Intelligence 38 (2010) 231–241

Contents lists available at ScienceDirect

Intelligence

Use of factor mixture modeling to capture Spearman's law of diminishing returns☆

Matthew R. Reynolds [a,]*, Timothy Z. Keith [b], S. Natasha Beretvas [b]

[a] Psychology and Research in Education, University of Kansas, Joseph R. Pearson Hall, 1122 West Campus Road, Lawrence KS, 66045-3101, United States
[b] The University of Texas at Austin, TX, United States

ARTICLE INFO

Article history:
Received 4 June 2009
Received in revised form 22 January 2010
Accepted 25 January 2010
Available online 18 February 2010

Keywords:
Factor mixture modeling
Intelligence
Spearman's law of diminishing returns

ABSTRACT

Spearman's law of diminishing returns (SLODR) posits that at higher levels of general cognitive ability the general factor (*g*) performs less well in explaining individual differences in cognitive test performance. Research has generally supported SLODR, but previous research has required the a priori division of respondents into separate ability or IQ groups. The present study sought to obviate this limitation through the use of factor mixture modeling to investigate SLODR in the Kaufman Assessment Battery for Children-Second Edition (KABC-II). A second-order confirmatory factor model was modeled as a within-class factor structure. The fit and parameter estimates of several models with varying number of classes and factorial invariance restrictions were compared. Given SLODR, a predictable pattern of findings should emerge when factor mixture modeling is applied. Our results were consistent with these SLODR-based predictions, most notably the *g* factor variance was less in higher *g* mean classes. Use of factor mixture modeling was found to provide support for SLODR while improving the model used to investigate SLODR.

1. Introduction

Test scores on measures of cognitive ability correlate positively. A general factor, *g*, is posited as one reason for these correlations, and thus is useful in explaining individual differences on cognitive ability test performance (Bartholomew, 2004; Jensen, 1998). A great deal of research involving human cognitive abilities has focused on understanding *g* and its real-world correlates (Gottfredson, 1997; Herrnstein & Murray, 1994; Jensen, 1998).

Researchers are often interested in assessing the contribution of the *g* factor in explaining individual differences in cognitive ability test scores. In a standard factor analytic model it is implicitly assumed that either the contribution of *g* is constant across individuals or that the factor loadings used to calculate the contribution of *g* represent an average of factor weights that vary somewhat across individuals (Wolfle, 1940). Spearman's law of diminishing returns (SLODR), or cognitive ability differentiation, however, posits that *g* does not contribute equally at different levels of ability; rather it decreases as the level of general ability increases (Detterman & Daniel, 1989; Spearman, 1927). If such a phenomenon exists, the assumption regarding constant *g* loadings across individuals is tenuous. Moreover, some findings related to *g* and its correlates may need to be revisited to account for its presence.

Spearman was the first to describe SLODR when he observed that correlations among mental ability tests were stronger in children with low general ability (1927). These stronger correlations in lower ability groups had been observed by others since Spearman (e.g., Maxwell, 1972), but Detterman and Daniel (1989) were credited with rediscovering SLODR. They provided one of the first rigorous tests of SLODR. Detterman's systems

☆ The manuscript was derived from a doctoral dissertation defended successfully in May 2007 by the first author, who would like thank his other dissertation committee members, Keenan Pituch, Dan Powers, and Janay Sander. We are grateful to Alan Kaufman, Mark Daniel, and the American Guidance Service for providing the data used in this research.

* Corresponding author.
E-mail address: mreynolds@ku.edu (M.R. Reynolds).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.01.002

theory of intelligence is also one of the few theoretical explanations of SLODR (Detterman, 1999).

SLODR has been investigated with a variety of methods. Typically though, interrelations among cognitive ability test scores have been compared across different levels of general cognitive ability. Evidence in support of SLODR is typically demonstrated via findings of cognitive ability scores correlating more strongly (Detterman & Daniel, 1989; Legree, Pifer, & Grafton, 1996) or a *g* factor (or first principal component) accounting for a greater proportion of subtest score variance (Deary, Egan, Gibson, Austin, Brand, & Kellaghan, 1996; Jensen, 2003, te Nijenhuis & Hartmann, 2006) in lower IQ groups. Although research in general supports the presence of SLODR, its presence has not been supported in all studies (e.g., Facon, 2004; Hartmann & Reuter, 2006; Saklofske, Yang, Zhu, & Austin, 2008). Possible explanations for the different findings across studies include mixing age differentiation with ability differentiation (e.g., Facon, 2004, 2008; Kane & Brand, 2006), issues related to the sample or tests used in the analysis (see Hartmann & Teasdale, 2004 for discussion), and different selection procedures used to choose ability groups (Carlstedt, 2001; Hartmann & Reuter, 2006; Nesselroade & Thompson, 1995; Reynolds & Keith, 2007).

1.1. SLODR and factor models

Researchers have increasingly turned to confirmatory factor analytic (CFA) methods to study intelligence. An important advantage of such model-based approaches is that researchers may match their theoretical models with their analytic models. Popular models of intelligence include higher-order factor models such as those based on three-stratum theory (Carroll, 1993) or verbal–perceptual–mental rotation abilities (Johnson & Bouchard, 2005). How should SLODR appear in such higher-order models? In an unstandardized CFA solution the second-order *g* factor variance would be expected to “shrink” at higher levels of *g*. If SLODR operated at the level of the first-order factors, or is somehow related to all of the common factors, then first-order factor residual variances would be expected to change as well.

If SLODR (or stronger correlations in lower IQ groups) is a *g*-related phenomenon, then in the standardized solution the loadings of the first-order factors on *g* (or *g* loadings in general) should decrease at higher ability levels. If it is related to first-order factors, then the loadings of the subtests on the first-order factors should decrease at higher ability levels. The relative contribution for each order of factors (in a standardized solution) can be decomposed via a Schmid–Leiman transformation so that comparisons of the relative contributions of *g* and first-order factors can be made. Because proportion of subtest variance explained is calculated from standardized factor loadings in this solution, if SLODR is present then subtest residual variances should be larger for higher ability groups (i.e., the proportion of subtest variance explained by the factors should decrease, namely, this should be due to the *g* factor).

