Psychology Science Quarterly, Volume 50, 2008 (2), pp. 283-294

# Does intellectual giftedness affect the factor structure of divergent thinking? Evidence from a MG-MACS analysis

HEINZ HOLLING[1] & JÖRG-TOBIAS KUHN

## Abstract

This study explored the latent structure of divergent thinking as a cognitive ability across gifted and non-gifted samples of students utilizing multiple-group analysis of mean and covariance structures (MG-MACS). Whereas Spearman's law of diminishing returns postulates lower *g* saturation of cognitive tests with increasing ability level and consequently, a lower correlation of cognitive abilities in more gifted samples, recent evidence from creativity research has shown that correlations of divergent thinking with intelligence are unaffected by ability level. In order to investigate this conflicting state of affairs with respect to divergent thinking, we utilized increasingly restrictive MG-MACS models that were capable of comparing latent variances, covariances, and means between gifted (IQ > 130) and non-gifted (IQ ≤ 130) groups of students. In a sample of 1070 German school students, we found that a MG-MACS model assuming partial strict measurement invariance with respect to the postulated factor model of verbal, figural, and numerical divergent thinking could not be rejected. Further, latent variances and covariances of latent divergent thinking factors did not significantly differ between groups, whereas the gifted group exhibited significantly higher latent means. Finally, implications of our results for future research on the latent structure of divergent thinking are discussed.

Key words: Giftedness; Confirmatory factor analysis

[1] Heinz Holling, Psychological Department IV, University of Muenster, Fliednerst. 21, 48149 Muenster, Germany; Phone: ++49 (0) 251-833 84 85, fax: ++49 (0) 251-833 94 69, email: holling@psy.uni-muenster.de

## Introduction

Many structural theories of intelligence incorporate a factor corresponding to creativity (e.g., Carroll, 1993; Jäger, 1984). Divergent thinking (DT), which has been defined as the capability to generate diverse and numerous ideas (Runco, 1991), can be considered as the core ability for creative achievements. In a classical article, Guilford (1950) identified three basic components as factors of DT: Fluency (the total number of ideas generated), flexibility (the number of categories in the ideas) and originality (the number of unique or unusual ideas). However, fluency is usually described as the central component of DT (Hargreaves & Bolton, 1972). In contrast to research on intelligence, DT tests reported in the literature focus on verbal or figural content, thereby neglecting the numerical domain (Cropley, 2000). However, numerical content plays an important role in research on reasoning and problem-solving, where DT is often of central importance (Mumford, Connelly, Baughman, & Marks, 1994). Further, Livne and Milgram (2006) have shown that DT is one important facet of mathematical achievement. Hence, an investigation of numerical DT, and its relationship with verbal and figural DT, seems necessary to elucidate the factorial structure of DT.

The concept of intellectual giftedness has been defined in different ways across the literature. Some approaches (e.g., Roznowski, Reith, & Hong, 2000) focus exclusively on high intellectual ability (*g*) as the sole determinant of intellectual giftedness, while others (e.g., Lubinski & Benbow, 2000) perceive giftedness as being multidimensional in nature. The role of creativity (and hence, DT) in models of giftedness varies as well, where some models of intellectual giftedness perceive creativity as a condition sine qua non for outstanding intellectual achievement (Renzulli, 1986), while others perceive creativity as an own form of giftedness (Gagné, 1993). Similar to Roznowski et al. (2000), we take a one-dimensional perspective on intellectual giftedness in this paper, in that subjects with a high level of fluid intelligence are defined as being intellectually gifted. Further, we assume that DT, as a core trait of creative performance, can be conceptualized as a latent cognitive ability that is part of a cognitive taxonomy (Carroll, 1993).

The empirical relationship of DT with intelligence has been intensively researched over the years (cf. Haensly & Reynolds, 1989; Sternberg & O'Hara, 1999). Kim (2005) recently conducted a meta-analysis, using 21 studies with 447 correlation coefficients and 45,880 participants overall. He found the average correlation between DT tests and intelligence to be $r = .17$. This relationship was moderated by several factors, however. For example, DT-intelligence correlations were significantly lower for younger students compared to older participants. Empirically, therefore, small to moderate correlations between intelligence and DT have been found.

