems expendable—nullifying it does not generate a poorer BIC index. Furthermore, constraining both the paths from MS and BO to G does not give rise to a worse BIC index. The outcomes shown in Table 3 and Table 4 appear to add further confusion about the roles of MS and BO.

Table 3 displays the parameter estimates of the working memory-to-G paths, including the MS-to-G and BO-to-G paths in the DS+LNS & Verbal+Performance+Achieve full model. All these paths estimates are statistically significant at p<0.01, although the BO-to-G path estimate is only marginally so.

Table 4 describes how the working memory factor mean scores are correlated with the G factor mean score. The two DS working memory traits, STS and the trait standing for the additional attention control demanded by the backward recall, both have significant correlations with the

---

[1]Model selection results based on another widely used index, AIC3 (Bozdogan, 1993; d=3; Andrews & Currim, 2003), also uniformly favored the ordinal-factor models over their continuous counterparts. The selection of the best-fitting ordinal-factor models based on AIC3 yielded largely the same results, with the exception for the ordinal trait underlying the LNS items. Model selection based on AIC3 led to a best-fitting 8-level model instead of a 9-level model, but the predictive strength of the two factors for the G factor in the extended model was nearly the same.

G factor mean score and the multiple correlation (R) is 0.62 between the working memory scores and the intelligence score. Both working memory trait scores add significantly to the prediction, with the R square changes due to the STS and the other factor as 0.16 and 0.17, respectively. The factor underlying LNS appears to be a strong predictor for intelligence, as evinced by the strong correlation of 0.80 between the mean factor scores for working memory and intelligence.

The four working memory factor mean scores from the full extended model of DS+LNS all have significant correlations (shown in parentheses) with the G trait score and they jointly have a multiple correlation of 0.85 with the latter. The factor mean score of MS makes a quite substantial unique contribution to the variability of the G score (R square change: 0.13), whereas the factor mean score presumably reflecting BO only adds 0.02 to the explained variability of the G score beyond the other factor mean scores. These seemingly puzzling outcomes regarding the roles of MS and BO in G may partly be caused by the relatively low factor reliability of the two factors, particularly that of MS. As can be seen in the diagonal of the correlation matrix (in parentheses), the standard R squared classification statistic reflecting the factor reliability is 0.34 for MS, notably lower than those for the other factors.

With the inconsistency surrounding the roles of MS and BO in G, and in order to obtain a more reliable posterior classification, the nested model featuring only unconstrained paths from STS and GAC to G was accepted to generate the classification results. This nested model not only generated a better BIC index, but also resulted in the STS and GAC factors that are more reliable (standard R squared indexes both greater than 0.80). The multiple R relating the factor mean scores of STS and GAC to that of G is 0.82, and both the STS and GAC trait scores share substantial unique variances with the G score.

The Discontinuity in Ability Distribution

Although the distribution of an ordinal latent trait is not bound to be normal, with its ordinal levels gradually increasing/decreasing, each level (modal) subcategory is still likely to be filled with non-zilch frequencies. A zero frequency in a modal subcategory would reveal a pronounced form of distributional irregularity, namely, a discontinuity in the distribution of the working memory trait in question, as the blank category would indicate a gap on a probabilistic metric between two nonempty subcategories that are actually adjacent in rank order. The joint classification outcome with GAC as the primary factor and STS as the secondary factor presented in Table 5 manifests such a discontinuity in the probability distribution of the GAC factor.[2]

The two lowest subcategories of GAC (Modals 1 and 2) are adjacent to the virtually empty third lowest subcategory (Modal 3, N=1) and are thus segregated from the next nonempty, higher subcategory (Modal 4). The validity of the discontinuity in classification seems to be supported by the composition of the two outlying low subcategories in terms of the MMR status, although MMR was not included as a classification standard. These two subcategories contain 80% of the 123 cases with MMR, and these cases with MMR constitute about 50% of the two subcategories. It should also be noted that the non-MMR cases classified into the two subcategories are not necessarily "misclassified", as the G scores of those without MMR in the two subcategories are rather low. For example, the average factor mean score of G for the 30 cases without MMR in the GAC Modal 1 subcategory is 0.43, close to two standard deviations below the average (0.72) of those without MMR in the next nonempty subcategory

[2]The discontinuity in the distribution of the attention control factors was also apparent in the DS & Verbal+Performance+Achieve and the LNS & Verbal+Performance+Achieve models, as both models resulted in one or more empty subcategories wedging between the lowest subcategory and the other higher, non-empty subcategories of the factors representing additional attention control (the Additional Attention Control factor and the LNS factor, respectively).