1.2. Group division in SLODR research

One nagging potential confound in studies of SLODR is the comparison of different ability or IQ groups that have been formed a priori. Groups are commonly formed by splitting the sample in half at the mean or median of a full scale IQ score (e.g., Jensen, 2003), a general factor score (e.g., Carlstedt, 2001; Reynolds & Keith, 2007), or a subtest score not included in the analysis (e.g., Detterman & Daniel, 1989) prior to submitting the data for analyses. One limitation with selecting groups based on such scores is that the cut-point between high and low ability groups, the number of groups, and various other selection decisions are left up to the discretion of the researcher. Such somewhat arbitrary division into groups is driven by pragmatics, and is neither model- nor theory-based. A second well-known limitation with this method is that the range restriction related to the dichotomization of a continuous distribution of IQ scores into low and high ability groups makes it difficult to find SLODR even if it is present (Cohen, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002). A third limitation is that even when groups are selected in this manner it is assumed that the constructs are measured similarly across ability groups, and rarely is this assumption explicitly tested (Nesselroade & Thompson, 1995). In fact, when selecting groups based on observed scores, the factor structure based on the whole sample may be distorted (see Muthén, 1989 for discussion).

The application of multi-group mean and covariance structure analysis (MG-MACS) can be used to overcome the third limitation (whether the constructs are being measured similarly in different ability groups) through tests of factorial invariance (Meredith, 1993; Nesselroade & Thompson, 1995; Reynolds & Keith, 2007). Invariance tests are important because if the measurement of the latent constructs differs between groups, it does not make sense to make comparisons between groups. Reynolds and Keith (2007) applied MG-MACS to test whether a higher-order factor model representing the three-stratum model of intelligence was invariant across low and high ability groups. The model was indeed invariant (i.e., equal unstandardized factor loadings) and allowed for defensible comparisons across groups (Reynolds & Keith, 2007). In the unstandardized CFA solution, the *g* factor variance was significantly less in the high ability group and supported the presence of SLODR.

MG-MACS is appealing because it allows for CFA models to be utilized and for tests of factorial invariance. The method, however, suffers from the shortcoming of forming ability groups a priori, and under the assumption of being able to select participants for these groups without error.

There is no theory that describes a specific cut-point for determining low and high ability groups, nor is there a theory that suggests how many ability groups exist along a latent *g* dimension. The focus of selecting different ability groups may in fact detract from studying how the phenomenon is related to the construct of *g*. SLODR is not about groups of people, it's about *g*, a continuous variable.

1.3. Factor mixture modeling

Factor mixture modeling (FMM) is a modeling technique that may be used to overcome the limitation of requiring researchers to define ability groups a priori, and it takes into account the possibility of error in group membership when groups are formed (Bauer, 2005). FMM represents the integration of confirmatory factor analysis and latent class (or finite mixture) modeling into a general latent variable modeling framework (Dolan & van der Maas, 1998; Lubke & Muthén, 2005; Yung, 1997). Specifically, continuous (factors) and categorical (classes) latent variables are utilized. FMM is appealing in that within each latent class a confirmatory factor model with continuous latent variables underlies responses on the measurement instruments. Moreover, the separation of the latent constructs from error and unique variance within each class means that those constructs are closer to the "true" constructs of interest (Jedidi, Jagpal, & DeSarbo, 1997).

Unlike MG-MACS models in which groups are defined before submitting the data for analysis, mixture models produce probabilities of class (or group) membership for each respondent so that the classification of individuals is based on the substantive model. In terms of SLODR, rather than defining groups a priori based on somewhat arbitrary cut-points along a continuous distribution, the classes are formed from model probabilities.

FMM is capable of identifying heterogeneity in a population, a direct approach, where the latent classes represent qualitatively different groups of people. Some have referred to these classes as "clusters" (Lubke & Spies, 2008). These classes are often interpreted as categorically different taxa.

Another approach, referred to as an indirect approach, is to use factor mixture modeling to capture non-normality in the distributions of latent variables within a homogenous population (Dolan, 2009; Lubke & Spies, 2008). Some refer to these latent classes as mixture components. Interpretation of the class as a cluster or a mixture component is left up to the theory and interest of the researcher. Here we do not assume that there will be qualitatively different latent groups of people within the overall population, rather we will present a novel method that may be used to potentially detect and understand SLODR (if it exists). That is, classes will not represent different groups defined by nature, rather they will represent mixture distributions that may be used to model potential non-normality in the distributions of the latent variables (e.g., negatively skewed *g* factor).

1.4. FMM and SLODR

If SLODR existed in a dataset, a predictable set of findings should emerge when using FMM. First, a model allowing for more than one latent class may better describe the data than a single class model.[1] Second, if different classes are identified, they should differ in their latent *g* means. Third, the classes should differ in their latent *g* variances, and the class with the higher level of latent *g* should show smaller variance on the latent *g* factor. Last, there may be differences in first-order residual variances and subtest residual variances.

The initial steps that would be taken to test these predictions are as follows. First, an acceptable higher-order single class confirmatory factor analysis (CFA) model of intelligence would be derived. This model would serve as the within-class model. Second, a two-class model would be specified, but the two classes' latent *g* means would be freely estimated. Hence, rather than a typical one-class CFA model, a two-class higher-order CFA model would be modeled, and the only difference between the two classes would be in the latent *g* means. Model fit indices would be used to assess which model fit better. Third, the restriction of equal subtest residual variances across classes would be released. Although there is not a necessary SLODR hypothesis related to the unstandardized subtest residuals, if differences exist, then constraining these to be equal would force variance unrelated to the common factors into the common factors and obfuscate an understanding of the true nature of SLODR. In a standardized solution, however, the subtest residual variances would be expected to be greater in the group with the higher *g* mean. Fourth, in a higher-order factor model, the first-order (residual) factor variances may change, and would be tested by releasing the equality constraints on these variances across classes. Last, the restriction of equal *g* variances across classes would be removed. If SLODR were present, the model fit would improve, and the *g* factor variance would be smaller in the higher ability class. Note that in a standardized solution, this variance difference would show up in the factor loadings. To support SLODR the standardized *g* factor loadings would be lower in the higher ability group indicating less influence of *g* at higher levels (also implying lower intercorrelations among subtests at higher levels of *g*).