A growing field of research pertains to the comparison of latent model structures across populations of varying levels of intelligence. For example, Spearman's "law of diminishing returns" (Spearman, 1927) basically states that the *g*-saturation of cognitive ability tests decreases with an increase of the subject's ability level. Conceptually, this implies a higher correlation of latent cognitive abilities at the lower end of the ability distribution than at the higher end. A large body of empirical evidence supports Spearman's assumption (e.g., Hartmann & Teasdale, 2004; Reynolds & Keith, 2007). In creativity research, a conceptually related theory ("threshold theory") has been proposed, which assumes that the correlation of intelligence with creativity is lower for subjects with higher intelligence (IQ > 120; cf. Barron, 1963). However, in the meta-analysis by Kim (2005), no support for the threshold theory was found. This result is in line with recent studies that were unable to find empirical support for the threshold theory (Preckel, Holling, & Wiese, 2006; Sligh, Conners, & Roskos-Ewoldsen, 2005). In fact, Sligh et al. (2005) even report an inverse threshold theory effect, implying a larger correlation of intelligence and creativity in the high ability group.

The inconsistencies between empirical results pertaining to the law of diminishing returns and the threshold theory are striking if DT is conceptualized as a latent ability that is part of a cognitive taxonomy of abilities (Carroll, 1993). However, this contradiction can partly be explained by the differing methodological analysis strategies between the intelligence and creativity research communities. Whereas many creativity researchers utilize zero-order correlations to analyse threshold theory, current intelligence research usually applies sophisticated latent variable modelling strategies that provide more thorough results. An important assumption for comparing observed test scores as well as correlations of latent abilities across groups is that measurement invariance (MI) holds. MI implies both that observed test scores have the same meaning across groups and that the postulated latent model structure is equivalent (Gregorich, 2006; Vandenberg & Lance, 2000). Assessing the degree of MI therefore helps to decide whether differences in observed test scores are equally attributable to latent abilities, or whether they are grounded in unrelated sources of variance (Wicherts, Dolan, & Hessen, 2005). Hence, in order to provide sound and unbiased evidence for either threshold theory or the law of diminishing returns, it is imperative to investigate MI across the compared groups. Only if MI is given can observed test scores be unequivocally interpreted and correlations of latent abilities be meaningfully compared.

Whereas MI assessments have become routine in intelligence research for the investigation of Spearman's law of diminishing returns, creativity researchers have largely refrained from analysing MI with respect to threshold theory. The only study that tested MI in creativity research we are aware of was conducted by Kim, Crammond, and Bandalos (2006), but these authors did not compare differentially gifted groups with respect to their DT performance. A study that systematically investigates MI and compares correlations of latent abilities across differentially gifted groups therefore seems lacking.

The purpose of the present study, therefore, was to compare the latent structure of verbal, figural, and numerical DT abilities, respectively, across levels of giftedness. Specifically, the main focus of this paper was an investigation of correlations of latent DT factors using MG-MACS (multiple-group means and covariance structures) analysis. In line with the methodologically more sound results from intelligence research pertaining to Spearman's law of diminishing returns compared to those relating to threshold theory, we hypothesized that correlations of latent DT abilities were lower in highly able subjects. Because it is a methodological prerequisite for comparing latent correlations, we also investigated MI across groups. Finally, we compared latent DT means between groups, hypothesizing that only negligible differences would exist because of the low correlations between DT and intelligence reported in the literature.

