Intelligence xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Spearman's law of diminishing returns. A meta-analysis

Diego Blum*, Heinz Holling

Westfälische Wilhelms-Universität Münster, Fliednerstraße 21, 48149 Münster, Germany

ARTICLE INFO

Keywords:
Spearman's law
Differentiation
g saturation
ability
age

ABSTRACT

The cognitive ability differentiation hypothesis, which is also termed Spearman's Law of Diminishing Returns, proposes that cognitive ability tests are less correlated and less *g* loaded in higher ability populations. In addition, the age differentiation hypothesis proposes that the structure of cognitive ability varies across respondent age. To clarify the literature regarding these expectations, 106 articles containing 408 studies, which were published over a 100-year time span, were analyzed to evaluate the empirical basis for ability as well as age differentiation hypotheses. Meta-analyses provide support for both hypotheses and related expectations. Results demonstrate that the mean correlation and *g* loadings of cognitive ability tests decrease with increasing ability, yet increase with respondent age. Moreover, these effects have been nearly constant throughout the century of analyzed data. These results are important because we cannot assume an invariant cognitive structure for different ability and age levels. Implications for practice as well as drawbacks are further discussed.

1. Introduction

1.1. Theoretical background

For more than a century, the psychometric tradition underlined the importance of a structural organization of abilities following a hierarchical order. According to the theory, hierarchical abilities can be placed in vertical direction from the most general at the head, until the most specialized at the bottom, where upper levels have an impact on lower levels (Burt, 1949; Carroll, 2003; Spearman, 1904; Thomson, 1919). Furthermore, this hierarchical theory has been used as a basis for the development of more specific theories, all of them giving particular importance to the role and properties of higher-order abilities. Examples are the broad constructs named as Fluid and Crystallized Ability by Cattell (1943) and Horn (1976), Thurstone's (1938) Primary Mental Abilities, the Berlin Intelligence Structure Model (BIS; Jäger, 1982, 1984), the three levels of ability proposed by Carroll (2003), and Spearman's (1904, 1927) General Intelligence or *g*. The latter is a single higher-order composite that is said to generate a positive manifold among abilities of different kinds.

Spearman demonstrated the existence of a *g* factor through positive correlations among the scores of individuals in several maximum-performance tests. Furthermore, he presented evidence suggesting that the amount of this general composite decreases as a function of ability (1927), i.e., the higher the ability, the lower the correlations among the tests will be. This evidence is based on research showing a tendency of reduced inter-test correlation coefficients for groups of higher-skilled individuals compared to those lower-skilled (i.e., normal vs. defective children, older vs. younger children, and adults vs. children).

Although these results suggest that an inverse relationship between the *g* saturation and ability is present, Spearman did not specify the type of relation (linear, curvilinear, logarithmic, etc.). He also did not constrain his study to any age or intelligence range in particular. Hence, the evidenced effect is a broad and unspecific interpretation of a few results. Nevertheless, he explained the phenomenon by stating that these higher-skilled groups comprised a higher amount of mental "energy" (p. 219), and that, as a consequence, ability was being less benefited from further increments of it. He considered this to be analogous to a fundamental principle of economics called the Law of Diminishing Returns.

This law states that, if the amount of input of a production process continuously increases and all of the other production factors stay constant, the rate of growth of the output will eventually decrease; this means that returns are diminished at a certain level as a consequence of expanding the volume of input. Thus, from a psychometric point of view, indicators of higher ability, such as an elevated *IQ* or increased age, could be accompanied with abilities being less dependent among each other and, therefore, less saturated with *g*.

This relationship was further explained as an increase of cognitive specialization, often known as the differentiation hypothesis. It states that factor patterns of intelligence are more differentiated when a personal condition, like being higher-skilled or older, is met (Reinert, 1970). With respect to ability, the differentiation hypothesis implies that factor structures are more differentiated for high-ability than for low-ability individuals, also called the ability-differentiation hypothesis

* Corresponding author.
E-mail addresses: blumworx@gmail.com (D. Blum), holling@uni-muenster.de (H. Holling).

http://dx.doi.org/10.1016/j.intell.2017.07.004
Received 29 December 2016; Received in revised form 17 June 2017; Accepted 11 July 2017

by Reinert, Baltes, and Schmidt (1965). With respect to age, it implies that "abstract or symbolic intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (Garret, 1946, p. 373). This was called the age-differentiation hypothesis by Anastasi (1958), with much research attempting to describe how factor structures of intelligence develop as a function of age (Reinert, 1970). A comparable theory of age differentiation extended to the entire life span states, however, that while this phenomenon occurs from childhood until early maturity, the opposite effect (i.e., the increase of the importance of the general ability composite) is expected from early to late maturity, thus suggesting the existence of a U-shaped relationship (Balinsky, 1941). On the contrary, Spearman (1927) did not seem to consider the entire life span; he only went as far as to compare adults with children, and no suggestion of a U-shaped effect was accomplished by him. Finally, an intersection point can be found between these two hypotheses because an earlier age level is usually marked by a lower ability level (Reinert, 1970).

1.2. Current state of knowledge

The interest in finding evidence about Spearman's expectations as well as the differentiation effect kept psychometricians busy throughout practically 100 years, and, according to Fogarty and Stankov (1995), researchers paid more attention to this topic during the last decades of the XX century. However, proof about such an effect, as well as against it, creates a contradictory state of knowledge (Hartmann & Nyborg, 2006; Reinert, 1970). As an example of ability differentiation, Detterman and Daniel (1989) revealed that individuals with low *IQ* comprise higher correlations among *IQ* tests than individuals with high *IQ*. As opposed to this finding, Amelang and Langer (1968) divided their sample in two groups distanced by 12 to 13 *IQ* points who responded to a number of ability tests, and the difference between the first unrotated eigenvalues of the two data sets was found to be non-significant by Hartmann and Nyborg (2006).