The purpose of this study was thus to use FMM to test for and describe SLODR. We used data from the norming sample of a popular individually administered intelligence test designed for children in our analyses.

2. Method

2.1. Instrument

The KABC-II is an individually administered test of cognitive abilities developed for children and adolescents ages 3 to 18. The extended battery takes about 90 to 100 minutes to administer. The KABC-II was developed to measure five stratum II abilities (Gc, Gv, Gf, Glr and Gsm) from Cattell–Horn–Carroll (CHC) theory, and a higher-order general intelligence (*g*) factor. The model shown in Fig. 1 is the second-order factor structure of the KABC-II supported in previous research (Reynolds, Keith, Fine, Fisher, & Low, 2007) and was used as the within-class model in this study.

According to the item level data analysis provided in the KABC-II manual (Kaufman & Kaufman, 2004), a Rasch calibration of the subtest items was performed. Results from this analysis provided the item difficulty distributions relative to the ability distributions so that any floor or ceiling effect problems could be identified and items that did not fit could be ascertained. Once the final set of items and rules were decided upon, the items were appropriately sequenced. Most tests showed an adequate floor and ceiling effect for the age levels intended. For one test, Story Completion, time bonus scoring was included so that a higher ceiling could be achieved for adolescents (although this resulted in lower reliability). Our analyses included the untimed scoring for all subtests because previous research had suggested the factor

[1] We say "may" rather than "would" because SLODR is a continuous phenomenon acting across all levels of *g*. Even if SLODR were operating, it might not result in better fitting multi-class models. The finding of better fitting multi-class models with the properties described, however, would be evidence for SLODR. We are grateful to a reviewer for elucidating this point.

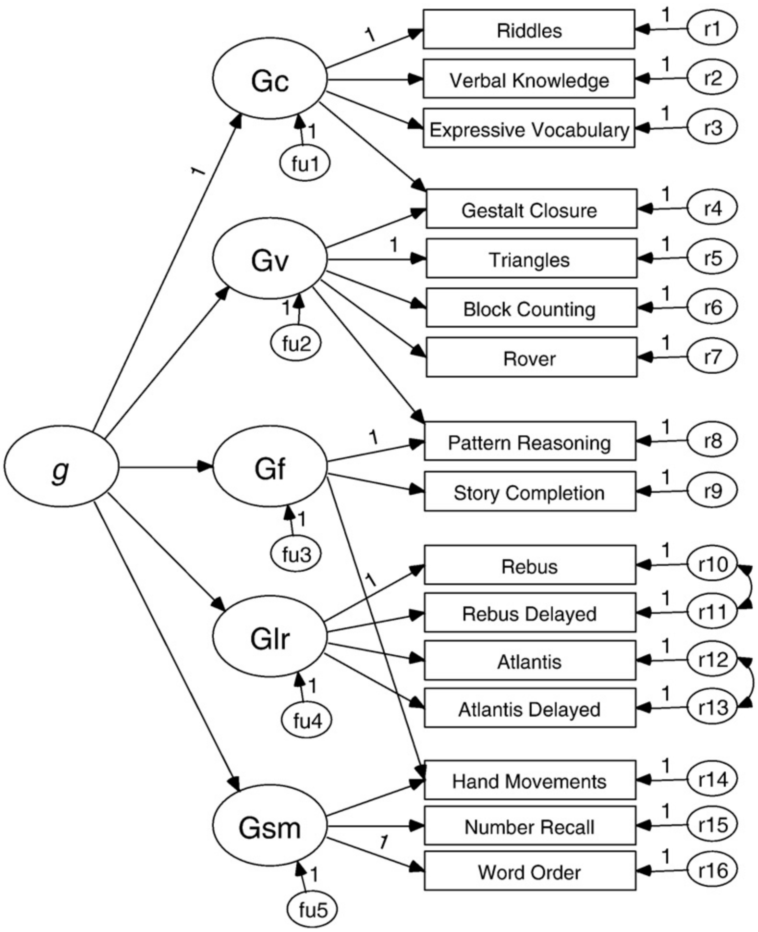

Fig. 1. The KABC-II factor structure.

structure of the KABC-II fit better with untimed scores (Reynolds et al., 2007). Because possible ceiling effects were only present for one subtest, we did not consider such ceiling effects to have an important influence on our findings. For more detailed information on item and test development see Kaufman and Kaufman (2004).

2.2. Participants

Participants included children and adolescents ages 6 through 18 drawn from the KABC-II standardization sample. The sample was stratified across the United States according to age, parental education, (i.e., maternal education if available or paternal education if maternal education was not available), ethnic group, geographic region, educational placement, and educational status (Kaufman & Kaufman, 2004). Overall sample characteristics are shown in Table 1. This overall sample, however, was split randomly into two subsamples to cross-validate findings.

2.3. Descriptive statistics

The sample sizes, means, and standard deviations for each randomly selected subsample are presented in Table 2. All of the scores reported in Table 2, and all scores used in the analysis, were age-standardized. Univariate skewness and kurtosis within each subsample were considered to be acceptable as the values all fell within a range of −0.50 to 0.50. Skewness values approaching and exceeding 2.00 and kurtosis values approaching

Table 1
Demographic characteristics for ages 6 to 18 of the normative sample for the KABC-II.

Variable	*N*
	Total
Total sample	2375
Sex	
Boys	1186
Girls	1189
Race/Ethnicity	
White	1475
Hispanic	420
African American	352
Other	128
Ages	
6–14	200 per year
15–17	150 per year
18	125 per year

Table 2
Descriptive statistics for the subtests in the subsamples.

