The issue of power in the identification of “g” with lower-order factors

Dora Matzke⁎, Conor V. Dolan, Dylan Molenaar

University of Amsterdam, The Netherlands

Article history:
Received 28 August 2009
Received in revised form 3 February 2010
Accepted 8 February 2010
Available online 4 March 2010

Abstract
In higher order factor models, general intelligence (g) is often found to correlate perfectly with lower-order common factors, suggesting that g and some well-defined cognitive ability, such as working memory, may be identical. However, the results of studies that addressed the equivalence of g and lower-order factors are inconsistent. We suggest that this inconsistency may partly be attributable to the lack of statistical power to detect the distinctiveness of the two factors. The present study therefore investigated the power to reject the hypothesis that g and a lower-order factor are perfectly correlated using artificial datasets, based on realistic parameter values and on the results of selected publications. The results of the power analyses indicated that power was substantially influenced by the effect size and the number and the reliability of the indicators. The examination of published studies revealed that most case studies that reported a perfect correlation between g and a lower-order factor were underpowered, with power coefficients rarely exceeding 0.30. We conclude the paper by emphasizing the importance of considering power in the context of identifying g with lower-order factors.

© 2010 Elsevier Inc. All rights reserved.

Keywords:
Covariance structure modeling
General intelligence
Log-likelihood difference test
Perfect correlation
Statistical power

1. Introduction

The positive intercorrelation among scores on cognitive ability tests is a well-established phenomenon, which is often explained by positing a general intelligence factor (g). The g-factor and the positive manifold of correlations may be viewed as synonymous, i.e., the positive manifold guarantees a dominant factor in principal component analyses (Basilevsky, 1983). However, we view the g-factor as a strong hypothesis, as the positive manifold may be attributable to causes other than a general factor (Bartholomew, Deary, & Lawn, 2009; van der Maas et al., 2006).

The g-factor is supposed to reflect the operation of a process that is common to all cognitive tasks (Jensen, 1998). There is, however, considerable disagreement as to what the nature of this general process might be. For instance, some researchers argued that g is in large part a reflection of frontal lobe functions such as inhibitory and control processes (e.g., Duncan, Burgess, & Emslie, 1995; Embretson, 1995; Dempster, 1991), while others stressed the importance of the speed of information processing (e.g., Demetriou et al., 2005; Kail & Salthouse, 1994; Jensen, 1998) or the efficiency of working memory (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002).

One source of information concerning the nature of g is higher-order factor modeling, in which g features as the highest order factor (Jensen, 1998). In these models, g is often found to correlate perfectly with a lower-order common factor, suggesting that g and some well-defined broad cognitive ability (represented by the lower-order factor) may be identical. However, the results of such studies are inconsistent. First, g has been found to be perfectly correlated with a wide variety of cognitive abilities, such as fluid reasoning (Gustafsson, 1984;
that g is highly correlated with both the equivalence of general intelligence and the proposed power of a variety of circumstances using realistic parameter values correctly reject the equivalence of two related factors under a description of the study in which we established the power to describe the study in which we constructed on the basis of realistic parameter values and the results of selected publications that reported a which we constructed on the basis of realistic parameter values and the results of selected publications that reported a perfect correlation between g and a lower-order common factor. Despite the theoretical importance of the issue of the exact nature of g, we are unaware of any order common factor. We focused exclusively on hierarchical factor models, with g as the single highest order factor, i.e., with g at the apex of the hierarchy (see Jensen, 1998). We suggest that the inconsistency in the results of studies that have addressed the equivalence of g and lower-order factors may in part be attributable to a lack of statistical power. In an underpowered study, the probability of rejecting the hypothesis that two highly correlated (e.g., 0.8 or 0.9) factors are in fact perfectly correlated is low. In such studies, one should be reticent to attach too great a meaning to the supposedly perfect correlation between g and the lower-order common factor. Despite the theoretical importance of the world, we are unaware of any study that has addressed the question of power in the context of identifying g with abilities represented by lower-order factors.