(Modal 4). These cases would be likely to fall into the group formerly known as "borderline mental retardation".

The GAC and STS joint classification also appears to give rise to an accurate diagnostic system of working memory deficiency for individuals with a subnormal level of intelligence, including MMR. For example, if one uses the diagonal of the table as the diagnostic standard and includes all cases in the upper-left triangle above the diagonal (i.e., ≤GAC Modal 1 and ≤STS Modal 4, ≤GAC Modal 2 and ≤STS Modal 3, ≤GAC Modal 4 and ≤STS Modal 2, and ≤GAC Modal 5 and ≤STS Modal 1), one can correctly identify 86% cases with MMR at a false alarm rate of 9%. The hit-to-false alarm ratio will be higher if one takes into account that many diagnosed cases without MMR actually had very low G scores and thus a subnormal level of intelligence.

Discussion

The finite mixture (FM) models fitted to the item responses of the Digit Span (DS) and Letter-Number Sequencing (LNS) tasks in the present study are illustrative in several ways. First, multiple latent working memory traits operating in the same task, such as those of Short-Term Storage (STS), General Attention Control (GAC), and Mental Sorting (MS) in LNS, and of STS, GAC, and Back-Ordering (BO) in DS, were extricated on the basis of these models. Such modeling would not be as effectively conducted on the task level where each task as an aggregate of multiple mechanisms is treated as one variable and disentangling separate sources of variance based on the correlations among these aggregated variables is intrinsically difficult.

The general mechanisms of STS and GAC were both found to be predictive whereas the explanatory power of the specific control factors MS and BO was less definitive. The ambiguity about the roles of MS and BO may have to do with the lower reliability of these factors, a limitation possibly remediable in future studies by adding more items to LNS and the backward part of DS. These specific control traits may also have differential predictive strengths for different criteria. It does seem safe to conclude, though, that the main predictive power of working memory comes forth from the general sources of STS and GAC, and working memory tasks varied in their specific aspects of executive control tap these two general sources to different degrees, some more intensively (e.g., LNS for GAC), and some less so (e.g., DS for GAC). This finding may help resolve the dispute about which subsystems of working memory are essential for predicting intelligence.

The underpinnings of DS that have caused some confusion can be conceptualized from this perspective. The task appears to tap GAC in addition to STS, although the loadings of its backward span items on GAC were lower than those of LNS items in the present study. The BO trait underneath the backward span items had a modest correlation with the proxy of g, but did not manifest itself as a highly distinctive predictor of g. These uncertainties about GAC and BO in the backward span items are probably why the findings about DS have been inconsistent. It also should be noted that, although BO failed to display a pronounced predictive power for G in the present study, it is probably premature to repudiate its distinctive predictive value. BO may still play an indispensible part in more specific intellectual criteria, for example, achievement. Stronger correlations of the backward span task with achievement tests than those of its MS counterpart has been found in certain large, representative samples (Luo, Thompson, & Detterman, 2006), suggesting that BO is likely to have a criterion-dependent value.

The two selected working memory tasks, DS and LNS, were both verbal-numerical tasks, but the same FM modeling approach can be adopted with visual tasks to distinguish various independent variance sources in these tasks and to evaluate how they bear on intelligence. The apparently more refined item-level FM modeling of working memory tasks also may provide

guidance about task-level analyses, so that the within-task independent sources of variance will not be masked in the task-level analyses.

The raw correlations between the working memory tasks of the present study and intelligence measures were about 0.40s-0.60s in binary and multiple correlations, which seem to be in the best expected range of linear correlations between manifest variables of working memory and intelligence (Ackerman et al, 2005). One could expect, justifiably, higher correlations among latent variables that are theoretically unaffected by error variances, but how much higher the "purified" correlations should be is open to question. As a result, the important issue of whether the majority of the *g* variability can be credibly ascribed to working memory has remained debatable.

