Intelligence 61 (2017) 130–139

Contents lists available at ScienceDirect

Intelligence

Checking the "Academic Selection" argument. Chess players outperform non-chess players in cognitive skills related to intelligence: A meta-analysis

Giovanni Sala [a,*,1], Alexander P. Burgoyne [b,1], Brooke N. Macnamara [c], David Z. Hambrick [b], Guillermo Campitelli [d], Fernand Gobet [a]

[a] University of Liverpool, United Kingdom
[b] Michigan State University, United States
[c] Case Western Reserve University, United States
[d] Edith Cowan University, Australia

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form 23 December 2016
Accepted 24 January 2017
Available online 3 February 2017

Keywords:
Chess
Intelligence
Cognitive ability
Deliberate practice
Meta-analysis

ABSTRACT

Substantial research in the psychology of expertise has shown that experts in several fields (e.g., science, mathematics) perform better than non-experts on standardized tests of intelligence. This evidence suggests that intelligence plays an important role in the acquisition of expertise. However, a counter argument is that the difference between experts and non-experts is not due to individuals' traits but to academic selection processes. For instance, in science, high scores on standardized tests (e.g., SAT and then GRE) are needed to be admitted to a university program for training. Thus, the "academic selection process" hypothesis is that expert vs. non-expert differences in cognitive ability reflect ability-related differences in access to training opportunities. To test this hypothesis, we focused on a domain in which there are no selection processes based on test scores: chess. This meta-analysis revealed that chess players outperformed non-chess players in intelligence-related skills ($\bar{d} = 0.49$). Therefore, this outcome does not corroborate the academic selection process argument, and consequently, supports the idea that access to training alone cannot explain expert performance.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Contents

1. Introduction . . . 131
1.1. Chess and expertise . . . 131
1.2. Purpose of the present study . . . 131
2. Method . . . 132
2.1. Literature search . . . 132
2.2. Inclusion criteria . . . 133
2.3. Description of the included studies (Table 1) . . . 133
2.3.1. Anderson (2004) . . . 133
2.3.2. Campitelli and Labollita (2016) . . . 133
2.3.3. Doll and Mayr (1987) . . . 133
2.3.4. Hänggi, Brütsch, Siegel, and Jäncke (2014) . . . 133
2.3.5. Unterrainer, Kaller, Halsband, and Rahm (2006) . . . 133
2.3.6. Unterrainer, Kaller, Leonhart, and Rahm (2011) . . . 133
2.3.7. Vidaurreta (2011) . . . 134
2.4. Effect size . . . 134
3. Results . . . 134

* Corresponding author at: Department of Psychological Sciences, Bedford Street South, University of Liverpool, L69 7ZA, United Kingdom.
E-mail address: Giovanni.sala@liv.uk.ac (G. Sala).
[1] Sala G. and Burgoyne A. P. are co-first authors of this article.

http://dx.doi.org/10.1016/j.intell.2017.01.013
0160-2896/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

3.1. Main model . . . 134
3.2. Outlier analysis . . . 134
3.3. Publication bias analysis . . . 134
3.4. Additional meta-analytic models . . . 136
4. Discussion . . . 137
4.1. What is the size of the "True Effect?". . . . 137
4.2. Limitations of the study. . . . 137
4.3. Recommendations for future research and conclusions. . . . 138
Acknowledgements . . . 138
References . . . 138

1. Introduction

Whether intelligence plays a significant role in determining expert performance has been intensely debated for more than a century. Starting with Galton's (1869) theory of hereditary genius, the idea that expert performance needs intelligence, and hence, that experts are on average more intelligent than laypeople, has received substantial support. Barron (1963) suggested that in some areas, such as physics and mathematics, the minimum threshold to achieve graduation is an IQ of 120. More recently, a longitudinal study showed that SAT scores of gifted children at the age of 13 predicted future success (e.g., number of scientific publications and patents) in their respective fields (Lubinski, 2009; Robertson, Smeets, Lubinski, & Benbow, 2010). Finally, in a meta-analytic review, Hunter and Hunter (1984) reported a strong positive correlation ($\bar{r} = 0.53$) between job performance and general intelligence (g). Moreover, the strength of this relationship was positively related to job complexity. These results were confirmed in successive meta-analyses (Hunter, Schmidt, & Le, 2006; Schmidt & Hunter, 1998; Schmidt, Shaffer, & Oh, 2008).