	Subsample one			Subsample two		
Subtest	*N*	*M*	*SD*	*N*	*M*	*SD*
Atlantis delayed	1174	9.93	2.90	1172	9.84	2.90
Atlantis	1188	9.96	3.12	1187	9.95	3.15
Block counting	1188	9.76	3.01	1187	10.06	3.02
Expressive vocabulary	1188	9.70	2.97	1187	9.92	2.97
Gestalt closure	882	10.11	2.92	856	10.02	2.97
Hand movements	1188	10.03	2.93	1187	10.00	2.84
Number recall	1188	10.04	2.91	1187	10.14	2.87
Pattern reasoning untimed	1188	9.81	3.03	1187	9.90	2.98
Rebus delayed	1157	10.04	2.91	1154	10.02	3.04
Riddles	1188	9.99	3.07	1187	10.14	3.08
Rebus	1188	10.10	3.05	1187	10.08	3.07
Rover	1188	10.04	2.97	1187	10.01	3.00
Story completion untimed	1188	9.92	2.89	1187	9.89	2.82
Triangles untimed	1188	9.96	2.91	1187	10.06	2.84
Verbal knowledge	1188	9.93	3.08	1187	10.02	2.96
Word order	1188	9.86	2.94	1187	9.88	2.81

and exceeding 7.00 may result in problems associated with non-normality (Curran, West, & Finch, 1996).

2.4. Models

FMMs were estimated using M*plus* software (Muthén & Muthén, 1998–2007). The within-class confirmatory factor model, as shown in Fig. 1, was consistent with the test structure and CHC theory (Carroll, 1993; Reynolds et al., 2007). Using the CHC-CFA model with continuous latent variables as the within-class model and a categorical latent class variable, several FMM models were estimated in which parameter constraints were released across classes. Model specifications for two-class models are detailed below. Each model was tested in both subsamples.

2.4.1. Single-class model

The baseline model was a one-class, second-order confirmatory factor model (see Fig. 1). This factor structure was the within-class factor structure specified in subsequent multi-class models.

2.4.2. Two-class models

2.4.2.1. g mean model. The second model estimated was similar to the baseline model, except that two classes were modeled and a second-order *g* factor mean difference was freely estimated. Specifically, for model identification purposes, one class served as the reference class with its latent *g* mean set to zero. The other class mean was then freely estimated. The resulting latent mean estimate represented the difference between the latent *g* mean for class two compared to the reference class. All of the other model parameters, second-order loadings, second-order *g* variance, first-order residual variances, and subtest residual variances were specified as class-invariant. In addition, all first-order factor and subtest intercepts were class-invariant, thus the only difference between the classes was that they differed in the level of *g*.

2.4.2.2. Subtest residuals model. The third model, the subtest residuals model, had the same specification as the *g* means model, however, in addition to the *g* mean difference, subtest residual variances were estimated freely within each class. If the classes differed in the subtest residual variances (i.e., specific variance and measurement error), the model fit indices should support the subtest residuals model over the *g* means model.

2.4.2.3. First-order residuals model. The first-order residuals model had the same specification as subtest residuals model, however, in addition to a latent *g* mean difference and freely estimated subtest residual variances, the first-order factor residual variances were estimated freely within each class. If the classes differed in first-order residual variances, the model fit indices should support the first-order residuals model over the subtest residuals model.

2.4.2.4. g variance model. In addition to modeling differences in *g* means, subtest residual variances and first-order residual variances (if significant) in the two-class *g* variance model, the equality constraint on the *g* factor variance was removed. This model was of particular interest because if SLODR was present, this particular model should have provided the best fit.

2.4.3. Three-class models

The same steps mentioned above were followed for this series of analyses, except three classes were allowed (rather than two) and the results were compared with those of the two-class models.

2.4.4. Model with covariates

To help validate interpretation of latent class membership (Lubke & Muthén, 2005), between- and within-class covariates were added to the final model and the resulting conditional FMMs were estimated. The background characteristic covariates were investigated as predictors of class membership and as potential sources of within-class *g* variance.

2.5. Model evaluation

The models were evaluated by examining the fit indices and interpreting the model parameter estimates. The fit indices used for evaluation included the Akaike Information Criteria (Akaike, 1987), the Bayesian Information Criteria (Schwarz, 1978), and the sample-sized adjusted Bayesian Information Criteria (Sclove, 1987). Lower AIC, BIC, and aBIC values indicate better fitting models. It is possible, however, that the AIC, BIC, and the aBIC provide conflicting evidence. The aBIC has been found to perform better in latent variable mixture models (Henson, Reise, & Kim, 2007). In addition to the information criteria, parameter estimates (latent mean differences, latent variances, etc.) were interpreted with regard to the SLODR-based predictions made in the Introduction.

Table 3
Fit indices for subsample one.

Model	Information criterion AIC	BIC	aBIC
Single class	85,219.7	85,509.3	85,328.3
Two-class			
g mean	85,162.0	**85,466.8**	85,276.3
Subtest residuals	85,106.7	85,492.7	85,251.3
First-order residuals	85,112.4	85,518.8	85,264.7
g variance	**85,092.7**	85,483.9	**85,239.3**
First-order residuals + *g*	85,099.7	85,511.2	85,253.9
Three-class			
g mean	85,166.0	**85,481.0**	85,284.1
Subtest residuals	85,046.5	85,524.1	85,225.5
First-order residuals	85,058.1	85,571.1	85,250.3
g variance	**85,028.3**	85,516.0	**85,211.1**
First-order residuals + *g*	85,042.1	85,581.2	85,244.5

Note. Bolded numbers represent the lowest information criterion (IC) value in the two- and three-class models for each of the three ICs. In the first-order residuals + *g* model, *g* variances and first-order residual variances were free to vary across class.

3. Results

3.1. Subsample one

3.1.1. Two-class models

As shown in Table 3, according to the BIC, the *g* mean model fit best, followed by the *g* variance model with the first-order residual variances constrained equal across classes. In contrast, the AIC and aBIC showed the best fit for the *g* variance model (i.e., one that allowed for *g* mean differences and class-specific subtest residual and *g* factor variances). The results based on the AIC and aBIC were those that would be predicted if SLODR was present. Research supports use of the aBIC with latent variable mixtures, and thus these results were given more interpretative weight (Henson et al., 2007). Although, for the most part, the information criteria-based model selection results were in line with SLODR-based predictions, the resulting class membership data and estimated parameters should also be consistently in the same direction as predicted by SLODR. The important parameter estimates associated with each latent class in the two-class *g* variance model were examined to assess whether the predictions were consistent with SLODR (see Table 4).