The item-level modeling of latent traits resulted in binary and multiple correlations between the factor mean scores of the working memory traits underlying DS, LNS, and DS+LNS and that of G noticeably higher (in the range of 0.60s-0.80s) than the linear relationship estimates between the raw working memory task scores and IQ. These correlations are not merely projected factorial correlations in conventional factor analysis and structural equation models that prevail only in theory. They are correlations calculated using actual scores produced by the model-based posterior classification, and these posterior trait scores are subjected to validation through other possible means, such as that provided by the effective identification of the children with mild mental retardation (MMR).

Although the working memory traits were determined in the present study without a specific reference to the subgroup with MMR, some of the discrete working memory traits, such as the GAC factor underlying DS+LNS, appeared to effectively differentiate children with MMR. In particular, about 80% cases with MMR were classified into the two lowest subcategories of the GAC factor, and the majority of the cases without MMR in these subcategories was also at least more than one standard deviation below the global mean in the G score. These results provided an independent validation for the virtue of the modeled working memory traits. Furthermore, the results seemed to underscore STS and GAC, particularly the latter, as the predominant cognitive determinants for the subnormal level of intelligence, especially for MMR, which is a finding mandating theoretical and practical interest in its own right.

The results of the study also spur an interesting question about the distributions of working memory traits. Discrete working memory factors were found to better fit the observed data than continuous ones, and there appeared to be a discontinuity in the distribution of these latent traits.[3] Because the latent traits in the present study were determined on the ratio scale of response probability, the discontinuity is unlikely to be an artifact of the chosen score metric. A discontinuity in ability distributions could plausibly account for the phenomenon often attributed to the mythic "Law of Diminishing Returns", namely, notably stronger between-test correlations in the lower end of the ability distribution, as many at the lower end of the distribution, although not so conspicuously outlying on the scale of observed scores, are literally outliers from the rest of the population in their latent ability traits.

Such a discontinuity could also be a sign of qualitative differences between ability subgroups, and the implicated population heterogeneity, albeit speculative at present, is worth careful examination, especially if the same discontinuity pattern is cross-validated in other samples with different age and cultural backgrounds. Whether the successive but disjoint levels of the working memory factors are merely a sample-based phenomenon, or whether they indeed grasp

---

[3]When the latent traits were treated as continuous factors, the correlations between the posterior scores of working memory traits and that of the G counterpart were similar to those reported in Table 3, with the highest multiple correlation between the continuous working memory factor scores and that of G counterpart above 0.80. The models with continuous factors, however, led to worse model fit indexes and were unable to account for possible distributional irregularities.

the essence of the ability traits in the population, the issue is of obvious significance to the very nature of the scientific research on working memory and the working memory-intelligence relationship, and should be given full attention in future investigations.

Acknowledgments

References

Ackerman PL, Beier M,E, Boyle MO. Working memory and intelligence: The same or different constructs? Psychological Bulletin 2005;131:30–60. [PubMed: 15631550]

Aitkin. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics 1999;55:218–234.

Andrews RL, Currim IS. A comparison of segment retention criteria for finite mixture logit models. Journal of Marketing Research 2003;40(2):235–243.

Baddeley, AD.; Hitch, GJ. Working memory. In: Bower, GA., editor. Recent Advances in Learning and Motivation. Vol. 8. Academic Press; New York: 1974. p. 47-89.

Bozdogan, H. Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In: Opitz, O.; Lausen, B.; Klar, R., editors. Information and Classification, Concepts, Methods and Applications. Springer; Berlin: 1993. p. 40-54.

Brody, N. Intelligence. Academic Press; San Diego, CA: 1992.

Deary, IJ. Looking Down on Human Intelligence. Oxford University Press, Inc; New York: 2000.

Carroll, JB. Human Cognitive Abilities: A Survey of Factor Analytic Studies. Cambridge University Press; Cambridge, UK: 1993.

Colom R, Abad FJ, Quiroga MA, Shih PC, Mendoza CF. Working memory and intelligence are highly related constructs, but why? Intelligence 2008;36:584–606.

Cowan, N. Attention and memory: An integrated framework. Oxford University Press; New York: 1995.

Detterman DK, Daniel MH. Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. Intelligence 1989;13:349–359.