The goal of the present paper is therefore to study the power to correctly reject the hypothesis of perfectly correlated factors in situations, where g and the lower-order factor of interest are strongly, but not perfectly, correlated. To this end, we investigated the power to detect a less than perfect correlation using exact population (summary) statistics, which we constructed on the basis of realistic parameter values and the results of selected publications that reported a perfect correlation between g and a lower-order common factor. We focused exclusively on hierarchical factor models, with g as the single highest order factor, i.e., with g at the apex of the hierarchy (see Jensen, 1998).

The outline of this article is as follows. In the first section, we introduce the concept of statistical power and present a brief overview of power calculations in the context of maximum likelihood estimation. In the second section, we describe the study in which we established the power to correctly reject the equivalence of two related factors under a variety of circumstances using realistic parameter values based on the literature. In the third section, we investigate the power of five published studies by reconstructing the original factor models using reported parameter values. We conclude the paper with a discussion.

2. Statistical power and the log-likelihood difference test

The concept of statistical power plays an important role in formal statistical testing. Statistical power represents the probability of rejecting a hypothesis, given that it is false (e.g., Cohen, 1988, 1992; Kraemer & Thiemann, 1989). We show how power calculations are carried out within the common factor model. In this presentation, it may be useful to note that this approach to power calculation within models for covariance structures is conceptually the same as that to power calculation within a more basic statistical test like the independent samples t-test. In case of a t-test, power represents the probability of correctly rejecting the null-hypothesis (H₀ no difference between the two samples) in favor of an alternative hypothesis (H₁, a difference between the two samples) using the t-statistic. Here, we are concerned with the power to reject a parsimonious model (H₀) in favor of a more complex model (H₁) using the T_diff-statistic (see below). We now present the details of this approach.

In covariance structure modeling, which includes higher order factor modeling, the H₀ concerns the true model, and the H₁ concerns the false model (Satorra & Saris, 1985; Saris, W. Satorra, A. 1993). Here H₀ is a special case of H₀, in that H₁ can be derived from H₀ by the imposition of constraints on the parameters (e.g., by fixing a given parameter in H₀ to zero): H₁ is thus nested under H₀. In the present situation, H₀ is the model which includes a less than perfect correlation between g and a first-order factor, and H₁ is the model in which the correlation is fixed to equal one, by the imposition of an appropriate constraint.¹ As shown in Table 1, the present article is thus concerned with the power to correctly reject the H₁ of perfectly correlated factors in favor of the H₀ of correlated, but distinct, factors. For any given statistical test, power is a function of the sample size, the Type I error probability (α), and the effect size, i.e., the discrepancy between the value of the parameter(s) of interest under the H₀ (correlation of say 0.8) and the H₁ (correlation of 1.00) models. A power of 0.80 is generally considered adequate: i.e., by consensus, a value lower than 0.80 implies an unacceptable risk of a Type II error. Power higher than 0.80 may, ceteris paribus, require unrealistically large sample size (Cohen, 1992).

In the setting of maximum likelihood estimation, the tenability of the parameter constraints associated with the H₁ (i.e., the constraints that renders the correlation perfect) can be determined using a number of asymptotically equivalent

<table>
<thead>
<tr>
<th>Statistical decision</th>
<th>Reject H₁</th>
<th>Accept H₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>True state of the world</td>
<td>H₀ is true (r = 0)</td>
<td>H₁ is true (r &gt; 0)</td>
</tr>
<tr>
<td>False state of the world</td>
<td>H₀ is true (r &lt; 0)</td>
<td>H₁ is true (r &gt; 0)</td>
</tr>
</tbody>
</table>

¹ Given the positive manifold, and its implication that the common factors are positively correlated (Krijnen, 2004), we do not consider the possibility of a perfect negative correlation.
To determine the power of the test, one has to consider the distribution of the test statistic $T_{\text{diff}}$. If the two factors of interest were truly perfectly correlated (i.e., if the $H_0$ model was true), $T_{\text{diff}}$ would asymptotically follow a central $\chi^2$ distribution.

$$T_{\text{diff}} \sim \chi^2(d_{\text{diff}}, \lambda = 0),$$  

where $d_{\text{diff}}$ is the difference in the number of estimated parameters under the $H_0$ and the $H_A$ models and $\lambda$ is the non-centrality parameter, which equals zero here.