The classes differed substantially in *g* level. The latent mean difference in *g* was statistically significant ($p<0.05$). The difference in the latent means was 1.56 with a corresponding effect size estimate of $d=0.77$ supporting a large difference in *g* means. For ease of presentation, the two classes will thus be referred to as “low” and “high” ability class.

The average probabilities for most likely class membership were 0.76 (low) and 0.73 (high). The relevant parameter estimates are shown in Table 4. As predicted, the estimated variance of *g* in the low ability class (with a value of 5.29) was larger than that for the high ability class (with an estimated variance of 1.90). The model fit indices supported selection of the model in which these factor variances were unconstrained across classes.

In the standardized solution, as expected, the average standardized subtest residual variance was also larger in the high ability class (0.61) compared to the low ability class (0.46). These standardized residuals represent the proportion of subtest variance not explained by the common factors. Hence, the average proportion of subtest variance explained by the common factors was 0.54 in the low ability class and 0.39 in the high ability class. This proportion of variance explained was decomposed to estimate the average variance explained by the second-order and first-order factors. In a higher-order model, this can be accomplished via the Schmid–Leiman transformation (or by subtracting the squared total standardized *g* effect on the subtest from the squared multiple correlation for that subtest). In the low ability class, the average proportion of subtest variance explained by *g* was 0.38 and by the broad abilities was 0.16. In the high ability class, the average proportion of variance explained by *g* was 0.17 and average proportion of variance explained by the broad abilities was 0.18. More descriptive output related to the classes is located in the Appendix.

Several statistical tests were performed to test whether the participants were distributed differently across classes based on background characteristics. First, we tested whether the age (in years) of the participants in each class was evenly distributed across the age range. The age distribution did not differ significantly across classes, (χ^2 (12) = 12.78, $p=0.38$). In addition, no significant relation was found between sex and class membership (χ^2 (1) = 0.37, $p=0.55$). Not surprisingly, however, a significant association between SES and class was found (χ^2 (12) = 66.65, $p<0.01$). That is, those who had a parent who was more educated tended to be in the higher ability class with a statistically significant correlation of 0.23 between SES and class membership (Kendall's tau-$b=0.21$). This finding is consistent with the positive correlation generally found between SES and *g* (see Jensen, 1998).

Table 4
Parameter estimates for two- and three-class solutions in subsample one.

	Two-class solution		Three-class solution		
	Low	High	Low	Mid	High
N	698	490	521	145	522
g variance	5.56*	2.61*	5.01*	3.75*	1.95*
(*SE*)	(0.72)	(1.29)	(0.64)	(0.79)	(0.64)
g mean	0.00	1.56*	0.00	1.25*	2.50*
(*SE*)		(0.67)		(0.46)	(0.36)
M of standardized residuals	0.46	0.61	0.47	0.60	0.60

Note. *M* = mean; *SE* = standard error; *$p<0.05$.

3.1.2. Three-class models

In general, based on the fit indices, the three-class models fit better than did the two-class models (see Table 3). According to the AIC and aBIC information criteria, the *g* variance model fit the best (with first-order residual variances constrained equal). Once again, the BIC supported the *g* means model.

Table 5
Fit indices for subsample two.

Model	Information criterion AIC	BIC	aBIC
Single class	84,864.4	85,153.9	84,972.8
Two-class			
g mean	84,797.6	85,102.3	84,911.8
Subtest residual	84,715.0	85,101.0	84,859.6
First-order residual	84,719.9	85,126.2	84,872.1
g variance	**84,698.2**	**85,089.2**	**84,844.6**
First-order residual + *g* variance	84707.4	85124.0	84863.5
Three-class			
g mean	84,801.6	85,116.5	84,919.6
Subtest residual	84,693.4	85,170.9	84,872.3
First-order residual	84,697.7	85,215.8	84,891.8
g variance	**84,680.5**	**85,168.1**	**84,863.2**
First-order residual + *g* variance	84,690.6	85,218.8	84,888.5

Note. Bolded numbers represent the lowest information criterion (IC) value in the two- and three-class models for each of the three ICs. In the first-order residuals + *g* model, *g* variances and first-order residual variances were free to vary across class.

The average latent class probabilities for most likely group membership in the respective class were 0.74(low), 0.69 (middle), and 0.71 (high). All of the parameter estimates were consistent with what would be predicted from SLODR: The larger the *g* mean (for which higher positive values represent higher levels of *g*), the smaller the *g* variance. The three-class *g* variance model estimates are shown next to the estimates for the two-class solution in Table 4. The findings from subsample one provided evidence of the presence of SLODR as well as demonstrated the potential for FMM to capture the phenomenon.

3.2. Subsample two

3.2.1. Two-class model

To serve as a validation of the results obtained using subsample one, the same analyses were performed with subsample two (see Table 5). The results from the analyses performed with subsample two were mostly consistent with those of subsample one. For this subsample, all fit indices, including the BIC, supported the two-class *g* variance model (with first-order residual variances constrained equal). This is the model that is most consistent with the presence of SLODR.

Table 6
Parameter estimates for two- and three-class solutions in subsample two.

	Two-class solution		Three-class solution		
	Low	High	Low	Mid	High
N	959	228	933	115	139
g variance	5.29*	1.90	5.41*	2.99	1.42
(SE)	(0.48)	(1.15)	(0.49)	(1.97)	(0.86)
g mean	0.00	1.56*	0.00	0.85	1.98*
(*SE*)		(*0.49*)		(*0.98*)	(*0.48*)
M of standardized residuals	0.46	0.61	0.46	0.60	0.67

Note. *M* = mean; *SE* = standard error; *$p<0.05$.