## Method

### *Sample*

The data used were a subsample from the standardization sample of the Berliner Intelligenzstruktur-Test für Jugendliche: Begabungs- und Hochbegabungsdiagnostik (BIS-HB [Berlin Structure-of-Intelligence test for Youth: Diagnosis of Talents and Giftedness]; Jäger et al., 2005). In this sample, a total of $n$ = 1328 students were tested (728 males and 598 females, two participants gave no information concerning gender). Mean age of the sample was 14.5 years ($s$ = 1.1 years). A subsample ($n$ = 1070) was additionally administered the short form of the German adaptation of the Culture Fair Intelligence Test (Cattell & Cattell, 1960) in order to assess IQ. Because we wanted to compare highly gifted students to a population with normal intelligence, we split this subsample into a gifted group (IQ > 130) and a non-gifted group (IQ ≤ 130). The gifted sample consisted of 250 students (mean IQ = 137, $s$ = 5.2), whereas the non-gifted sample comprised 820 individuals (mean IQ = 112, $s$ = 10.7). The large amount of gifted students in the sample can be explained by the fact that the BIS-HB offers specific testnorms for gifted students.

### *Measures*

The BIS-HB contains 12 DT tests overall, with each falling into one of three content domains (verbal, figural, and numerical, respectively). Similar to the assumption of Carroll (1993), in the BIS model, DT is seen as a cognitive ability that represents a facet of cognitive functioning. Table 1 provides an overview of the 12 DT tests investigated in this study.

**Table 1:**
Description of the 12 DT tests of the BIS-HB

| Name | Abbreviation[1] | Description |
|---|---|---|
| Symbol completion | ZF (F) | Drawing pictures from pre-specified objects |
| Symbol combining | ZK (F) | Combination of geometric objects into different figures |
| Object design | OJ (F) | Conversion of geometric figures into real objects |
| Layout | LO (F) | Designing logotypes |
| Specific traits | EF (V) | Listing traits a salesman should not have |
| Possible object use | AM (V) | Listing uses for given objects |
| Masselon | MA (V) | Building sentences as possible from three given words |
| Insight Test | IT (V) | Giving explanations for a presented social situation |
| Puzzles with numbers | ZR (N) | Inventing patterns of numbers according to rules |
| Divergent computing | DR (N) | Finding numbers that produce a given sum |
| Equations with numbers | ZG (N) | Producing equations from given numbers and operations |
| Inventing telephone numbers | TN (N) | Inventing telephone numbers according to pre-specified rules |

*Note.* [1]Content domain given in brackets: F = figural, V = verbal, N = numerical.

Each DT test was administered with a pre-specified time-limit and scored for fluency. The unadjusted intraclass-correlation-coefficient as a measure of objectivity of scoring between the ratings of two independent raters showed satisfactory values for all DT tests ($M$ = .94, $s$ = .04). The overall internal consistency (Cronbach's alpha) over the twelve DT tests scored for fluency was found to be satisfactory (α = .84; verbal DT: α = .76; figural DT: α = .65; numerical DT: α = .60).

## *Statistical modelling procedures*

Establishing MI requires fitting a sequence of nested, increasingly restricted CFA models (Vandenberg & Lance, 2000). MI is best assessed using MG-MACS (multiple-group analysis of mean and covariance structures), because this approach allows all central aspects of MI to be statistically tested. The MIMIC model (multiple-indicator multiple-cause; Muthén, 1989) can be used to assess specific aspects of MI as well, but this model assumes all factor loadings and residual variances to be constant across groups, an assumption that can be explicitly tested with MG-MACS. We therefore utilized MG-MACS in this study.

Let $\mathbf{y}_{ig}$ denote the observed $p \times 1$ random vector of subject $i$, $i = 1, \ldots, n_g$ in group $g$. The following model for $\mathbf{y}_{ig}$ was specified:

$$\mathbf{y}_{ig} = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_{ig} + \boldsymbol{\varepsilon}_{ig}, \tag{1}$$

where $\boldsymbol{\nu}_g$ is a $p \times 1$ vector of measurement intercepts, $\boldsymbol{\Lambda}_g$ is a $p \times q$ matrix of factor loadings ($q < p$), $\boldsymbol{\eta}_{ig}$ is a $q \times 1$ vector of common factor scores and $\boldsymbol{\varepsilon}_{ig}$ is a $p \times 1$ vector of residuals. Further, the observed $p \times p$ variance-covariance matrix $\boldsymbol{\Sigma}_\mathbf{g}$ can be decomposed as