If the correlation between the two factors of interest was truly less than one (i.e., if the $H_0$ model was true), $T_{\text{diff}}$ would follow the non-central $\chi^2$ distribution (Satorra & Saris, 1985; Saris, W. Satorra, A. 1993),

$$T_{\text{diff}} \sim \chi^2(d_{\text{diff}}, \lambda > 0).$$

To put the non-central $\chi^2$ distribution at use, one has to determine its shape. The shape of the non-central $\chi^2$ distribution depends on $d_{\text{diff}}$ and the non-centrality parameter $\lambda$. To obtain a numerical estimate of $\lambda$, one first assigns plausible values to the parameters of the $H_0$ model and calculates the associated population covariance matrix. The parameters of the $H_0$ model are chosen such that (inter alia) the correlation between $g$ and the lower-order factor of interest is less than one (say 0.8 or 0.9). As the population covariance matrix is generated according to the $H_0$ model (i.e., true model), it obtained features as the true population covariance matrix. Second, one expresses the $H_A$ model by assigning plausible values to its parameters. The parameters of the $H_A$ model are chosen such that the correlation between $g$ and the lower-order factor of interest equals one. Note that both the $H_0$ and the $H_A$ models must be fully specified; all model parameters must be assigned plausible values. Third, one chooses a realistic, but arbitrary, sample size $N$, fits the $H_0$ and the $H_A$ models to the population covariance matrix and establishes the values of $T_{\text{diff}}$ and $d_{\text{diff}}$. Note that the likelihood ratio associated with $H_0$ ($T_0$) is zero, because $H_0$ is the true model, i.e., the model used to obtain the population covariance matrix. The likelihood ratio associated with $H_A$ ($T_A$) is greater than zero, as the $H_A$ model is false and does not fit the population covariance matrix. The value of $T_{\text{diff}}$ (Eq. (1)) equals the non-centrality parameter $\lambda$ (Satorra & Saris, 1985; Saris, W. Satorra, A. 1993).

Once the value of $\lambda$ is obtained, one can calculate the power of the test as follows. Choose the Type I error rate ($\alpha$) and the associated critical value ($C$) based on the central $\chi^2$ distribution. Next, determine the Type II error rate ($\beta$) by calculating the probability of observing a value of the test statistic $T_{\text{diff}}$ that is smaller than the chosen critical value $C$, given the non-central $\chi^2$ distribution,

$$\beta = P\left[\chi^2(d_{\text{diff}}, \lambda) < C\right].$$

The power of the test is then given by $1 - \beta$ (see Table 1). Note that this approach to power calculation does not rely on Monte Carlo simulations, rather it uses analytic methods to determine the probability to correctly reject the $H_A$.

Thus the steps towards analytical power calculation are 1) choose the parameter values of the $H_0$ model (i.e., the true model in which the correlation between $g$ and the first-order factor is less than perfect) and calculate the associated population covariance matrix; 2) choose an arbitrary sample size $N$; 3) fit the $H_0$ model to obtain $T_0$, which should equal zero (a useful check); 4) fit the $H_A$ model (i.e., the false model in which the correlation between $g$ and the first-order factor equals one) to obtain $T_A$, which will assume a value greater than zero; 5) calculate $\lambda$ and $d_{\text{diff}}$; 6) choose the Type I error rate ($\alpha$) and calculate the associated critical value $C$; 7) calculate the power given, $\lambda$, $N$, and $\alpha$. As noted below, the power for other values of $N$ can be calculated easily, i.e., does not require refitting the model.

### 3. Power analysis

The objective of the power study was to establish the power to correctly reject the $H_A$ of perfectly correlated factors under a variety of circumstances using realistic parameter values. The power to detect the distinctiveness of $g$ and a lower-order factor depends on a number of aspects of the factor model, including the number and the reliability (i.e., explained variance in the factor model) of the indicators, and the effect size (i.e., the discrepancy between the correlation of $g$ and the lower-order factor under the $H_0$ and the $H_A$ models). Therefore, we systematically manipulated these features of the factor models to study their influence on power.