Regarding the age differentiation, Hertzog and Bleckley (2001) studied the performance of three groups of individuals with respect to a battery of cognitive tests. One of the groups consisted of undergraduate students, and the other two comprised age ranges of 43–62 and 63–78. The general tendency was that both test and factor intercorrelations were higher for older groups than for younger ones. On the contrary, Escorial, Juan-Espinosa, García, Rebollo, and Colom (2003) found practically no change regarding the percentage of variance accounted for by the first unrotated principal component between groups with a respective age range of 16–24 and 35–54 that responded to a set of ability measures.

It is important to notice that nearly all of the existing studies related to differentiation arrive at results which can be deemed as partialized, given that they more or less depend on the selected sample, the chosen test battery, the place of conduction, among other variables. Thus, a reasonable doubt about the possibility to be able to compare results among publications is present. In fact, Molenaar, Dolan, Wicherts, and van der Maas (2010) mention some interesting critiques to current research reporting correlation matrices among subtests, which can be summarized as follows:

- Most research methods are ad hoc and lack of explicit framework. Therefore, they have not enough quality.
- Creating subgroups, each with different predictor levels, and studying the *g* differences among them can be problematic because the factor structure may be different for the whole population than it is for each subgroup, which could lead to the distortion of the subtest correlations within subgroups as well.
- To solve the latter issue, some other researchers created subgroups based on the latent variable (i.e., second-order factor) in order to distort the factor structure of the subgroups in a lower degree. The problem is that the structure is not preserved with many types of factor scores because "the covariances among the common factors based on the calculated factor scores are not equal to the covariances as estimated in fitting the factor model" (p. 613).
- Researchers are generally not unanimous about which specific *g* measure to use.

To our knowledge, only one recent attempt has been made by Hartmann and Nyborg (2006) to perform a detailed review of the existing literature pertaining the differentiation effect with the purpose of establishing a comprehensive understanding of it. By following some of Reinert's (1970) previous short review, Hartmann and Nyborg accomplished an enormous summary and categorization of empirical research from 1923 until 2004. They showed how several measures of the *g* saturation, such as the mean correlation among ability tests, the first unrotated eigenvalue or its total explained variance, change depending on ability or age, in accordance with previous research. They assessed statistical differences regarding the mean correlation between groups differing on ability or age, and for that purpose they used the *t*-test on Fisher's *Z*-transformed correlations in consonance with Lynn and Cooper's (1993, 1994) approach. They also scored the selected studies with a maximum of 4 quality points depending on whether researchers published one of the expected *g* measures such as the average intercorrelation (1 point), controlled or manipulated ability (1 point) or age (1 point), and whether they controlled the *SD* of ability (1 point).

According to Hartmann and Nyborg's results regarding ability effects, out of the 9 ability-related studies considered by them to be of highest relevance (i.e., those within the range of 3½–4 quality points), 3 show a tendency of decreasing *g* saturation as a function of ability, only significant in one case, other five show no relation or a very modest relation between the variables, and the remaining study reveals an increasing effect. After recalculating scores by giving also one point to studies not controlling for age, Hartmann and Nyborg were left with 18 papers in the subgroup, 10 out of which show a decreasing effect, six of them being significant. Hartmann and Nyborg concluded that there seems to be a moderate support for a small effect of ability on the *g* saturation.

With respect to studies of the impact of age on the *g* composite, most research reveals a *g*-saturation increase in later life, and almost 50% of the studies "point in the direction of a non-significant tendency towards a U-curvilinear relationship between *g* saturation and age, with the minimum *g* saturation from ages 18-34" (p. 110). This seems to confirm the extended theory of age differentiation already described (Balinsky, 1941). Another 25% of all studies does not reveal a relation whatsoever between both variables, and 15 to 20% of the remaining literature shows an inverted U-shaped relation where the *g* saturation is the highest within the age period 10–14. Furthermore, according to a subsample of age-related studies which were given 3½–4 quality points, Hartmann and Nyborg found their results to be even more contradictory than the ones of the broader sample. Therefore, it was their point of view that outcomes are not supportive of the age differentiation effect.

Hartmann and Nyborg argued that the effect of age on the *g* saturation is perhaps mediated by the effect of ability. In fact, Spearman (1927) did not ever mention that *g* could be a function of age. He only gave examples where groups of younger and older individuals were being compared with respect to *g*, given that the ability level was expected to be different across these groups as well. Moreover, ability is known to drastically change along the earlier years of life, and then the relation between ability and age diminishes (Kalveram, 1965; Merz & Kalveram, 1965). Finally, Reinert et al. (1965) found that, when ability is kept constant, no effect of age on the *g* saturation is present. If results supporting a U-shaped relationship between age and *g* loadings are considered, this relation approximately fits with an inverted U-shaped relation between age and ability, where ability is maximal at middle ages (Hartmann & Nyborg, 2006).

All in all, according to Hartmann and Nyborg's review, 10 out of the

18-mentioned studies with the ability predictor reveal a decrease of the *g* saturation as ability advances, therefore being supportive for such an effect. On the other hand, practically 50% of studies with the age predictor shows little evidence of a U-shaped effect where the *g* saturation is the lowest between 18 and 34 years, whereas other studies reveal contradictions regarding the latter outcome. Moreover, the subgroup of papers considered of highest relevance with respect to age shows no support for a unanimous tendency. One of Hartmann and Nyborg's conclusions was that a meta-analysis might help in order to estimate the true effect size, even after considering the heterogeneity of approaches seen in the literature.