Ericsson (2014) has recently offered a competing view of expertise in which he argues that the difference in cognitive ability between experts and non-experts may be due to academic selection processes. Admission into undergraduate and postgraduate education programs is often based on test scores correlating with general intelligence measures (e.g., SAT for undergraduate studies, GRE for graduate studies, LSAT for law school). Therefore, Ericsson (2014) claims that if experts—such as scientists or musicians—score better on intelligence tests than non-experts, this may only be because individuals with higher cognitive ability had access to the training that made them experts.

Both views recognize the evidence that experts tend to have greater cognitive ability than their non-expert counterparts in many fields. However, they differ in what is assumed to account for the relationship between expertise and cognitive ability. The traditional view holds that cognitive ability is directly predictive of expertise. That is, experts have greater cognitive ability than non-experts because cognitive ability accounts for some of the variance in performance. The competing view (Ericsson, 2014) holds that cognitive ability predicts access to training, which in turn predicts expertise. That is, according to this view, experts have greater cognitive ability than non-experts because those with greater cognitive ability were selected for training; cognitive ability is not directly predictive of expert performance.

To test these competing views, we chose a field—chess—in which there is no selection process that limits access to training. According to the academic selection process hypothesis (Ericsson, 2014), if there are no barriers to training based on cognitive ability, then there should be no difference in cognitive ability between chess players who are training in this area of expertise and non-chess players who are not engaged in training.

1.1. Chess and expertise

Thanks to a reliable indicator of players' strength (Elo rating; Elo, 1978) and a balance between tactical calculation and strategic thinking (Hambrick et al., 2014), chess is widely considered an archetypical domain for the study of expertise, i.e., the ability to perform better than the majority of people in a particular domain (Gobet, 2016). In fact, to emphasize the idea that chess is an ideal model environment for research in the psychology of expertise, Simon and Chase (1973) indicated that the role chess plays in cognitive science is comparable to that of *Drosophila* (i.e., fruit fly) for the field of genetics.

Chess is also an excellent model environment to investigate the importance of cognitive ability and deliberate practice––the engagement in specific structured activities designed to improve performance in a field––in the acquisition of expertise (Gobet, 2016; Gobet, De Voogt, & Retschitzki, 2004; Grabner, 2014). Achieving mastery in chess requires intensive practice, but other factors seem to play an important role, too. For example, Campitelli and Gobet (2011) found that the amount of deliberate practice necessary to reach expertise in chess varied massively between players, suggesting that deliberate practice alone cannot explain individual differences in expert performance in chess. Since chess is a complex and intellectual activity, general intelligence and more specific cognitive abilities—such as visuospatial ability, short-term and working memory, planning, processing speed, and problem-solving skills—have been considered as potential predictors of individual differences in chess skill (Bilalić, McLeod, & Gobet, 2007; Burgoyne et al., 2016; de Bruin, Kok, Leppink, & Camp, 2014; Frydman & Lynn, 1992; Jastrzembski, Charness, & Vasyukova, 2006; Schneider, Gruber, Gold, & Opwis, 1993; Waters, Gobet, & Leyden, 2002).

Unlike other fields of expertise, access to chess practice and instruction does not depend on performance on standardized tests. That is, chess is an activity open to many people without restriction. Joining a chess club does not require any preliminary testing, and participation in courses and tournaments is often open to players of any level and relatively affordable. Certainly, a few chess players are funded by chess federations and admitted to elite tournaments, but the selection is always based mainly (if not completely) on players' chess performance. Chess is thus an ideal domain to test Ericsson's (2014) argument that differences in cognitive ability between experts and non-experts are due to performance on standardized tests limiting access to the domain, given that no such limitation applies to the field of chess.