#### 3.1. Design

We focused exclusively on hierarchical factor models with five first-order factors and with $g$ as a single second-order factor. First, we specified the $H_0$ models in which the correlations between $g$ and all first-order factors were less than one. The parameter values of the models were chosen to span the range of values reported in published studies. In each model, the correlations (i.e., standardized second-order factor loadings) between $g$ and the first-order factors were set to 0.95, 0.90, 0.85, 0.80, and 0.75, with corresponding explained variances of 0.90, 0.81, 0.72, 0.64, and 0.56. To study the effects of the number and the reliability of the indicators, the $H_0$ models were created by systematically manipulating these aspects of the models. The $H_0$ models featured either two, three, or four indicators per first-order factor, where the reliabilities of the indicators were either relatively high (ranging from 0.55 to 0.80) or relatively low (ranging from 0.3 to 0.4).
This design resulted in $3 \times 2 = 6$ different $H_0$ models. Fig. 1 presents an example of a hierarchical factor model used in the study. Note that each first-order factor had the same number of indicators and that the residuals of the indicators as well as the residuals of the first-order factors were uncorrelated.

Second, we specified the $H_A$ models that implied perfect correlations between $g$ and a given first-order factor. Each $H_0$ model had five corresponding $H_A$ models, where each $H_A$ model was created by constraining the correlation between $g$ and one of the five first-order factors to one. This design resulted in $6 \times 5 = 30$ different $H_A$ models. Because each first-order factor correlated differently with $g$, this approach allowed us to investigate the influence of various effect sizes, ranging from 0.05 to 0.25.

Third, we obtained the non-centrality parameter $\lambda$ for each $H_A$ model. To this end, we first calculated the population covariance matrices associated with the six $H_0$ models. Next, we fitted each $H_A$ model to the data generated according to its corresponding $H_0$ model, using an arbitrary sample size of $N = 200$. To obtain perfectly correlated factors, each $H_A$ model was fitted by constraining the residual variance (i.e., unexplained variance) of the first-order factor of interest to equal zero (van der Sluis, Dolan, & Stoel, 2005). Discarding subject indices, let
\begin{equation}
\eta_i = \gamma_i g + \zeta_i
\end{equation}
denote the regression of the $i$th first-order factor ($\eta_i$) on $g$, with $\zeta_i$ representing the residual. Scaling the variance of $g$ to equal one, the correlation between $\eta_i$ and $g$ equals
\begin{equation}
\rho_{\eta g} = \frac{\gamma_i}{\sqrt{\gamma_i^2 + \sigma_\zeta^2}}.
\end{equation}

If $\sigma_\zeta^2$ is constrained to equal zero, then clearly the correlation equals
\begin{equation}
\rho_{\eta g} = \frac{\gamma_i}{\sqrt{\gamma_i^2}} = 1.
\end{equation}

The goodness-of-fit statistics of the $H_A$ models ($T_A$) were used as approximations for the non-centrality parameters.

Finally, we calculated the power to reject each $H_A$ model using the obtained non-centrality parameter and $df_{diff} = 1$. Note that because the $H_A$ models were obtained by constraints on the variance components, the Type I error probability ($\alpha$) was doubled to correct for the violation of the admissible parameter space (e.g., Dominicus, Skrondal, Gjessing, Pedersen, & Palmgren, 2006; Stoel, Garre, Dolan, & Van den Wittenboer, 2006). The power was therefore computed given $\alpha = 0.05 \times 2 = 0.1$. Note also that the results reported in the following section are based on non-centrality parameters computed with $N = 200$. However, the non-centrality parameters and the power can easily be computed for any other sample size using:
\begin{equation}
\lambda_{\text{new}} = \left( \frac{\lambda_{\text{original}}}{200} \right) \times N_{\text{new}}.
\end{equation}

where $\lambda_{\text{original}}$ is the value of the non-centrality parameter reported in the present article (based on $N = 200$) and $N_{\text{new}}$ is the new sample size of interest. The models were fitted using LISREL (Jöreskog & Sörbom, 2001). The data generation and the power calculations were carried out using the R package (R Development Core Team, 2006). The R code for the power calculation is presented in the Appendix.
3.2. Results

Table 2 summarizes the results of the power calculations. Over the $H_A$ models that we considered, power varied from 0.17 to 1. In general, the three manipulations (i.e., the number and the reliability of the indicators and the effect size) all influenced the power to detect the distinctiveness of $g$ and the lower-order factors.