1.3. Aims of the current review

Hartmann and Nyborg's work reflects the current state of knowledge on the topic. Even though it is not a meta-analysis, it prepares the framework for further and more rigorous approaches. As a response, we develop in this paper a detailed meta-analytic study influenced by Hartmann and Nyborg's previous work. Here we will prove whether the same relations as those described by Hartmann and Nyborg can be replicated or if rather different trends than the ones stated by them are evidenced. We will also compare these results with Spearman's (1927) expectations to see if they are in line with our data from a broad perspective. The compound questions that arise from the collected works describe the relation between predictors and a common criterion, the latter being the *g* saturation seen in the scores of maximum-performance tests.

The questions to be addressed in this meta-analytic study are the following:

1. How does the *g* saturation relate to ability indicated by *IQ*? Is there evidence of a decreasing effect of ability on *g* loadings?
2. How does the *g* saturation relate to age? Is there evidence of a linear relation, U-shaped relation, inverted U-shaped relation or any other curvilinear relationship?

These questions will be analyzed by performing a meta-analytic review. It provides the possibility of taking both independent variables as continuous predictors, as opposed to most research related to the differentiation hypothesis that simply compared lower- vs. higher-skilled or younger vs. older groups of testees. Furthermore, as Hartmann and Nyborg (2006) explained, the confirmation of such effect would imply that the cognitive structure of abilities is not the same for different ability or age levels. As a consequence, the question of validity of the *g*-factor theory would be raised from a developmental perspective. Additionally, the confirmation of decreasing *g* loadings as a function of ability, which was investigated by many psychological researchers over the course of a century, would be in line with Spearman's statements as well as with the ability-differentiation hypothesis.

2. Method

2.1. Literature search and inclusion criteria

We performed a keyword online search by using PsycARTICLES, PsycINFO and PSYNDEX in joint basis; the chosen time span was from 1916 until June 2016. The search term is the following: (test OR battery OR batterie OR scale OR skala) AND (intelligence OR intelligenz OR g factor OR g faktor OR g-factor OR g-faktor) AND (IQ OR age OR alter) AND (correlation OR korrelation OR factor analysis OR faktorenanalyse).

It returned 3396 results for a further inspection. From the latter amount, a subgroup of 253 articles in English or German languages was preselected after screening only by the title and, in case of doubt, also by the abstract. A more specific inspection of each study belonging to this subgroup allowed the identification of 75 final publications which

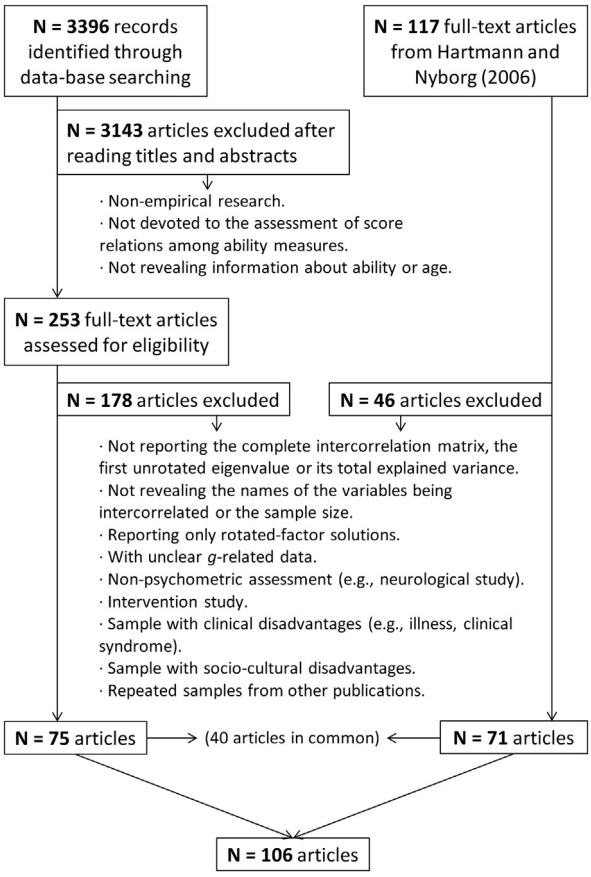

Fig. 1. Study selection criteria.

met every single aspect of our interests for the meta-analytic study. Then we performed a backward research by screening Hartmann and Nyborg's (2006) research selection. We looked for articles that were not found by the online search engine and that were also in line with our interests. As a result, we added 31 more papers from Hartmann and Nyborg's work, thus returning 106 articles in total. Search criteria followed the guides of the Cochrane Handbook for Systematic Review of Interventions, version 5.1.0 (Higgins & Green, 2011). Fig. 1 shows the study selection step by step.

Additional criteria also helped us select the final 106 manuscripts in both steps. As for the very few publications including two or more assessments of the same group in relatively similar moments in time, we decided to select only one of such studies per group to avoid dependent data (Gleser & Olkin, 2009). On the other hand, the assessment of different groups by the same researchers, or the multiple assessment of the same group with a distance of at least one year between study moments, was not considered problematic because cluster-robust-data estimation did not show that the variance of the overall effect size is increased in an important way due to the inclusion of more than one study per paper. Another important criterion was to eliminate studies applying only one or two of the types of tests detailed in Appendix 1 because we wanted the *g* saturation to represent a variety of abilities. This criterion based on a test-type taxonomy led us to eliminate 8 articles as well as some studies of another two articles (in sum, 32 studies).