$$\boldsymbol{\Sigma}_\mathbf{g} = \boldsymbol{\Lambda}_\mathbf{g} \boldsymbol{\Psi}_\mathbf{g} \boldsymbol{\Lambda}'_\mathbf{g} + \boldsymbol{\Theta}_\mathbf{g}, \tag{2}$$

where $\boldsymbol{\Psi}_\mathbf{g}$ is a $q \times q$ vector of latent factor (co-)variances, and $\boldsymbol{\Theta}_g$ represents a diagonal $p \times p$ matrix of residual variances.

In this study, several increasingly restrictive forms of MI were investigated. The relevant MG-MACS models were nested and therefore could be compared using likelihood-ratio tests. *Configural MI* requires equal patterns of factor loadings across groups, constraining construct dimensionality to be equivalent. *Metric MI* constrains factor loadings to be invariant ($\Lambda_g = \Lambda$). A further model additionally assumes *equal residual variances* across groups ($\Lambda_g = \Lambda$, $\Theta_g = \Theta$). The next two models impose restrictions on the measurement intercept structure: *Strong MI* assumes factor loadings and intercepts to be invariant across groups ($\Lambda_g = \Lambda$, $\nu_g = \nu$), whereas *strict MI* imposes constraints on intercepts, factor loadings and residuals ($\Lambda_g = \Lambda$, $\Theta_g = \Theta$, $\nu_g = \nu$). Further, in order to evaluate our hypothesis pertaining to the equality of latent correlations across groups, we specified a sixth model that kept latent (co-)variances constant across groups ($\Lambda_g = \Lambda$, $\Theta_g = \Theta$, $\nu_g = \nu$, $\boldsymbol{\Psi}_\mathbf{g} = \boldsymbol{\Psi}$). This model was termed *strict MI with equal latent (co-)variances*. Finally, we fitted a model that constrained the vector of latent factor means to be equal across groups in order to statistically test whether latent mean differences existed between groups ($\Lambda_g = \Lambda$, $\Theta_g = \Theta$, $\nu_g = \nu$, $\boldsymbol{\Psi}_\mathbf{g} = \boldsymbol{\Psi}$, $\boldsymbol{\alpha}_\mathbf{g} = \boldsymbol{\alpha}$). Figure 1 illustrates the model investigated in this study. We did not test a second-order model here, because with three latent factors, the model is underidentified for multiple-group comparisons (Chen, Sousa, & West, 2005).

According to Meredith (1993), strict MI is required for unequivocally ascribing observed mean differences to latent mean differences. However, Byrne, Shavelson, and Muthén (1989) have argued that releasing selected measurement intercepts still allows for latent mean comparisons (cf. Thompson & Green, 2006). In line with these authors, we will call a model that holds all factor loadings and residual variances, but only selected measurement intercepts fixed across groups a *partial strict MI* model. We regard strict MI as a desideratum for latent mean comparisons, albeit we assume that comparing latent mean differences under partial strict MI is possible. In order to standardize latent mean comparisons, we computed effect sizes of latent mean differences (Hancock, 2001).

The following indices were used for assessing model fit (abbreviations given in brackets): (1) the Satorra-Bentler $\chi^2$-statistic (SB-$\chi^2$; Satorra & Bentler, 2001); (2) the Comparative Fit Index (CFI); (3) the Bayesian information criterion (BIC); (4) the Root Mean Square Error of Approximation (RMSEA), adjusted for multiple groups; (5) the $\chi^2$-difference test; and (6) Steiger's (1989) gamma hat ($\hat{\gamma}$), as well as McDonald's (1989) noncentrality index (Mc), both of which are sufficiently robust in MI evaluation (Cheung & Rensvold, 2002). CFI values close to .95 and RMSEA values close to .06 indicate an adequate fit (Hu & Bentler, 1999).