Dolan CV, van der Maas HLJ. Fitting multivariate normal finite mixtures subject to structural equation modeling. Psychometrika 1998;63:227–253.

Engle RW, Tuholski SW, Laughlin JE, Conway ARA. Working memory, short term memory and general fluid intelligence: A latent variable approach. Journal of Experimental Psychology: General 1999;128:309–331. [PubMed: 10513398]

Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK. Not all executive functions are related to intelligence. Psychological Science 2006;17:172–179. [PubMed: 16466426]

Gelman A, Rubin DB. Evaluating and using statistical methods in the social sciences: A discussion of "A critique of the Bayesian Information Criterion for model selection". Sociological Methods Research 1999;27:403–410.

Gong Y, Cai T. Chinese Wechsler Intelligence Scale for Children – Revised. Journal of Chinese Clinical Psychology 1994;2(1):1–6.

Heinen, T. Latent class and discrete latent trait models: Similarities and differences. Sage; Thousand Oaks, CA: 1996.

Hunt E. The role of intelligence in modern society. American Scientist 1995:356–367.

Jensen, AR. The g factor. Praeger Publishers; 88 Post Road West; Westport; CT 06881: 1998.

Jensen AR. Regularities in Spearman's law of diminishing returns. Intelligence 2003;31:95–105.

Lazarsfeld, PF.; Henry, NW. Latent structure analysis. Houghton Mifflin; Boston: 1968.

Lubke GH, Muthén B. Investigating population heterogeneity with factor mixture models. Psychological Methods 2005;10(2):21–39. [PubMed: 15810867]

Luo D, Thompson LA, Detterman DK. The criterion validity of tasks of basic cognitive processes. Intelligence 2006;34:79–120.

Maij - de Meij AM, Kelderman H, Flier H. van der. Fitting a mixture IRT model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. Applied Psychological Measurement 2008;32:611–631.

Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–174.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A. The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. Cognitive Psychology 2000;41:9–100.

McLachlan, GJ.; Peel, D. Finite Mixture Models. John Wiley & Sons, Inc.; New York: 2000.

Muthén, BO. Second-generation structural equation modeling with a combination of categorical and continuous latent variables. New opportunities for latent class/latent growth modelling. In: Collins, LM.; Sayer, A., editors. New Methods for the Analysis of Change. APA; Washington, DC: 2001a. p. 291-322.

Muthén, BO. Latent variable mixture modelling. In: Marcoulides, GA.; Schumacker, RE., editors. New Developments and Techniques in Structural Equation Modelling. Lawrence Erlbaum Associates; Mahaw, NJ: 2001b. p. 1-33.

Muthén BO. Beyond SEM: General latent variable modeling. Behaviormetrika 2002;29(1):81–117.

Oberauer K. Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition 2002;28:411–421.

O'Connell, AA. Logistic regression models for ordinal response variables. SAGE Publications; Thousand Oaks, California: 2006.

Rosen VM, Engle RW. Forward and backward serial recall. Intelligence 1997;25:37–47.

Schwarz G. Estimating the dimension of a model. Annals of Statistics 1978;6(2):461–464.

Spearman C. General intelligence, objectively determined and measured. American Journal of Psychology 1904;15:201–293.

Vermunt JK. The use of restricted latent class models for defining and testing nonparametric and parametric IRT models. Applied Psychological Measurement 2001;25:283–294.

Vermunt, JK.; Magidson, J. Latent class cluster analysis. In: Hagenaars, JA.; McCutcheon, AL., editors. Applied Latent Class Analysis. Cambridge University Press; Cambridge: 2002. p. 89-106.

Vermunt, JK.; Magidson, J. Factor analysis with categorical indicators: a comparison between traditional and latent class approaches. In: Van der Ark, A.; Croon, MA.; Sijtsma, K., editors. New Developments in Categorical Data Analysis for the Social and Behavioral Sciences. Erlbaum; Mahwah: 2005a. p. 41-62.

Vermunt, JK.; Magidson, J. Latent GOLD 4.0 User's Guide. Statistical Innovations Inc.; Belmont, Massachusetts: 2005b.

Vermunt, JK.; Magidson, J. LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module. Statistical Innovations Inc.; Belmont, MA: 2008.