2.2. Obtained studies

The selected 106 publications describe 408 studies altogether, with 98 of these studies showing data with respect to *IQ* and 394 informing respondent age. Appendix 2 presents the basic data for each of the included studies, ordered from former to later year of publication. In total, results from 188,489 individuals were assessed here. The overall

mean of the sample size is 469.8 (SD = 1733.21), the overall mean of the mean IQ for studies reporting it is 101.97 (SD = 12.92), the overall mean of the mean age for studies informing it is 21.57 (SD = 20.48) with a life span ranging from 2 to 90 years, and the overall mean number of the included variables or (sub)tests is 11.71 (SD = 5.51). For studies that inform the gender distribution, the overall mean percentage of women is 45.47 (SD = 31.57). As for missing data, sample sizes had to be estimated only in a few cases; when the age of school pupils was not revealed, it was estimated through the informed school grade level, provided that the age-grade correspondence was checked regarding the educational system of the respective country of origin.

2.3. Meta-analytic procedure

All data were analyzed by means of the 'metafor' (Viechtbauer, 2010) package available for R software environment (R Core Team, 2015). As effect-size measure (i.e., the g-saturation estimate), the average correlation ($\bar{r}$) among all correlations (r) between variables was chosen. For each correlation matrix, Fisher's Z-transformed r's were averaged, and the result was back-transformed to obtain $\bar{r}$.

The average correlation is equivalent to the total explained variance of the first unrotated eigenvalue resulting from the above-mentioned matrix. It is calculated as λ/p, where λ is the first unrotated eigenvalue and p is the number of variables. When only the first eigenvalue (λ) was reported, or whenever it could be calculated from other data, the transformation was done by $\bar{r} = (\lambda - 1) / (p - 1)$ (see Kaiser, 1968).

A three-level multilevel meta-analysis was applied to consider the variance between articles ${\tau_s}^2$, the variance between different effect sizes of the same article ${\tau_w}^2$ and the sampling variance ${\tau_e}^2$ (see e.g., Cheung, 2014, for multilevel meta-analysis).

Studying how moderators affect the variation of the effect size is the main purpose of the present review. Two main moderators were here assessed as they are in line with the addressed questions, namely, mean IQ of the group ($\overline{IQ}$), and mean age of the group ($\bar{a}$). Another moderator considered here was the year of publication.

The funnel plot was inspected to see whether a publication bias was present. In meta-analysis, higher effect sizes might be overrepresented because studies with non-significant effect measures or small effect sizes may have a lower probability to be published. Furthermore, a visual inspection may reveal that some studies yield rather different observed effects than the rest of the data, thus potentially influencing the results (Borenstein, Hedges, Higgins, & Rothstein, 2009). For these cases, case-deletion diagnostics were computed such as standardized residuals and Cook's distances to remove these studies if necessary.

2.4. Quality assessment

Based on Detterman and Daniel's (1989) ideas, as well as on Detterman's personal communications,[1] we graded each selected study on the methodological criteria described below. Thus, an overall quality score was calculated for every study to further analyze how the methodologically best studies perform with respect to effects. Grades were the following:

- 1 point if the sample had certain characteristics that made it more likely to represent the entire population. Such characteristics were a sample variance with no more than a 0.05-point difference with respect to the population variance and the usage of a previously-standardized test battery.
- A score that is dependent on the total number of IQ- and/or age-related categories or subgroups. The fewer the number of categories, the less chance of finding an effect among them will occur, even if it is large because it will be obscured by the large number of cases from the middle of the distribution. Those papers reporting two to four subgroups obtain a score of 0.5, those reporting five to eight obtain a score of 0.75, and those reporting more than eight acquire a score of 1 with this criterion.
- 1 point if a formula was applied for range restriction after generating the subgroups, in order to get correlation matrices which are valid for the whole population. If the predictor is divided into subgroups of reduced ranges, this restricts the observed correlation of a range to be less than would be observed in the predictor's full range; therefore it should be corrected for range restriction.
- 1 point for the specific case of IQ categories if subtest scores were used as criterion to generate the IQ subgroups and then not included in further calculation, because the inclusion of such a subtest in further calculation could attenuate the correlations or even force them to be negative.

3. Results

We used $\bar{r}$ as measure of effect size. The variance of this statistic has been derived using the sampling distribution of λ (see Anderson, 2003, p. 474), given that λ is equivalent to $\bar{r}$. In some studies correlations based on a Confirmatory Factor Analysis (CFA) were reported. As these correlations are corrected for measurement error, they are higher than those based on observed variables. To attenuate the correlations based on CFA, an attenuation factor had to be estimated and was applied to the mean correlations of factors of a CFA (see details in Appendix 3). The following results are divided in four subsections: (a) meta-analyses without moderators, (b) meta-analyses with the ability moderator, (c) meta-analyses with the age moderator, and (d) additional results.

3.1. Meta-analyses without moderators

The multilevel meta-analysis yields the overall estimate $\hat{\rho}$ = .3374 (SE = .0118, $p < .001$) and ${\tau_s}^2$ = .0122. The total explained variance corresponding to $\hat{\rho}$ = .3374 amounts to 40.95%. τ_s = .1104, thus representing an important standard deviation of the overall effect $\bar{r}$; this amount should be expected because of the heterogeneity of approaches about the topic. Appendix 4 shows a large Forest Plot with the $\bar{r}$ of every study at the right side plus $\hat{\rho}$ at the bottom, along with their respective confidence intervals between square brackets. The weights of the sample effect sizes, which are $1/\mathrm{Var}(\bar{r}_i)$, are graphically represented by the thickness of the squares, whereas the width of the diamond-shaped figure at the bottom represents the confidence interval of $\hat{\rho}$.