1.2. Purpose of the present study

To test the academic selection hypothesis, we investigated whether cognitive ability differences emerge between those who have entered a field and those who have not in a field where there is no selection test: chess. We conducted a meta-analysis across all the studies meeting our inclusion criteria that measured cognitive ability. This meta-analysis examines studies of natural groups—those who have chosen to enter the field and those who have not[2]—and thus our results cannot address the causal structure of a relationship between chess player status and overall cognitive ability. However, finding a relationship would support the notion that

[2] For an analysis of the correlation between chess skill and intelligence in chess players of different skill levels, see Burgoyne et al. (2016).

cognitive ability is a strong candidate factor to play a role in expert performance and that expert vs. non-expert differences in cognitive ability cannot be assumed to only be due to academic selection processes.

2. Method

We designed the random-effects meta-analysis and report the results in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, & Altman, 2009). The entire procedure is summarized in Fig. 1.

2.1. Literature search

A systematic search strategy was used to find the relevant studies. Google Scholar, ERIC, Psyc-Info, JSTOR, Scopus, and ProQuest Dissertations & Theses databases were searched to identify potentially relevant studies. We used two specific combinations of keywords: (a) chess AND

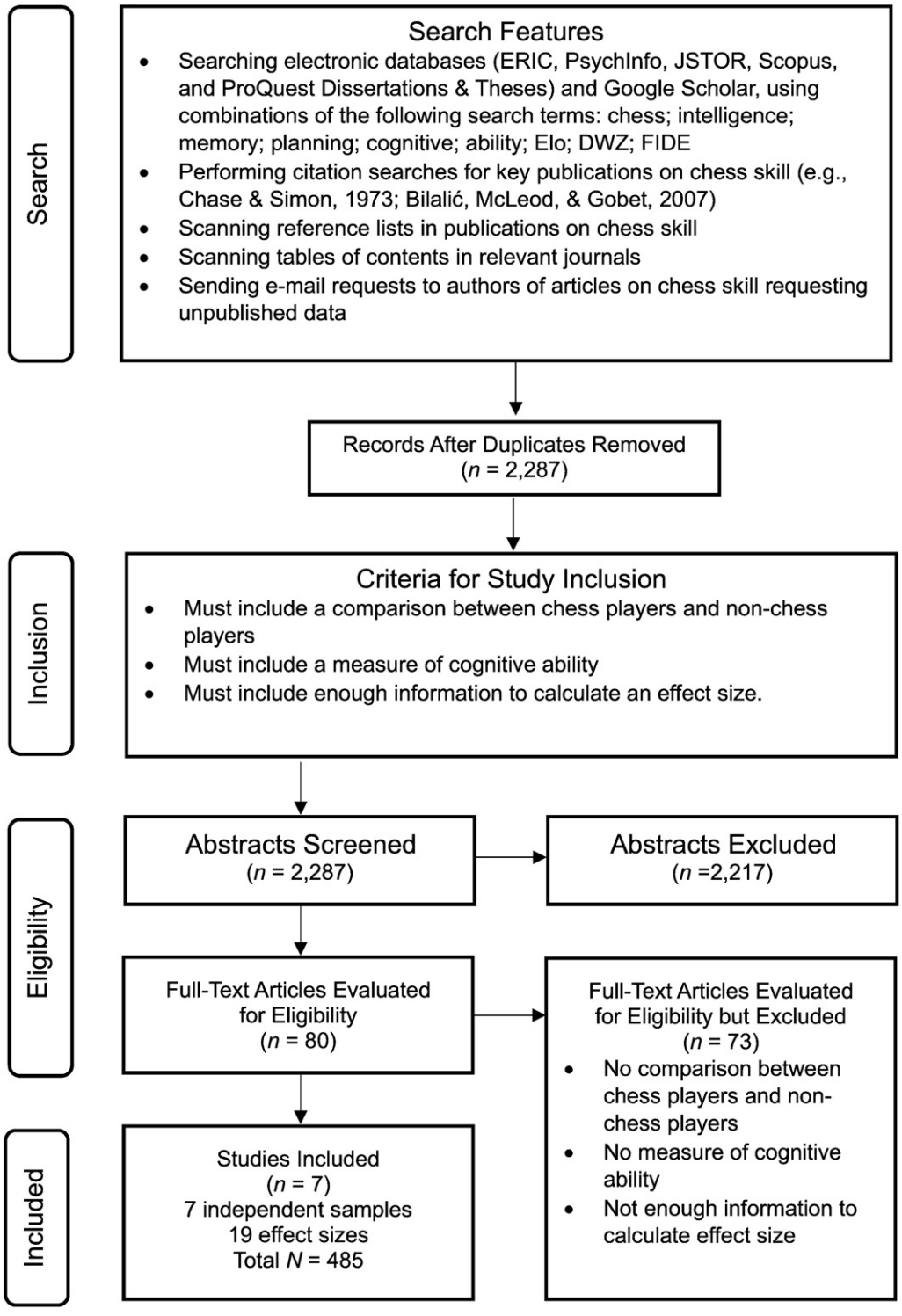

Fig. 1. Flow diagram of the studies included in the meta-analysis.