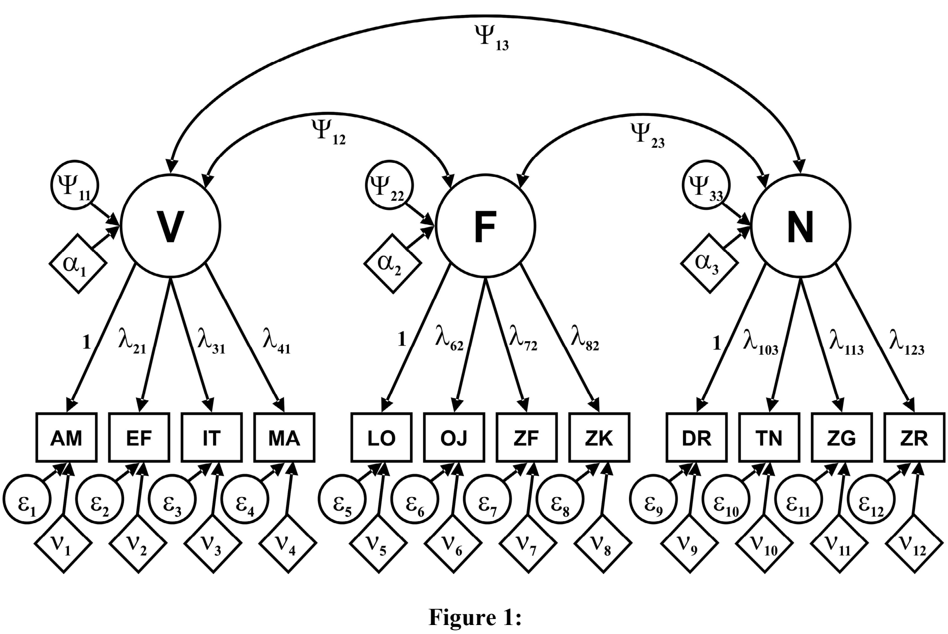


**Figure 1:**
Postulated factor structure of the 12 BIS-HB DT tests.
V = Verbal DT factor, F = Figural DT factor, N = numerical DT Factor

## Results

Descriptive statistics of the 12 DT tests investigated in this study can be found in Table 2. As can be seen, Cohen's *d* (Cohen, 1988) indicated small to medium standardized mean differences between groups with respect to the observed means. However, these results are only meaningfully interpretable after MI has been established.

Except for DT test Masselon, univariate skewness of all DT tests significantly deviated from normality. In addition, univariate kurtosis of six DT tests revealed significant departure from normality as well. As could be expected, statistical tests indicated the absence of multivariate normality (Mardia's multivariate skewness: $b_{1,12} = 5.83$, $A = 1040.47$, $p < .01$; Mardia's multivariate kurtosis: $b_{2,12} = 184.44$, $Z = 14.67$, $p < .01$). Therefore, all standard errors, fit indexes as well as $\Delta$SB-$\chi^2$ were rescaled utilizing the procedure outlined in Satorra and Bentler (2001).

The results of all MI comparisons can be found in Table 3. As can be seen, no significant deterioration in model fit occurs up to Model 3, assuming metric MI as well as equal error variances. However, strong as well as strict MI were not tenable here (Models 4 and 5, respectively). We therefore investigated the univariate Lagrange multiplier tests (LM) that reveal information as to where model misspecification might have occurred. We found that the latent intercepts of three DT tests varied across groups, with LM = 17 for divergent computing, LM = 11 for symbol combining and LM = 7 for Masselon. Expected parameter change indexes revealed that the latent intercepts in the gifted group were significantly higher than for the normal group. Hence, we freed these intercepts and estimated a model representing partial strict MI (Model 6), which did not significantly deviate from Model 3. In a next step, we compared a model that additionally constrained latent variances and covariances to be equal across groups (Model 7) with the partial strict MI model. No significant difference was found, indicating that latent variances as well as covariances can be assumed to be equal across groups. In a final step, we compared a model constraining latent means to be equal across groups (Model 8) with Model 7, and found a significant deterioration in model fit, indicating that the latent means differed significantly between groups. In order to quantify these differences, we computed standardized effect sizes (ES; Hancock, 2001) that are comparable in meaning to Cohen's *d*. *ES* indicated large latent mean differences in favor of the gifted group, *ES*(verbal) = .69, *ES*(figural) = .66, *ES*(numerical) = .50. Hence, gifted individuals showed significantly better DT skills.