Inspection of the funnel plot gives no hint for a publication bias. This result seems to be reasonable because in the present meta-analysis the effect measure $\bar{r}$ is usually significant. Furthermore, with respect to the present research questions, studies with lower mean correlations should not have a higher probability to be rejected. Also, only three potential outliers are present throughout the 408 studies, which, after their elimination, do not affect the estimation of $\hat{\rho}$ or other relevant measures. For this reason, the interpretation of results does not change either, and we therefore decided to keep such studies.

3.2. Meta-analyses with the ability moderator

Next, we studied the mean Intelligence Quotient ($\overline{IQ}$) as predictor of $\hat{\rho}$. To this respect, the following regression equation and its parameters are significant ($p < .001$): $\hat{\rho} = .6493 - .0033\ \overline{IQ}$. Then we were interested in studying the presence of a U-shaped effect together with other effects by also adding a quadratic (i.e., $\overline{IQ}^2$) and, furthermore, a cubic (i.e., $\overline{IQ}^3$) term in the latter equation. Results of these more-complex regressions are, respectively, $\hat{\rho} = 1.0328 - .0110\ \overline{IQ} + 4 \times 10^{-5}\ \overline{IQ}^2$, and $\hat{\rho} = .9334 - .0076\ \overline{IQ} - 8 \times 10^{-5}\ \overline{IQ}^2 + 1 \times 10^{-7}\ \overline{IQ}^3$, all parameters of both equations being non-significant, except for the intercepts ($p < .05$). Moreover, the squared and cubic terms get very small and negligible regression

[1] We kindly thank Prof. Dr. Detterman for his valuable advice on quality assessment.

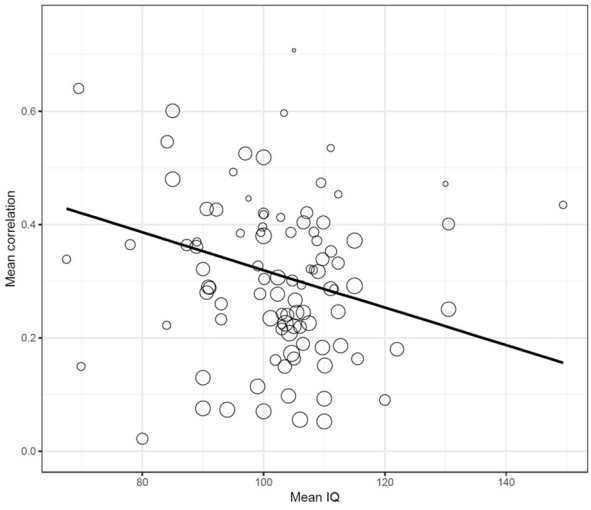

Fig. 2. Plot of the relation between $\overline{IQ}$ levels (*x*-axis) and $\overline{r}$ levels (*y*-axis) with thick linear regression line. Dot sizes are a function of model $\overline{r}$ weights.

weights. Therefore, the simplest equation excluding these two additional terms is considered to be an adequate model. This means that a linear relation is present, with a tendency of $\hat{\rho}$ being smaller for higher $\overline{IQ}$'s. Fig. 2 plots the relation between $\overline{IQ}$ values (*x*-axis) and $\overline{r}$ values (*y*-axis) with a linear regression line.

From Fig. 2 it can be seen that there is one study with $\overline{IQ}$ above 140. This is an influential observation, and the question arises whether such a high *IQ* was really given because intelligence tests usually do not allow for a reliable measurement of such *IQ*. Leaving out that observation leads to the regression equation $\hat{\rho} = .6730 - .0036\ \overline{IQ}$ with both estimates being significant ($p < .001$). Thus, this slightly higher relationship between the *g* saturation and $\overline{IQ}$ seems to be reasonable. Inclusion of the quadratic and cubic terms when omitting the influential case once again leads to non-significant slopes as in the case where all observations were included.

Then we reanalyzed the data regarding the $\overline{IQ}$ predictor by considering subsamples that differ on the basis of methodological quality. The quality-scale distribution that can be seen on Table 1 has a mean of 1.25 ($SD = 0.72$), a median of 1.50, and a first and third quartiles of 0.50 and 1.75 respectively. Even though the scale can reach a maximum level of 4, all given values are lower than 3.

We decided to rerun the results by considering three different subgroups, where subgroup 1 (SG1) comprises studies with quality values below the median, subgroup 2 (SG2) acquires studies with quality scores equal to or greater than the median, and subgroup 3 (SG3) consists of the methodologically best studies (i.e., those above the third quartile). Reanalyzing the data with the mentioned subsamples leads to a non-significant $\overline{IQ}$ regression weight for SG1 ($\hat{\rho} = .1007 - .0023\ \overline{IQ}$), and to significant regression equations as well as regression weights for the other two subgroups (SG2: $\hat{\rho} = .8265 - .0051\ \overline{IQ}$; SG3: $\hat{\rho} = .8550 - .0052\ \overline{IQ}$; $p < .001$). When the influential case is removed, regressions do not change greatly. Hence, even though quality scores are not generally high, results reflect a significant tendency towards greater study quality, which is expected. The $\overline{IQ}$ slope which the authors of the present paper attribute to be of highest confidence is that of SG2: −.0051. Therefore, Fig. 3 plots the same relation seen in Fig. 2 but only considering SG2. Table 2 shows the predicted effect sizes $\hat{\rho}$ for given $\overline{IQ}$ values, stemming from the earlier simple regressions.