We first consider the effects of the reliability of the indicators. The results indicated that power increased as the reliability of the indicators increased. The increase in power was, however, more pronounced when the number of indicators and the effect sizes were relatively low. For low effect sizes, power coefficients varied by as much as 0.6 across the two reliability conditions. This difference was reduced for higher effect sizes, especially when the number of indicators was relatively high. With respect to the sample size, the results followed the same pattern. The sample size required to reach sufficient power (i.e., 0.80) was substantially lower for reliable indicators than for unreliable indicators; increases in the reliability of the indicators resulted in an average decrease of 90% in the necessary sample size.

Turning to the effects of the number of indicators, the results indicated that for reliable indicators power was high regardless of the number of indicators. For unreliable indicators, power increased as the number of indicators increased, with power coefficients varying by about 0.15 over the levels of this factor. Similarly, for reliable indicators, the necessary sample size was relatively low regardless of the number of indicators. For unreliable indicators, an increase in the number of indicators was accompanied with an average decrease of 45% in the required sample size.

Lastly, with respect to the influence of the effect size, the results indicated that for reliable indicators power was high regardless of the value of the effect size. For unreliable indicators, power increased as the effect size increased. Note, however, that the increase in power was generally more pronounced – reaching nearly 0.40 – for lower effect sizes, especially when the number of indicators was high. Similarly, for reliable indicators, the necessary sample size was relatively low regardless of the value of the effect size. For unreliable indicators, the necessary sample size decreased as the effect size increased. This decrease varied between 70% and 30% over the levels of this factor.

To summarize, the results of the analyses indicated that power was substantially influenced by the effect size and the number and the reliability of the indicators. For reliable indicators (i.e., from 0.55 to 0.80), power was high regardless of the effect size and the number of indicators. For less reliable indicators (i.e., from 0.20 to 0.45), power increased as the effect size and the number of indicators increased, with a corresponding decrease in the sample size required to reach sufficient power.

### 4. Power analysis of selected case studies

In this section, we investigate the approximate power of five published studies in order to highlight the importance of power in the context of identifying $g$ with lower-order factors. We reconstructed the original factor models using reported parameter values and determined the power to reject the equivalence of $g$ and the proposed lower-order factors. We focused on studies that used hierarchical factor models – with $g$ as a second or third-order factor – and reported a (near) perfect correlation between $g$ and a lower-order factor.

#### Table 2

Power to detect the distinctiveness of $g$ and the lower-order factors.

<table>
<thead>
<tr>
<th>Effect size</th>
<th>2 indicators</th>
<th>3 indicators</th>
<th>4 indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reliable (0.75)</td>
<td>Unreliable (0.38)</td>
<td>Reliable (0.68)</td>
</tr>
<tr>
<td>0.05</td>
<td>Power at $N = 200$</td>
<td>0.77</td>
<td>0.17</td>
</tr>
<tr>
<td>N at power = 0.8</td>
<td>221</td>
<td>3016</td>
<td>160</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>5.60</td>
<td>0.41</td>
<td>7.76</td>
</tr>
<tr>
<td>0.10</td>
<td>Power at $N = 200$</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>N at power = 0.8</td>
<td>66</td>
<td>853</td>
<td>46</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>18.76</td>
<td>1.45</td>
<td>26.98</td>
</tr>
<tr>
<td>0.15</td>
<td>Power at $N = 200$</td>
<td>1.00</td>
<td>0.55</td>
</tr>
<tr>
<td>N at power = 0.8</td>
<td>35</td>
<td>401</td>
<td>23</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>36.23</td>
<td>3.09</td>
<td>54.39</td>
</tr>
<tr>
<td>0.20</td>
<td>Power at $N = 200$</td>
<td>1.00</td>
<td>0.72</td>
</tr>
<tr>
<td>N at power = 0.8</td>
<td>24</td>
<td>249</td>
<td>16</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>52.84</td>
<td>4.98</td>
<td>82.42</td>
</tr>
<tr>
<td>0.25</td>
<td>Power at $N = 200$</td>
<td>1.00</td>
<td>0.85</td>
</tr>
<tr>
<td>N at power = 0.8</td>
<td>18</td>
<td>172</td>
<td>12</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>69.28</td>
<td>7.23</td>
<td>111.77</td>
</tr>
</tbody>
</table>