Finally, Figure 2 reveals the standardized factor loadings and factor correlations of Model 7, where these values are constrained to be equal across the gifted and non-gifted group. As can be seen, the latent DT factors correlated substantially, especially the verbal and figural DT factor. Figure 2 shows only the latent means of the gifted group because all latent means were fixed to zero in the non-gifted group for comparison purposes (cf. Thompson & Green, 2006).

**Table 2:**
Means, standard deviations and Cohen's *d* of DT tests

| DT Test | Abbr. | Domain[1] | Subsample | | *d* |
|---|---|---|---|---|---|
| | | | Normal | Gifted | |
| Symbol completion | ZF | F | 5.93 (2.33) | 6.66 (2.40) | 0.31 |
| Symbol combining | ZK | F | 4.98 (2.17) | 5.87 (2.06) | 0.42 |
| Object design | OJ | F | 5.68 (2.07) | 6.32 (1.96) | 0.32 |
| Layout | LO | F | 3.97 (1.56) | 4.34 (1.51) | 0.24 |
| Specific traits | EF | V | 7.35 (3.60) | 8.39 (3.52) | 0.29 |
| Possible object use | AM | V | 6.12 (2.87) | 7.38 (2.78) | 0.45 |
| Masselon | MA | V | 3.79 (1.21) | 4.32 (1.28) | 0.43 |
| Insight Test | IT | V | 9.62 (3.84) | 10.85 (4.04) | 0.31 |
| Puzzles with numbers | ZR | N | 4.48 (1.92) | 4.67 (1.82) | 0.10 |
| Divergent computing | DR | N | 8.66 (3.22) | 10.13 (3.37) | 0.45 |
| Equations with numbers | ZG | N | 6.62 (2.92) | 6.95 (2.74) | 0.12 |
| Inventing telephone numbers | TN | N | 9.03 (3.17) | 9.53 (3.15) | 0.16 |

*Note*. Standard deviations are given in brackets. [1]F = figural, V = verbal, N = numerical.

**Table 3:**
Fit indices for MI model comparisons across groups

| Model | EQC[a] | *df* | SB-$\chi^2$ | Compare | $\Delta$SB-$\chi^2$ | CFI | BIC | $\hat{\gamma}$ | Mc | RMSEA (95 % CI) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 102 | 242.862 | | | .949 | 56956 | 0.979 | 0.94 | 0.051 (0.041-0.061) |
| 2 | $\Lambda$ | 111 | 253.337 | 2 vs. 1 | 10.03 | .949 | 56904 | 0.978 | 0.94 | 0.049 (0.039-0.058) |
| 3 | $\Lambda, \Theta$ | 123 | 265.791 | 3 vs. 2 | 9.86 | .948 | 56838 | 0.978 | 0.94 | 0.047 (0.037-0.056) |
| 4 | $\Lambda, \nu$ | 120 | 298.136 | 4 vs. 2 | 43.72*) | .936 | 56888 | 0.973 | 0.92 | 0.053 (0.044-0.062) |
| 5 | $\Lambda, \Theta, \nu$ | 132 | 309.346 | 5 vs. 3 | 42.51*) | .936 | 56821 | 0.973 | 0.92 | 0.050 (0.041-0.059) |
| 6 | $\Lambda, \Theta, \nu^b$ | 129 | 274.507 | 6 vs. 3 | 8.63 | .947 | 56804 | 0.978 | 0.93 | 0.046 (0.037-0.055) |
| 7 | $\Lambda, \Theta, \nu^b, \Psi$ | 135 | 275.427 | 7 vs. 6 | 0.92 | .949 | 56764 | 0.979 | 0.94 | 0.044 (0.035-0.053) |
| 8 | $\Lambda, \Theta, \nu^b, \Psi, \alpha$ | 138 | 308.728 | 8 vs. 7 | 33.30*) | .938 | 56779 | 0.974 | 0.92 | 0.048 (0.040-0.057) |