The average probabilities for most likely class membership were 0.84 (low ability) and 0.75 (high ability).Relevant parameter estimates are shown in Table 6. All findings were in the expected direction (assuming SLODR was present and FMM was a reasonable model through which to capture SLODR). In the low ability class the average proportion of subtest variance explained by *g* was 0.36 and by the broad abilities was 0.17. In the high ability class the average proportion of subtest variance explained by *g* was 0.12, while the average explained by the broad abilities was 0.20.The only other difference between the second and first subsample results was the proportional distribution of membership across the classes. There were more people in the low ability class in subsample two than there were in subsample one.

As with the first subsample, sex (χ^2 (1) = 0.41, p = 0.52) and age (χ^2 (12) = 8.72, p = 0.76) were not distributed significantly differently across classes. SES was distributed significantly differently across classes (χ^2 (3) = 18.45, $p<0.01$), and in the expected direction. The statistically significant correlation between SES and class was 0.11 (Kendall's' tau-*b* = 0.10).

3.2.2. Three-class model

In general, and as with the first subsample's results, the information criteria supported the fit of the three-class over the corresponding two-class models. Of the three-class models, the *g* variance model (with first-order residual variances constrained equal) fit the best. All of the parameter estimates' values were consistent with SLODR predictions, including a decrease in *g* variance associated with the class with a higher *g* factor mean. All of the model fit indices are presented in Table 5 and parameter estimates in Table 6. The average latent class probabilities for most likely group memberships in the respective class were 0.79 (low), 0.66 (middle), and 0.68 (high). As with the two-class models, one difference between the subsamples was the proportion of participants in each class. There were more people in the lowest ability class and fewer in the higher ability classes compared to subsample one. Taken together, the findings are mostly in support of the presence of SLODR using both subsamples.

3.3. FMMs with covariates

Results from post hoc analyses reported earlier suggested that age and sex were distributed evenly across the classes, but those of higher SES were more likely to be in the higher ability classes. Because age differentiation has been found to explain SLODR in some studies (Facon, 2004, 2008; Kane & Brand, 2006), additional analyses were performed where age was modeled as a

Table 7
Fit indices for models with age as a covariate.

Model	Information criterion AIC	BIC	aBIC
Two-class			
Within/Between	84,699.5	85,100.8	84,849.8
Within/Between & class-specific age effects	84,701.4	85,107.8	84,853.7

covariate in the two-class solution from subsample one. Age was included to explain differences between and within classes, therefore, the latent class variable was regressed on age, and within each class *g* was regressed on age (in years). Two models were specified, one allowing for class-specific age effects on *g* and one with the effects of age on *g* set equal across classes.

The information criteria are contained in Table 7. The model with age effects on *g* fixed equal across classes fit better than a model where class-specific effects were allowed. No statistically significant age effects, however, were found between or within classes. Age was not a significant predictor of class membership, and age did not explain a significant amount of variation in *g* within classes. These findings in addition to the post hoc class comparisons of age suggest that for this particular dataset our findings were not influenced by age differentiation. Such findings were also consistent with a previous study using these data where the covariance matrices were found to be invariant across age (Reynolds et al., 2007).

Last, additional models were tested using SES as a covariate. In the previous analysis without covariates the classes were found to differ in terms of SES (i.e., a greater number of higher SES participants were found in the higher ability group). Given the well-established relation between SES and *g* and the results from post hoc class membership analysis, it was reasonable to include SES as a covariate. Therefore, the latent class variable and the within-class *g* factor were regressed on SES. Models allowing for *g* to regress on SES within-class would not converge properly. SES, however, did significantly predict class membership such that those who were of higher SES were more likely to be in the high ability class. The average latent class probabilities for most likely class membership were improved in this model (0.93 for low ability and 0.94 for high ability).

4. Discussion

SLODR refers to the phenomenon that *g* loses some of its explanatory value and shows less variability relative to other cognitive abilities at higher levels of *g*. In this study, the presence of SLODR in the KABC-II norming sample data was investigated with factor mixture modeling (FMM). Factor mixture modeling was considered an improved model to investigate SLODR because groups did not have to be selected by researchers prior to the study and higher-order CFA models consistent with psychometric models of intelligence could be used to clarify the nature of SLODR.

The findings, which were mostly replicated across two subsamples, supported the presence of SLODR and the utility of FMM to investigate it. If SLODR was present and found, there would have been a set of predictable findings. The latent classes should have differed in the latent mean of *g* and *g* factor variance (and potentially first-order residual variances). A class with a higher *g* level should have been associated with less *g* variance in the unstandardized solution, which would have manifested as lower *g* loadings in the standardized solution. In addition, given a decreased proportion of subtest variance explained by *g* (or common factors in general), the standardized subtest residual variances should have been larger in the higher *g* class.

The results from this study were consistent with this set of predictions (except that the BIC for subsample one indicated the *g* variance model fit second best). For both subsamples, the latent classes differed in mean levels of *g*, and the higher ability class (higher on *g*) showed less variance in *g*. Comparing parameter estimates from the standardized solutions indicated that the proportion of subtest variance explained by the common factors was lower in the higher ability group (standardized subtests residual variances were greater). This decrease in proportion of subtest variance explained by the common factors was decomposed via the Schmid–Leiman transformation. The decrease corresponded to a decrease in the variance explained by *g*, and not the broad abilities (similarly model fit indices indicated the models with first-order residuals constrained equal across classes fit better). Put differently, the standardized *g* factor loadings were lower in the higher ability class.

Our findings related to SLODR (which is also referred to as ability differentiation) did not appear to be a result of age differentiation. Age was evenly distributed across classes, and age did not have an effect between or within classes. Moreover, previous research with these data indicated that the covariance matrices were invariant across age (Reynolds et al., 2007). Recent findings have suggested that age differentiation may have a role in SLODR (Facon, 2004, 2008; Kane & Brand, 2006), so future research will be needed to clarify the potential role of age in ability differentiation.

SES, on the other hand, was found to predict class membership. Children who had a parent with more education had a higher probability of being in the higher ability group. This finding was not surprising given the positive correlation found between SES of origin and *g*/IQ typically ranges from about 0.30 to 0.40 (see Jensen, 1998; Lubinski, 2004). Because SES tends to be a known source of heterogeneity in IQ test data, it may be worthwhile to include an SES variable in factor models so that more accurate parameter estimates are obtained.