Note. The effect size reflects the discrepancy between the correlation of $g$ and the lower-order factor under the $H_A$ (i.e., correlation of one) and the $H_0$ models. The average reliability of the indicators is shown in brackets. The reliabilities of the indicators in the $H_0$ models were chosen as follows. In the two $H_0$ models that featured two indicators per first-order factor, the reliabilities of the indicators were set to 0.7 and 0.8 in the reliable condition and to 0.3 and 0.45 in the unreliable condition. In the two $H_0$ models that featured three indicators per first-order factor, the reliabilities of the indicators were set to 0.55, 0.7 and 0.8 in the reliable condition and to 0.2, 0.3 and 0.45 in the unreliable condition. In the two $H_0$ models that featured four indicators per first-order factor, the reliabilities of the indicators were set to 0.55, 0.7, 0.75 and 0.8 in the reliable condition and to 0.2, 0.3, 0.35 and 0.45 in the unreliable condition.
4.1. Design

The selected case studies and some details of the investigated factor models are listed in Table 3. It is important to note that some of these studies (e.g., Gustafsson, 1984; Undheim & Gustafsson, 1987) formally tested the presence of a perfect correlation and clearly emphasized the equivalence of g and the proposed lower-order factors, while others (e.g., Johnson & Bouchard, 2005) merely reported the perfect correlation between the two factors and did not draw further conclusions about their equivalence. Nevertheless, the results of these publications provided interesting case studies to investigate the power to detect the distinctiveness of highly correlated factors.

The power analyses of the case studies were conducted as follows. First, we reconstructed the H0 models and the corresponding population covariance matrices using the original factor structures and the exact parameter values reported in the articles. Although we attempted to approximate the original factor models as closely as possible, we did not allow for cross-factor loadings and residual correlations. Further, if the reported correlation between g and the lower-order factor of interest equaled one – either because it was fixed to one or estimated to be one – we set the standardized factor loading between the two factors to 0.95. Note also that in some models the residual variance of the lower-order factor of interest was negative (i.e., Heywood case; Heywood, 1931). Although a Heywood case raises important questions related to model mis-specification and/or over-parameterization (Jöreskog & Sörbom, 1988), the issue of the adequacy of these models is beyond the scope of the present article and will not be considered further. For the purposes of the present investigation, if the reported residual variance of a lower-order factor was negative, we set the standardized residual variance to 0.10, corresponding to a standardized factor loading of 0.95. Next, we specified the H1 models of perfect correlation by constraining the residual variances of the lower-order factors of interest to equal zero. The H1 models were then fitted to the generated datasets using the original sample sizes. Lastly, we computed the power to reject the various H1 models using the obtained non-centrality parameters, \( df_{diff} = 1 \), and \( \alpha = 0.05 \times 2 = 0.1 \).

4.2. Results

The results of the case studies are shown in Table 3. Across the various models, power varied from 0.135 to 0.99. In most cases, however, power did not exceed 0.3, indicating that the power to reject the equivalence of g and a given lower-order factor was generally very low. In the light of the results reported above, this result was not unexpected. Most case studies featured extremely low effect sizes and factor models with only a few (two or three) relatively unreliable indicators per first-order factor. Also in line with our results, power was substantially higher for studies that used relatively reliable or a large number of indicators, reaching 0.6 for Gustafsson’s (1984) model and exceeding 0.9 for the first model of Johnson and Bouchard (2005). In summary, the results suggested that the selected case studies, with a very few exceptions, were underpowered to detect the distinctiveness of g and the proposed lower-order factors.