Yao S, Gong Y. The Adaptive Behavior Scale for Children and its norm in the urban and rural areas. Acta Psychologica 1993;1:38–42. In Chinese.
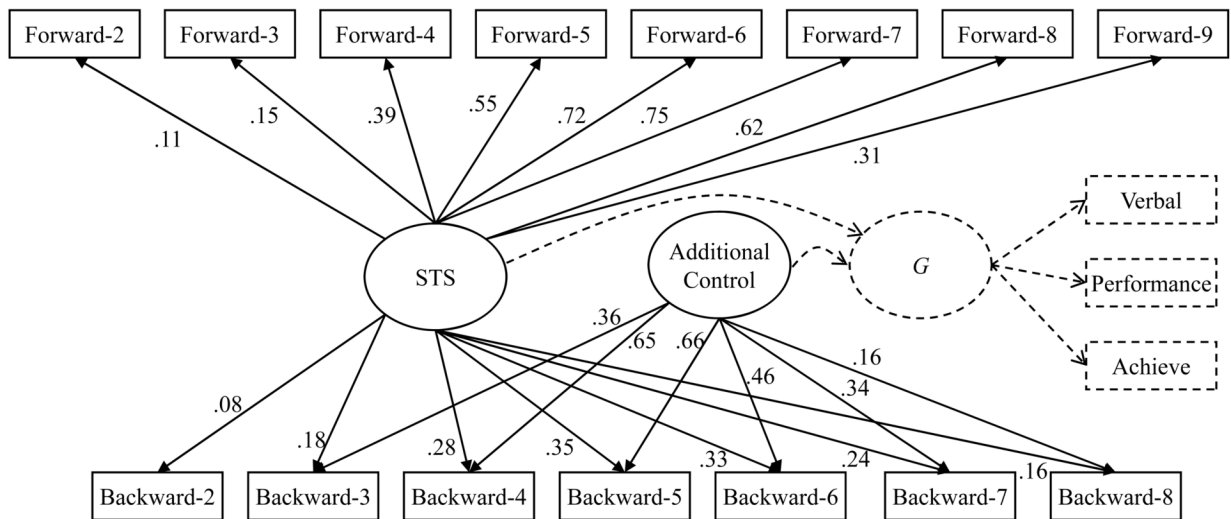
Figure 1.
2-Discrete-Factor Model of Digit Span (Levels: 7 for Short-Term Storage (STS), 6 for Additional Control).