Table 1
Frequency distribution of the quality scale.

Value	Frequency	Percentage
0	33	8.09
0.50	78	19.12
0.75	24	5.88
1	65	15.93
1.50	76	18.63
1.75	42	10.29
2	62	15.2
2.50	10	2.45
2.75	18	4.41

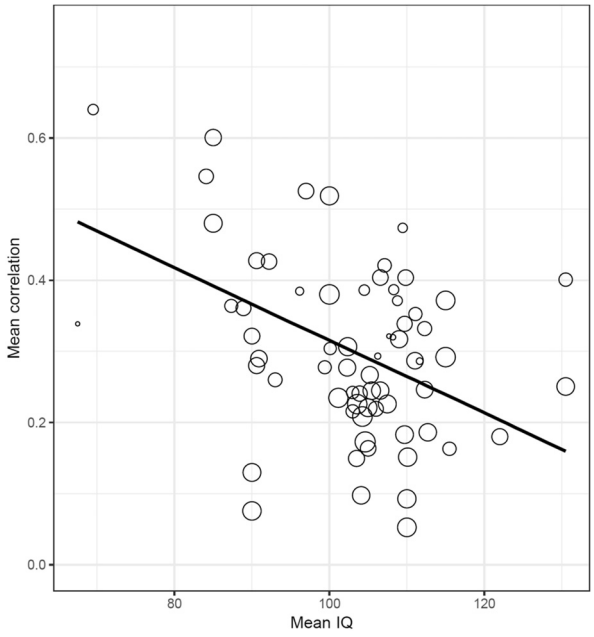

Fig. 3. Plot of the relation between $\overline{IQ}$ levels (*x*-axis) and $\overline{r}$ levels (*y*-axis) with thick linear regression line, when SG2 is only considered. Dot sizes are a function of model $\overline{r}$ weights.

3.3. Meta-analyses with the age moderator

The second continuous moderator considered was mean age ($\overline{a}$). The following regression model as well as its parameters are significant ($p < .001$): $\hat{\rho} = .3085 + .0012\,\overline{a}$. The inclusion of $\overline{a}^2$ in the regression equation returns the following result: $\hat{\rho} = .3365 - .0013\ \overline{a} + 3 \times 10^{-5}\ \overline{a}^2$, and adding $\overline{a}^3$ leads to: $\hat{\rho} = .3400 - .0018\ \overline{a} + 4 \times 10^{-5}\ \overline{a}^2 - 1 \times 10^{-7}\ \overline{a}^3$. Practically all of the parameters of these last two regression equations are not significant, except for the intercepts ($p < .001$). Therefore, a model with $\overline{a}$ as single linear predictor is considered to be adequate. Fig. 4 contains a plot of this relationship between $\overline{a}$ values (*x*-axis) and $\overline{r}$ values (*y*-axis).

Then we established a similar methodological quality assessment as the one described above. In all subgroups, results are nearly the same as before and also significant with respect to regression equations as well as the parameters (SG1: $\hat{\rho} = .3081 + .0012\,\overline{a}$; SG2: $\hat{\rho} = .3168 + .0011\,\overline{a}$; SG3: $\hat{\rho} = .2891 + .0009\,\overline{a}$; $p < .05$).

Table 2
Predicted effect sizes $\hat{\rho}$ for fixed $\overline{IQ}$ values.

$\overline{IQ}$	All 98 studies	All but $\overline{IQ} > 140$	Subgroup 1	Subgroup 2	Subgroup 3
70	.4183	.4210	−.0603	.4695	.4910
85	.3688	.3670	−.0948	.3930	.4130
100	.3193	.3130	−.1293	.3165	.3350
115	.2698	.2590	−.1638	.2400	.2570
130	.2203	.2050	−.1983	.1635	.1790
145	.1708	.1510	−.2328	.0870	.1010

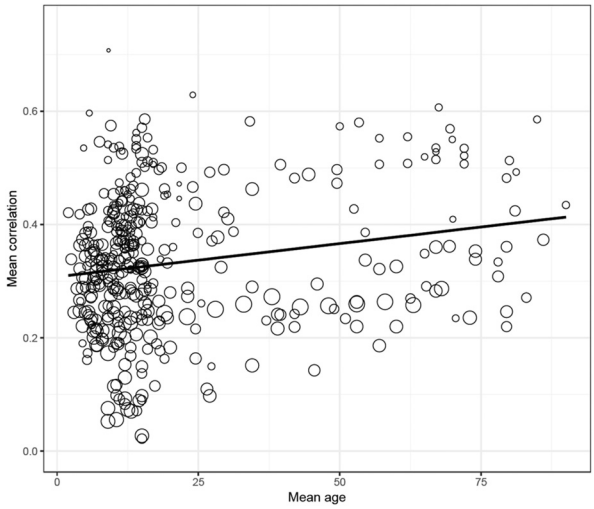

Fig. 4. Plot of the relation between $\bar{a}$ levels (*x*-axis) and $\bar{r}$ levels (*y*-axis) with thick linear regression line. Dot sizes are a function of model $\bar{r}$ weights.

3.4. Additional results

An important remark is that $\overline{IQ}$ and $\bar{a}$ correlate with $r = -.05$, meaning that they are nearly independent. As for the year of publication (*YP*), this predictor does not show a significant effect with $\hat{\rho}$ as criterion, nor is there evidence of a significant effect when *YP* is added to any of the aforementioned regression equations. This means that the effect size is not apparently affected by the moment in time where publications were released.