5. General discussion

The goal of this study was to determine the power to reject the hypothesis that g and a lower-order factor are perfectly correlated, given that the correlation is relatively high, but
In the light of these results, a large number of indicators (i.e., Johnson & Bouchard, 2005) may be quite large. Such correlations do certainly require an examination of power before concluding that \( g \) and a lower-order factor would feature a substantially perfect relationship. The failure to do so reveals that the correlation between height and weight is equal to one. The perfect correlation between height and weight does not imply that the two variables are identical nor that they necessarily share a common causal substrate. Nevertheless, it is possible that \( g \) may indeed be identical to a certain lower-order factor. However, given the very mixed results and the lack of power of most published studies, we are reluctant to accept this. We point out that if such an identity did truly exist, it would imply, by the application of Occam’s razor, the demise of \( g \) as a causal factor in the study of individual differences in cognitive abilities: why entertain the notion of a single essentially ill-defined higher order factor, if it is in fact identical to a well-defined lower-order factor (say, working memory)?

We also note that the focus on individual differences, which characterizes studies of \( g \), has its inherent interpretational limitations: the presence of a perfect correlation between two variables is not sufficient to conclude that the variables are identical or that they share a common causal substrate. Suppose, for the sake of argument, that in a sample of children, who vary sufficiently in age, an appropriate statistical test reveals that the correlation between height and weight is equal to one. The perfect correlation between height and weight does obviously not imply that the two variables are identical nor that they necessarily share a common causal substrate.

In conclusion, the goal of the present study was to highlight the importance of considering power in the context of identifying \( g \) with lower-order factors. Our results provide useful guidelines on the ideal dataset and the necessary sample size required to reach sufficient power in a variety of realistic situations. Furthermore, the procedure used here to investigate power is easy to implement and the R code presented in the Appendix provides a helpful tool to establish the power and the necessary sample size in the particular situation at hand. As pointed out above, the failure to do so might result in mistakenly concluding that \( g \) and a lower-order factor are perfectly correlated and therefore – at least from the perspective of individual differences – can be considered identical.

**Acknowledgements**

Dylan Molenaar is supported by a grant from the Netherlands Organization for Scientific Research (NWO).

**Appendix A**

**R code for power calculations**

This appendix presents the R code that can be used to calculate power for various sample sizes. The code takes as inputs the goodness-of-fit statistic of the \( H_0 \) model (\( \Lambda \)), the chosen Type I error probability (\( \alpha \)), the sample size (\( N \)) used to obtain the non-centrality parameter \( \lambda \), and the minimum (\( \text{minN} \)) and maximum (\( \text{maxN} \)) sample sizes of interest. The output provided by the code consists of the power coefficients corresponding to sample sizes ranging from the minimum and the maximum sample size of interest.
#Goodness-of-fit statistic of HA model (i.e., non-centrality parameter lambda)
TA = 7.23
#Type I error probability
alpha = 0.05 * 2
# Degrees of freedom (i.e., df.diff)
df = 1
# Sample size used to calculate the non-centrality parameter lambda
N = 200
# Critical value
C = qchisq(alpha, df=df, ncp=0, lower.tail=F)
# Minimum sample size of interest
minN = 100
# Maximum sample size of interest
maxN = 2000
power = matrix(0, maxN - minN + 1, 2)
# Nnew is new sample size of interest
for (Nnew in minN:maxN) {
    # lambda.new is the value of the non-centrality parameter corresponding to the new sample size
    lambda.new = (TA/N) * Nnew
    # Calculate power
    power[Nnew-minN+1,1] = pchisq(C, df=df, ncp=lambda.new, lower.tail=F)
    power[Nnew-minN+1,2] = Nnew
}
# Power plot
plot(power[,2],power[,1], type='l', xlab="Sample size", ylab="Power")
# Power for original N
print(power[,2]==N,1])

References