Note: Forward-2 thru -9: Lengths 2 thru 9 of Forward Digit Span. Backwar-2 thru -8: Lengths 2 thru 8 of Backward Digit Span. G (5 levels): The General factor of the Verbal, Performance, and Achievement Total summary scores. Loading estimates displayed are the linear approximates (Vermunt & Magidson, 2005a) of the model slope parameter estimates that indicate the strength of the relationship between the latent traits and the specific task items in the style of the traditional factor analysis. All listed loading estimates are significant at $p<0.05$. The dashed lines indicate the elements to be included in the extended model.
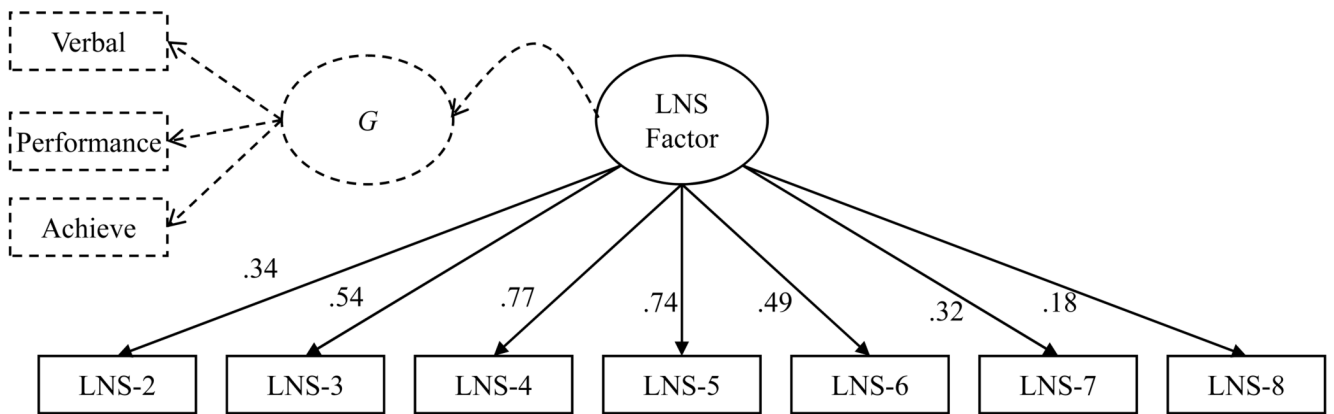
Figure 2.
1-Discrete-Factor Model of Letter-Number Sequencing (Levels: 9)
Note: LNS-2 thru -8: Lengths 2 thru 8 of Letter-Number Sequencing. G (5-levels): The General factor of the Verbal, Performance, and Achievement Total summary scores. Displayed loading estimates are the linear approximates of the corresponding slope estimates. All listed estimates are significant at p<0.05. The elements to be included in the extended model are indicated by dashed lines.
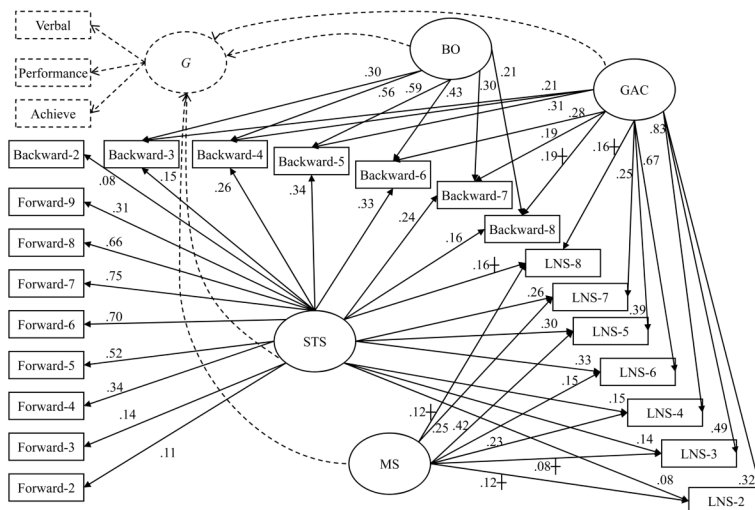
Figure 3.
4-Discrete-Factor Model of Digit Span + Letter-Number Sequencing (DS+LNS; Levels: 5 for Short-Term Storage (STS); 6 for General Attention Control (GAC), 5 for Mental Sorting (MS), 5 for Back-Ordering (BO). G (5-levels): The General factor of the Verbal, Performance, and Achievement Total summary scores.

Note: The loading estimates are the linear approximates of the related slope estimates.

Estimates labeled by ┼ are insignificant at p<0.05. Elements depicted with dashed lines are the additional entities for the extended model.

Table 1

Correlations, Reliability Estimates and Descriptive Statistics of the Task-Level DS and LNS scores and the Intelligence Test Scores

	DS	LNS	Verbal	Performance	Achieve	Full IQ
DS	(0.77)					
LNS	0.48	(0.71)				
Verbal	0.45	0.48	(0.92)			
Performance	0.35	0.45	0.57	(0.84)		
Total Achieve	0.48	0.53	0.58	0.50	(0.91)	
Full IQ	0.58	0.56	0.88	0.82	0.63	(0.95)
Mean	16.48	9.12	41.51	42.06	0.01	105.09
SD	3.75	2.61	11.13	8.42	1.88	16.48

Note: Acronyms for the variables are, DS for Digit Span total, LNS for Letter-Number Sequencing total, Verbal for the sum of four WISC-R Verbal subscales, Performance for the sum of four WISC-R Performance subscales, and Achievement for the total standard scores of Chinese and Mathematics achievement tests. Except for the WISC-R Full IQ (Full IQ), the reliability estimates displayed in the diagonals are item-based Cronbach Alpha coefficients. The reliability coefficient for the WISC-R Full IQ is the estimate reported for the normative sample.