In all, factor mixture modeling was a useful method to capture SLODR. Because higher-order factor models were used as within-class models, the SLODR-related "action" was better elucidated with regards to current conceptions of intelligence.

4.1. Latent class interpretation

Identification of a multiple-class model requires the interpretation of the latent categorical variable. There are generally two interpretations of latent classes using factor mixture models. One interpretation is that the classes are indeed qualitatively different groups of people, categories, or "clusters." As in a classic analogy explained by Meehl, there are gophers and chipmunks, each a class, but there are not gophermunks (1992). Here we do not view low and high ability classes as chipmunks and gophers (distinct classes), nor do we consider *g* to be a "low–high" dichotomy. The assumption of categorically different "classes" of people could potentially be dangerous if factor mixture models were used to formally classify individuals into "high" and "low" ability groups. Here, the different number of individuals found in classes varied across the subsamples. These classes, however, show potential for capturing and describing the phenomenon (see Johnson, Hicks, McGue, & Iacano, 2007).

An alternative interpretation of latent classes is that the classes represent some type of mixture component (Bauer, 2007; Dolan, 2009; Lubke & Spies, 2008). This interpretation is more plausible and fits nicely with the study of SLODR. Hence, FMM can be viewed as a method to capture non-normality in the distribution of the latent variables. In this study, the *g* variance difference across classes indicated a non-normal distribution of *g* (as predicted by SLODR). The *g* variance was higher in the low ability class compared to the high ability class. In the standardized solution this difference in variance showed up in the standardized *g* loadings. The loadings were lower in the high ability class, which was interpreted as *g* having less influence in this class. It should also be noted that the subtest residuals were not invariant across classes. Heteroscedastic residual variance, or residual variances that change depending on the level of the factor, have also been implicated as a possible cause of lower correlations in higher ability groups (Hessen & Dolan, 2009).

4.2. Implications

The findings from this study suggest that standardized *g* factor loadings depend on the level of the factor. This possibility is interesting because a long-standing criticism of factor analysis is that factor loadings are assumed to be static, or represent an average loading, across individuals (Wolfle, 1940). SLODR-related findings imply that there is a systematic pattern related to standardized factor loadings: *g* factor loadings decrease as a function of *g*.

One method to allow loadings to vary across groups is multiple group factor analysis (Jöreskog, 1970). For example, high and low ability groups could be formed based on *g* factor scores, and the *g* loadings could be compared across groups. Factor mixture models are considered to be an alternative to multiple group models in that the groups are unknown and thus do not require the researcher to create ability groups prior to the study, allowing for error in selection of group membership. Factor mixture modeling and multiple group factor analysis (groups were split into low and high ability groups on the basis of *g* factor scores; Reynolds & Keith, 2007) have been performed on the KABC-II norming data using identical CFA models. Both techniques supported the presence of SLODR. Using both techniques, the average standardized *g* loadings (obtained via a Schmid–Leiman transformation) were lower in the high ability groups suggesting that loadings may be a function of the general factor. Thus, the common assumption of fixed factor loadings, as questioned by Wolfle (1940) and Carroll (1993), may well be an erroneous one in intelligence test data.

One practical implication of SLODR is related to the interpretation of intelligence test scores. Findings from SLODR indicate that the explanatory power of the general factor decreases at higher levels of ability. If measurement issues at the item level can be ruled out as a cause (most modern test IQ test publisher put items through rigorous analyses), then this finding, in turn, suggests that broad and specific abilities are relatively more important for people with high IQ scores. Perhaps more interpretative weight should be given to those broad and specific abilities for those with higher IQs. Alternatively, other influences such as personality or interests may become more important for those of higher ability, and such possibilities should be tested in future research. Research on *g* and its important correlates may also need to consider the possibility that *g* may become less important with higher levels of *g*. Is it the case, for example, that *g* also has smaller effects on other outcomes depending on one's level of *g*? It may be, for example, that *g* likewise has smaller effects, relative to the specific abilities, on achievement and other academic outcomes for those with high levels of *g*. Clearly understanding SLODR is important in understanding *g* and may have important applied implications.

4.3. Limitations

The presence of SLODR should continue to be investigated in datasets with representative samples and an adequate number of subtests. Our sample included children and adolescents, so we are limited in our generalizations. FMM should be considered as a method to investigate SLODR in other intellectual measurement batteries, especially with batteries that include adult samples. In addition, alternative model specifications should be tested (e.g., factor loadings and intercepts). In this study, we maintained equality constraints on the intercepts to make sure differences in the class levels were a function of the *g* factor. Researchers may want to investigate the validity of constraining intercepts equal across classes (see Reynolds & Keith, 2007 for tests of intercept constraints using a high and low ability multi-group model with these KABC-II data).

Last, differences in the class membership proportions were found across the subsamples. There was a higher proportion of individuals in the low ability class in subsample two than there was for the corresponding class estimated using subsample one. Note that in the two-class samples the *g* mean differences were similar, suggesting the difference may be related to sampling issues. Future research may be needed to address this issue in more detail.

4.4. Summary

Factor mixture modeling was used to capture and describe SLODR in a higher-order model of intelligence using the norming sample from the KABC-II. The results are consistent with predictions based on what would be expected if SLODR was present. FMM was useful in capturing and describing the phenomenon. Rather than the traditional CFA model (i.e., one-class model) the KABC-II data were best described by a two- or three-class confirmatory factor model with class-specific *g* means, *g* variances, and subtest residuals. The higher ability class was characterized by less *g* factor variance, which showed up as lower standardized *g* factor loadings in the standardized solution, supporting the role of *g* in this phenomenon. Researchers may want to consider the use of factor mixture modeling as a novel method that obviates the problems that come with selecting ability groups prior to analyzing SLODR. If SLODR is found in other popular individual intelligence test batteries, then applied implications such as those related to interpretation of test scores and *g* factor loadings need to be included in SLODR research.