*Note*. [a]Equality constraint across groups, [b]intercepts of three DT tests (DR, ZK, MA) were allowed to vary across groups. *) $p < .01$.

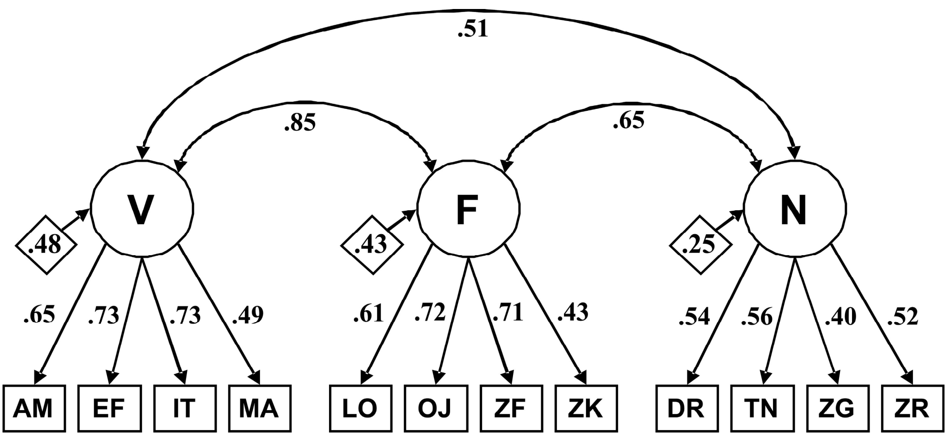


**Figure 2:**
Standardized loadings, correlations and latent means of Model 7.
Latent means are given for the gifted group only (see text)

## Discussion

As stated in the outline of the paper, conflicting evidence in the literature pertaining to the relationship of latent cognitive variables across subgroups have been reported, either in support of a lower relationship between cognitive abilities in populations of higher intelligence (Spearman's law of diminishing returns) or against such an assumption (as in the meta-analysis by Kim, 2005). In order to unequivocally assess the relationship of verbal, numerical, and figural DT across groups of differing giftedness, we utilized MG-MACS models that allow the successive statistical testing of model assumptions. We found that concerning DT, the postulated measurement structure of latent abilities remains largely equal across groups of differing abilities. That is, with the exception of three latent intercepts, all other model parameters (excluding latent means), i.e. factor loadings, error variances, factor variances and covariances as well as nine latent intercepts were found to be equal. The domain-specific latent structure of DT postulated in this model therefore has been shown to be remarkably stable. Further, in contrast to most evidence reported in the literature, we found large latent mean differences between the two groups investigated. That is, gifted students showed better DT skills in all content domains, especially with respect to verbal DT. A reason for this finding might reside in the fact that the DT tests applied here were administered with a time-limit, therefore introducing speededness into the tests, which might have been an advantage for the gifted students. However, most psychometric tests need time-limits due to administrative reasons. The finding that latent variances were equal for gifted and non-gifted samples in this study underlines the fact that DT shows sufficient variability even in gifted subjects.

The fact that latent correlations of DT factors are equal across groups differing in giftedness contrasts the vast literature in support of Spearman's law of diminishing returns. It might be presumed that a reason for this finding resides in the nature of DT as a cognitive ability that is less g-loaded than other cognitive tests. For example, in Carroll's (1993) taxonomy of cognitive abilities, the factor representing creativity is assumed to be much less dominated by the *g* factor than a factor representing fluid intelligence. However, other work utilizing parcelling strategies (e.g., Süß & Beauducel, 2005) has shown that DT shows substantial loadings on the *g* factor. Apparently, DT as operationalized in our study is less vulnerable to the law of diminishing returns than cognitive abilities in the domain of fluid intelligence.