4. Discussion

The purpose of this paper is to establish a comprehensive study about the differentiation effect through meta-analytic modelling. For that matter, we dealt with data coming from heterogeneous approaches seen in the existing literature, which involves practically a century of research on the topic. The results reveal some tendencies which can be discussed in light of the questions governing this paper. Overall, the results are in line with the differentiation effect and with Spearman's expectations, as well as with some results of previous research.

An analysis without moderators provides support for the *g*-factor theory of intelligence developed by Spearman in 1904. It further suggests that abilities are only partially determined by the general composite (i.e., to approximately 41%). This is a reasonable result because recent evidence suggests the presence of about 10 group-level factors of intelligence which are independent from each other and, at the same time, dependent on a higher-hierarchy factor (*g*). These group factors have an impact on 50 to 60 narrower abilities too (Carroll, 2003). However, it should be remembered that this has historically been a matter requiring further investigation.

The first compound questions of this paper delved into the relation between *IQ*-indicated ability and the *g* saturation, and it is therefore essential to unmask the nature of ability differentiation. Spearman's idea was that the general ability composite declines with higher ability, with some evidence from Hartmann and Nyborg's review about ability-related effects. Our results show a linear decrease of the *g* saturation as a function of ability, which is even greater if only the 50% methodologically best studies are considered. According to this subgroup, for samples with a mean intelligence that is two standard deviations (i.e., 30 *IQ*-points) higher, the mean correlation to be expected is decreased by approximately .15 points. This is in line with Spearman's expectations as well as with some of Hartmann and Nyborg's results. Therefore, as ability advances, individuals tend to be less dependent on their general capacity to solve tasks, with abilities being therefore more differentiated.

The second compound questions of the present review delved into the relation between age and the *g* saturation, and it is relevant to understand the nature of age differentiation. Spearman (1927) revealed examples where older persons comprised a narrower estimate of this saturation, whereas a parallel theory extended to the entire life span suggested the presence of a U-shaped effect (Balinsky, 1941). The review of Hartmann and Nyborg (2006) showed the latter effect to be present, although they described it as a non-significant tendency, with the lowest *g* loadings occurring between ages 18 and 34. They also pointed out that contradictory results exist throughout the body of research since 1923 regarding the age predictor. On the other hand, our results suggest that an increasing linear effect of age on the *g* saturation amounting to approximately .001 is present, meaning that older people comprise a higher *g* saturation than younger people. E.g., for samples with a mean age that is 10 years higher than for others, an increased mean correlation of .01 is to be expected.

Hartmann and Nyborg underlined that effects with age as predictor could be attributed to ability effects and not necessarily to age itself, because of the reasons explained in the introduction section of the present paper. They also reminded that there is an inversely U-shaped trend implicated in the relation between age and ability, where ability is maximal at middle ages and then it starts to decline as age advances. For the period between middle age and senescence, the declination of ability could be explaining the rise of the *g* saturation seen in our data, which would, in principle, be in line with Spearman's (1927) expected trend as well as with the ability-differentiation hypothesis defined by Reinert, Baltes, and Schmidt in 1965. But for the period that goes from the first years of life until early maturity, there is a contradiction to this statement, because an important ability increase normally occurs within this period (Kalveram, 1965; Merz & Kalveram, 1965), and it follows that a parallel decrease of the *g* saturation should be expected. In fact, this decreasing effect was seen between younger and older children in Spearman's (1927) as well as in some of Hartmann and Nyborg's (2006) reviewed literature. However, our results show a linear and positive effect for the whole life span.

Two main disadvantages should be considered that restricted us from obtaining more representative age estimations for the period until 18 years, thus possibly clouding the real effects of age on the *g* saturation as a consequence. As for the first disadvantage, many studies reported the school grade level of students instead of their age, and we had to estimate the latter variable based on the age-grade correspondence according to the school system of the country of origin. This might be somewhat problematic because age levels and their percentages can present variations from grade to grade. Second, only working with the age mean of each group could have hidden the impact of within-group ability variations, given that individuals of a group still had different age levels. Unlike the time period above 18 years, during the first years of life ability abruptly changes as age advances, and two children or adolescents with similar ages can be at more distinct ability phases. However, the impact of these within-group differences on *g* loadings was not the main concern of researchers in most of the meta-analyzed bibliography, as between-group results were often given more attention. To summarize, the described disadvantages may have introduced some noise with respect to the mentioned age period, and this could have made results less reliable.

5. Conclusions and drawbacks

The present meta-analytic study provides support for the differentiation hypothesis. Results demonstrate that the mean correlation and *g* loadings of cognitive ability tests decrease with increasing ability, yet increase with respondent age. As for Spearman's (1927) expected trend, he did not evidently propose that a declination of *g* loadings as a function of age is present, but only as a function of ability; therefore, our results seem to broadly confirm this trend, indicated by a *g*-saturation decrease as a function of *IQ* as well as a *g*-saturation increase from middle ages to senescence.

As evidenced, a declination of the *g* saturation amounting to approximately .15 is expected from lower- to higher-ability groups distanced by 30 *IQ*-points. The question remains whether a difference of this magnitude could result in a greater apparent factorial complexity when cognitive data are factored for the higher-ability sample, as opposed to the lower-ability sample. It seems likely that greater factor dimensionality should tend to be observed for the case of higher ability, but the magnitude of this effect (i.e., how much more likely and how many more factors) remains uncertain. Furthermore, if we acknowledge that the *g* saturation (and, therefore, the cognitive structure of abilities) is not the same for different ability as well as age levels, this leads us to the question whether theories of intelligence need to be altered. In other words: is this difference between ability groups enough to impact the intelligence theory?