Table 2

Model Fit Indexes for Competing Models of DS, LNS, DS+LNS, and Verbal+Performance+Achieve Scores

Separate Models of DS, LNS, DS+LNS, and Verbal+Performance+Achieve

DS

Model	L^2	df	p (Bootstrap)	LL	LL-based BIC
1-continuous-factor	1962.24	833	0.00	−5835.17	11968.49
2-continuous-factor	1384.57	817	0.00	−5545.99	11498.54
3-continuous-factor	1336.69	801	0.00	−5522.05	11559.07
2-continuous-factor with constraints	1388.84	826	0.00	−5548.12	11441.82
2-discrete-factor (levels: 7, 6)	1304.17	817	0.00 (0.13)	−5505.77	11418.10

LNS

Model	L^2	df	p (Bootstrap)	LL	LL-based BIC
1-continuous-factor	499.34	852	1.00	−4057.20	8290.62
2-continuous-factor	431.39	844	1.00	−4023.42	8277.29
3-continuous-factor	409.90	836	1.00	−4012.48	8309.63
1-discrete-factor (levels: 9)	375.87	845	1.00 (0.04)	−3995.46	8214.60

DS+LNS

Model	L^2	df	p (Bootstrap)	LL	LL-based BIC
1-discrete-factor (levels: 8)	7923.77	795	0.00	−9852.72	20213.07
2-discrete-factor (levels: 7, 6)	7197.93	778	0.00	−9489.79	19602.29
4-discrete-factor (levels: 5, 6, 5, 5)	6786.08	759	0.00 (0.72)	−9327.23	19319.36

Verbal+Performance+Achieve

Model	LL	LL-based BIC
1-continuous-factor	−7142.38	14345.72

	LL	LL-based BIC
1-discrete-factor (levels: 5)	−7109.78	14307.40

Extended Models of Working Memory & Verbal+Performance+Achieve

DS & Verbal+Performance+Achieve

Model	LL	LL-based BIC
3-discrete-factors (levels: 7, 6, 5) full	−14575.81	29658.94
nested, STS-to-G path fixed	−14626.16	29752.88
nested, Add. Ctrl.-to-G path fixed	−14621.18	29742.92

LNS & Verbal+Performance+Achieve

Model	LL	LL-based BIC
2-discrete-factors (levels: 8, 5) full	−13063.78	2638.76
nested, LNS factor-to-G fixed	−13214.74	26733.93

DS+LNS & Verbal+Performance+Achieve

Model	LL	LL-based BIC
5-discrete-factors (levels: 5, 6, 5, 5, 5) full	−18137.08	37139.06
nested, STS-to-G path fixed	−18195.98	37250.97
nested, GAC-to-G path fixed	−18168.51	37195.15
nested, MS-to-G path fixed	−18130.79	37119.73
nested, BO-to-G path fixed	−18141.97	37142.07
nested, MS&BO-to-G paths fixed	−18133.36	37118.33

Note: DS: The Digit Span task. LNS: The Letter-Number Sequencing task.

Verbal+Performance+Achieve: Summary scores of the four WISC-R Verbal subscales, four WISC-R Performance subscales, and the total standard scores of Chinese and Mathematics achievement tests. STS: Short-Term Storage. Add. Ctrl: Additional Attention Control of the Digit Span task. G: The general factor of the Verbal, Performance, and the Achievement Total summary scores. GAC: The General Attention Control factor. MS: The Mental Sorting factor. BO: The Back-Ordering factor. Models with continuous indicators do not yield the chi-square (L^2) index, and the L^2-based bootstrapped p values for model fit are thus not available for models with the continuous Verbal, Performance, and Achieve indicators. As a result, the L^2-based bootstrapped p values (displayed in parentheses) were only provided for the besting fitting DS, LNS, and DS+LNS models. All Bayesian Information Criterion (BIC) indexes obtained for model selections were derived from the log-likelihood (LL) values.