To conclude, we did not find large qualitative differences in latent DT structures between differentially gifted groups concerning DT, and gifted students showed significantly better DT skills than a normally-gifted comparison group. In order to corroborate our findings across different DT tests, additional research is required.

## References

Barron, F. (1963). *Creativity and psychological health*. Princeton, NY: Van Nostrand.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456-466.

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

Cattell, R. B., & Cattell, A. K. S. (1960). *Culture Fair Intelligence test: Scale 2.* Champaign, IL: Institute for Personality and Ability Testing.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order models. *Structural Equation Modeling, 12,* 471-492.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233-255.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review, 23,* 72-79.

Gagné, F. (1993). Constructs and models pertaining to exceptional human abilities. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 69-87). Oxford: Pergamon.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44,* S78-S94.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5,* 444-454.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66,* 373-388.

Haensly, P. A., & Reynolds, C. R. (1989). Creativity and intelligence. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 111-132). New York: Plenum.

Hargreaves, D. J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology, 63,* 451-462.

Hartmann, P., & Teasdale, J. W. (2004). A test of Spearman's „law of diminishing returns" in two large samples of Danish military draftees. *Intelligence, 32*, 499-508.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Research on the structure of intelligence: Competing models, new developments, perspectives]. *Psychologische Rundschau, 35,* 21-35.

Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H.-M., & Beauducel, A. (2005). *Berliner Intelligenzstruktur-Test für Jugendliche: Begabungs- und Hochbegabungsdiagnostik (BIS-HB)* [Berlin Structure-of-Intelligence test for Youth: Diagnosis of Talents and Giftedness]. Göttingen, Germany: Hogrefe.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *The Journal of Secondary Gifted Education, 16,* 57-66.

Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking – Figural. *Educational and Psychological Measurement, 66,* 459-477.

Livne, M. L., & Milgram, R. M. (2006). Academic versus creative abilities in mathematics: Two components of the same construct? *Creativity Research Journal, 18*, 199-212.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist, 55*, 137-150.

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

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97-103.

Mumford, M. D., Connelly, M. S., Baughman, W. A., & Marks, M. A. (1994). Creativity and problem solving: Cognition, adaptability, and wisdom. *Roeper Review, 16*, 241-246.

Muthén, B. O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika, 54,* 557-585.

Preckel, F., Holling, H., & Wiese, M. (2006). Relationship of intelligence and creativity in gifted and non-gifted students: An investigation of threshold theory. *Personality and Individual Differences, 40*, 159-170.

Renzulli, J. S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R. Sternberg & J. Davidson (Eds.), *Conceptions of giftedness* (pp. 332-357). New York: Cambridge University Press.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence, 35*, 267-281.

Roznowski, M., Reith, J., & Hong, S. (2000). A further look at youth intellectual giftedness and its correlates: Values, interests, performance, and behavior. *Intelligence, 28*, 87-113.

Runco, M. A. (1991). *Divergent thinking.* Norwood, NJ: Ablex.

Satorra, A., & Bentler, P. M. (2001). A scaled difference $\chi^2$ test statistic for moment structure analysis. *Psychometrika, 66,* 507-514.

Sligh, A. C., Conners, F. A., & Roskos-Ewoldsen, B. (2005). Relation of creativity to fluid and crystallized intelligence. *Journal of Creative Behavior, 39*, 123-136.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.

Steiger, J. H. (1989). *EzPATH: Causal modeling.* Evanston, IL: SYSTAT.

Sternberg, R. J., & O'Hara, L. A. (1999). Creativity and intelligence. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 251-272). Cambridge: Cambridge University Press.

Süß, H.-M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 313-332). Thousand Oaks, CA: Sage.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & G. R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 119-169). Greenwich, CT: Information Age Publishing.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4-70.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89,* 696-716.