In this line of thinking, if a group of researchers claims to have identified new intelligence constructs with minimal *g* loadings, they should put their results into perspective by considering the outcomes here presented. Particularly, constructs have been found to be less *g*-loaded in higher-ability samples, so this could be a consequence of a mere sample effect (Legree, Mullins, Laport, & Roberts, 2016). Researchers at a major research university may find it much easier to collect data from a high-ability sample (e.g., university students), which does not necessarily represent the general population. Therefore, results may reflect sample effects, as opposed to the identification of a new construct.

Two main drawbacks of the present research should be discussed. First, with respect to ability, the reduced amount of publications reporting a common as well as standardized ability measure such as *IQ* is notorious, and it implies that results here presented about ability as predictor may not necessarily reflect what happens in the population taken as a whole. This inconvenience leads us to think that researchers should publish well-known ability measures like *IQ* more often in the future in order to make meta-analytic studies of the present kind more valid.

As for the second drawback, some researchers managed to publish results about the impact of several ability measures on performance themselves, but they did not publish the most elemental information about the *g* estimate (e.g., correlation matrix, first unrotated eigenvalue or its total explained variance). Whenever this information was absent or inconsistent, or whenever it could not be calculated by means of other existing data, the work in question had to be excluded from our meta-analytic investigation, which also means a loss of information. Hence, to help future meta-analytic studies attain a much more reliable state, any researcher intending to undergo a study on the present topic should, from now on, not forget to publish classic *g*-related statistical data, because this will enable results be comparable to most of the research performed across the XX century.

Acknowledgements

This work was financially supported by the following organizations:

- German Academic Exchange Service (Deutscher Akademischer Austauschdienst, DAAD).
- Universität Münster, Fachrichtung Psychologie, Arbeitseinheit Statistik und Methoden.

Appendices of the Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2017.07.004.

References

Amelang, M., & Langer, I. (1968). CZur Kritik der Divergenzhypothese der Intelligenz. *Archiv für die Gesamte Psychologie*, *120*, 203–217.
Anastasi, A. (1958). *Differential psychology*. New York: The Macmillan Company.
Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Stanford, CA: Wiley.
Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*, *23*, 191–234.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
Burt, C. (1949). The structure of the mind. A review of the results of factor analysis. *British Journal of Educational Psychology*, *19*(3), 176–199.
Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). New York: Pergamon Press.
Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.
Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*, 211–229.
R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.
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.
Escorial, S., Juan-Espinosa, M., García, L. F., Rebollo, I., & Colom, R. (2003). Does *g* variance change in adulthood? Testing the age de-differentiation hypothesis across sex. *Personality and Individual Differences*, *34*, 1525–1532.
Fogarty, G., & Stankov, L. (1995). Challenging the Law of Diminishing Returns. *Intelligence*, *21*(2), 157–176.
Garret, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, *1*(9), 372–378.
Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 357–376). (2nd ed.). New York: Russell Sage Foundation.
Hartmann, P., & Nyborg, H. (2006). Spearman's "Law of Diminishing Returns": A critical eye on a century of methods, results, and current standing of the theory. In P. Hartmann (Ed.), *Investigating Spearman's "Law of Diminishing Returns"* (pp. 31–190). Department of Psychology, University of Aarhus PhD thesis.
Hertzog, C., & Bleckley, M. K. (2001). Age differences in the structure of intelligence. Influences of information processing speed. *Intelligence*, *29*, 191–217.
Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions, version 5.1.0*. The Cochrane Collaboration. Available from www.handbook.cochrane.org.
Horn, J. L. (1976). Human abilities. A review of research and theory in the early 1970s. *Annual Review of Psychology*, *27*, 437–485.
Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, *28*(3), 195–226.
Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, *35*(1), 21–35.
Kaiser, H. F. (1968). A measure of the average intercorrelation. *Educational and Psychological Measurement*, *28*(2), 245–247.
Kalveram, K. T. (1965). Die Veränderung von Faktoren Strukturen durch "simultane Überlagerung" *Archiv für die Gesamte Psychologie*, *117*, 296–305.
Legree, P. J., Mullins, H. M., LaPort, K. A., & Roberts, R. D. (2016). SLODR-house rules: EI tests less *g* loaded in higher ability groups. *Intelligence*, *59*, 32–38.
Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's *g* in France. *Learning and Individual Differences*, *5*(1), 43–48.
Lynn, R., & Cooper, C. (1994). A secular decline in the strength of Spearman's *g* in Japan. *Current Psychology*, *13*(1), 3–9.
Merz, F., & Kalveram, K. T. (1965). Kritik der Differenzierungshypothese der Intelligenz. *Archiv für die Gesamte Psychologie*, *117*, 287–295.
Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*, 611–624.
Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In L. R. Goulet, & P. B. Baltes (Eds.), *Life-span Developmental Psychology* (pp. 467–484). London, UK: Academic Press.
Reinert, G., Baltes, P. B., & Schmidt, L. R. (1965). Faktorenanalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: die Leistungsdifferenzierungshypothese. *Psychologische Forschung*, *28*(3), 246–300.
Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292.
Spearman, C. (1927). *The abilities of man*. New York: MacMillan.
Thomson, G. H. (1919). The hierarchy of abilities. *British Journal of Psychology*, *9*(*3–4*), 337–344.
Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
Viechtbauer, W. (2010). Conducting meta-analysis in R with the metaphor package. *Journal of Statistical Software*, *36*(3), 1–48.