Appendix A

Subsample one	Low ability class			High ability class		
Subtest	*N*	*M*	*SD*	*N*	*M*	*SD*
Atlantis delayed	690	9.49	3.09	484	10.56	2.47
Atlantis	698	9.32	3.28	490	10.88	2.62
Block counting	698	9.07	3.03	490	10.76	2.66
Expressive vocabulary	516	9.07	2.80	366	10.83	2.63
Gestalt closure	516	9.59	2.92	366	10.85	2.94
Hand movements	698	9.42	3.03	490	10.89	2.55
Number recall	698	9.52	2.93	490	10.78	2.72
Pattern reasoning untimed	681	8.91	2.88	476	11.05	2.82
Rebus delayed	681	9.29	2.88	476	11.10	2.62
Riddles	698	9.16	3.03	490	11.17	2.73
Rebus	698	9.36	2.93	490	11.15	2.91
Rover	698	9.40	2.88	490	10.95	2.86
Story completion untimed	698	9.31	3.01	490	10.81	2.43
Triangles untimed	698	9.31	3.17	490	10.87	2.18
Verbal knowledge	698	9.10	2.93	490	11.11	2.90
Word order	698	9.11	2.65	490	10.93	2.99
Subsample two	Low ability class			High ability class		
Subtest	*N*	*M*	*SD*	*N*	*M*	*SD*
Atlantis delayed	949	9.57	2.95	223	11.01	2.37
Atlantis	959	9.61	3.20	228	11.38	2.46
Block counting	959	9.90	3.04	228	10.72	2.88
Expressive vocabulary	959	9.64	2.91	228	11.12	2.92
Gestalt closure	695	9.93	3.05	161	10.85	2.58
Hand movements	959	9.81	2.81	228	10.38	2.89
Number recall	959	9.97	2.93	228	10.88	2.47
Pattern reasoning untimed	959	9.57	2.81	228	11.30	3.26
Rebus delayed	934	9.96	3.00	220	11.35	2.78
Riddles	959	9.79	3.04	228	11.64	2.81
Rebus	959	9.71	2.97	228	11.63	3.00
Rover	959	9.72	2.98	228	11.24	2.76
Story completion untimed	959	9.64	2.82	228	10.91	2.56
Triangles untimed	959	9.84	2.88	228	10.99	2.39
Verbal knowledge	959	9.71	2.92	228	11.36	2.77
Word order	959	9.59	2.60	228	11.08	3.33

References

Akaike, H. A. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317—332.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies.* New York: Cambridge University Press.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*, 513—535.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757—786.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, *36*, 589—609.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249—253.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16—29.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 349—359.

Detterman, D. K. (1999). The psychology of mental retardation. *International Review of Psychiatry*, *11*, 26—33.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349—359.

Dolan, C. V. (2009). Structural equation mixture modeling. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 568—591). Thousand Oaks, CA: SAGE.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227—253.

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence*, *32*, 391—401.

Facon, B. (2008). How does the strength of the relationships between cognitive abilities evolve over the life span for low-IQ vs. high-IQ adults? *Intelligence*, *36*, 339—349.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79—132.

Hartmann, P. H., & Reuter, M. (2006). Spearman's "Law of Diminishing Returns" tested with two methods. *Intelligence*, *34*, 47—62.

Hartmann, P. H., & Teasdale, T. W. (2004). A test of Spearman's "law of diminishing returns" in two large samples of Danish military draftees. *Intelligence*, *32*, 499—508.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modelling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*, 202—226.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hessen, D. J., & Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood. *British Journal of Mathematical and Statistical Psychology*, *62*, 57—77.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39—59.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Prager.

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence*, *31*, 95—105.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393—416.

Johnson, W., Hicks, B. M., McGue, M., & Iacano, W. G. (2007). Most of the girls are alright, but some aren't: Personality trajectory groups from ages 14 to 24 and some associations with outcomes. *Journal of Personality and Social Psychology*, *93*, 266—284.

Jöreskog, K. (1970). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409—426.

Kane, H. D., & Brand, C. R. (2006). The variable importance of general intelligence (*g*) in the cognitive abilities of children and adolescents. *Educational Psychology*, *26*, 751—767.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children-Second Edition: Technical manual.* Circle Pines, MN: American Guidance Service.

Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, *23*, 45—57.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) General intelligence, objectively determined and measured. *Journal of Personality and Social Psychology*, *86*, 96—111.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21—39.

Lubke, G. H., & Spies, J. R. (2008). Choosing a "correct" factor mixture model: Power, limitations, and graphical data exploration. In G. R. Hancock & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 343—362). Charlotte, NC: Information Age.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19—40.

Maxwell, A. E. (1972). The WPPSI: A marked discrepancy in the correlations of subtests for good and poor readers. *British Journal of Mathematical and Statistical Psychology*, *25*, 283—291.

Meehl, P. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, *60*, 117—174.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525—543.

Muthén, B. O. (1989). Factor structure in groups selected on observed scores. *British Journal of Mathematical and Statistical Psychology*, *42*, 81—90.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide. Fifth edition.* Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R., & Thompson, W. W. (1995). Selection and related threats to group comparisons: An example comparing factorial structures of higher and lower ability groups. *Psychological Bulletin*, *117*, 271—284.

Reynolds, M. R., & Keith, T. Z. (2007). A test of Spearman's law of diminishing returns in hierarchical models of intelligence. *Intelligence*, *35*, 267—281.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. F., & Low, J. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second Edition: Consistency with Cattell–Horn–Carroll theory. *School Psychology Quarterly*, *4*, 511—539.

Saklofske, D. H., Yang, Z., Zhu, J., & Austin, E. L. (2008). Spearman's law of diminishing returns in normative samples for the WISC-IV and WAIS-III. *Personality and Individual Differences*, *29*, 57—69.

Schwarz, G. (1978). Estimating the dimensions of a model. *The Annals of Statistics*, *6*, 461−464.
Sclove, S. L. (1987). Applications of some model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333−343.
Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.
te Nijenhuis, J., & Hartmann, P. (2006). Spearman's "law of diminishing returns" in samples of Dutch immigrant children and adults. *Intelligence*, *34*, 437−447.
Wolfle, D. (1940). *Factor analysis to 1940.* Chicago: The University of Chicago Press.
Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297−330.