Table 3

Parameter Estimates for the Working Memory-to-G Paths in the Extended Models

Parameter Term	Coefficients	s.e.	z-value	p-value
Model: DS & Verbal+Performance+Achievement				
STS to G	7.86	1.15	6.82	0.00
Add. Control to G	5.44	0.82	6.65	0.00
Model: LNS & Verbal+Performance+Achievement				
LNS Factor to G	23.27	3.34	6.97	0.00
Model: DS+LNS & Verbal+Performance+Achievement, (1) Full-Initial				
STS to G	17.50	2.69	6.50	0.00
GAC to G	12.67	1.84	6.90	0.00
MS to G	16.12	4.57	3.52	0.00
BO to G	3.90	1.45	2.69	0.01
Model: DS+LNS & Verbal+Performance+Achievement, (2) Constrained-Accepted				
STS to G	16.02	2.33	6.88	0.00
GAC to G	16.74	2.34	7.11	0.00

Table 4

Correlations among Working Memory and Intelligence Factor Mean Scores and WISC-R Full IQ

Factor Mean Scores of DS and Verbal+Performance+Achieve		
	STS	Additional Control	G
STS	0.85		
Additional Control	0.13	0.75	
G	0.46	0.47	0.91
Full IQ	0.49	0.37	0.72

Factor Mean Scores of LNS and Verbal+Performance+Achieve	
	LNS Factor	G
LNS Factor	0.86	
G	0.80	0.81
Full IQ	0.65	0.90

Factor Mean Scores of DS+LNS and Verbal+Performance+Achieve				
	STS	GAC	MS	BO	G
STS	0.85(0.85)				
GAC	0.13 (0.14)	0.83(0.77)			
MS	0.01 (0.06)	0.09 (0.31)	0.49(0.34)		
BO	0.13 (0.09)	0.13 (0.16)	0.06 (0.05)	0.65(0.64)	
G	0.54 (0.53)	0.68 (0.62)	0.02 (0.53)	0.13(0.21)	0.86(0.87)
Full IQ	0.50 (0.49)	0.53 (0.47)	0.01 (0.47)	0.14(0.20)	0.89(0.88)

Note: Italic values in the diagonal are the standard R squared classification statistics reflecting the reliability of the related factors (Vermunt & Magidson, 2005b). Values in the parentheses are based on the DS+LNS and Verbal+Performance+Achieve full model with free paths from all four working memory factors to G.

Tble 5

Frequency Counts Based on the Joint Classification of the STS and GAC Factor Modals Resultant from the Extended Model of LNS+DS & Verbal +Performance+Achieve

STS Modal:	1	2	3	4	5	Subtotal (G-mean/sd)
GAC Modal: 1						
MMR	24	14	36	3	1	78 (0.07/0.07)
Non-MMR	0	0	11	13	7	31 (0.43/0.19)
GAC Modal: 2						
MMR	7	2	10	1	0	20 (0.12/0.09)
Non-MMR	1	2	28	23	15	69 (0.56/0.20)
GAC Modal: 3						
MMR	0	0	0	0	0	0
Non-MMR	0	0	0	0	1	1 (0.71/NA)
GAC Modal: 4						
MMR	5	3	10	1	0	19 (0.22/0.08)
Non-MMR	8	11	89	89	58	255 (0.72/0.16)
GAC Modal: 5						
MMR	2	1	1	0	0	4 (0.24/0.06)
Non-MMR	4	9	120	148	63	344 (0.82/0.13)
GAC Modal: 6						
MMR	1	0.	1	0	0	2 (0.41/0.27)
Non-MMR	0	3	39	66	25	133 (0.91/0.10)
GAC Modals: All						
MMR	39	20	58	5	1	123 (0.11/0.10)
(G-mean/sd)	(0.07/0.09)	(0.09/0.07)	(0.15/0.11)	(0.17/0.10)		(0.19/NA)
Non-MMR	13	25	287	339	169	833 (0.77/0.18)

STS Modal:	1	2	3	4	5	Subtotal (G-mean/sd)
(G-mean/sd)	(0.39/0.13)	(0.64/0.20)	(0.68/0.18)	(0.82/0.15)		(0.86/0.15)
Total:	52	45	345	344	170	956 (0.68/0.28)

Note: MMR: The subgroup with mild mental retardation. Non-MMR: The subgroup without mild mental retardation. G-mean/sd: The mean and standard deviation of the factor mean scores of the G factor. The case weight used in the analysis for participants with MMR was removed for the tabulation. Modals 1, 2, …, 5 of the STS factor and 1, 2, …, 6 of the GAC factor are ranked from the lowest to the highest in the related abilities. The G-mean/sd values are means and standard deviations of the factor mean scores of the G factor for the marginal subgroups of the GAC and STS factors.