Table 1
The descriptive features of the seven included studies.

Study	Variable	Test	Age	Rating	Control group matching
Anderson (2004)	Processing speed	d2 Test of Attention	11–14	Unranked	Age and level of education
Campitelli and Labollita (2016)	Planning; cognitive reflection	Tower of London; cognitive reflection test	15–55; M = 27.8 (9.70)	Ranked (Elo): 2159–2607; M = 2405.4 (122.0)	Age and level of education
Doll and Mayr (1987)	Numerical; verbal; figural intelligence; processing speed; memory; creativity; information processing capacity	Berlin Intelligence Structure	18–51; M = 25.7 (4.9)	Ranked (Elo): 2220–2425; M = 2301.1 (54.3)	Age
Hänggi et al. (2014)	Fluid intelligence; mental rotation ability; visuospatial short-term memory	RAPM; V&K; block-tapping test	19–41; M = 28.9 (6.02)	Ranked (Elo): 2187–2560; M = 2366 (107)	Age and level of education
Unterrainer et al. (2006)	Planning; fluid intelligence; verbal working memory; visuospatial short-term memory	Tower of London; SPM; digit span; Corsi block-tapping test	M = 29.3 (8.6)	Ranked (DWZ[a]): 1250–2100; M = 1683.32	Age and level of education
Unterrainer et al. (2011)	Planning	Tower of London	20–50; M = 32.9 (9.1)	Ranked (DWZ[a]): 1209–2303; M = 1808.6 (272.8)	Age, level of education, and general intelligence
Vidaurreta (2011)	Problem-solving	WCST	7–11	Unranked	Age and level of education

[a] DWZ (Deutsche Wertungszahl) is the German equivalent of Elo rating.

(intelligence OR memory OR planning OR cognitive OR ability); and (b) (chess OR Elo OR DWZ OR Fide) AND (intelligence OR memory OR planning OR cognitive OR ability). In addition, previous narrative reviews were examined, and we e-mailed ($n = 7$) researchers in the field asking for unpublished studies and data.

2.2. Inclusion criteria

To be included in our meta-analysis the study needed to:

1. report a comparison between chess players and non-chess players;[3]
2. include at least one measure of intelligence;[4] and
3. report an effect size or the authors needed to provide enough information to calculate an effect size.

We found seven studies, conducted between 1987 and April, 1, 2016, that met all the inclusion criteria. These studies included seven independent samples and 19 effect sizes,[5] with a total of 485 participants.

2.3. Description of the included studies (Table 1)

2.3.1. Anderson (2004)

In this study, chess players ($n = 80$) and non-chess players ($n = 46$), aged 11–14 years, were administered the d2 Test of Attention (Brickenkamp & Zillmer, 1998). This test measures processing speed and discrimination of similar visual stimuli. No measure of chess skill was provided because the chess players, probably due to their young age, did not have an Elo rating (or equivalent). The participants were from two middle schools with open chess courses to their students. The chess players were those who had selected to enroll in a chess course. The chess players and non-chess players were matched on age and years of education.

2.3.2. Campitelli and Labollita (2016)

In this study, chess players ($n = 25$) and non-chess players ($n = 25$), aged 15–55 years, were administered the Tower of London (a measure of planning ability; Shallice, 1982) and the Cognitive Reflection Test (Campitelli & Gerrans, 2014). According to a commonly accepted metric (Elo > 2000; see Gobet & Simon, 2000), all the chess players can be considered chess experts, and the sample included some international grandmasters (Elo > 2500). The chess players and non-chess players were matched on age and level of education.[6]

2.3.3. Doll and Mayr (1987)

In this study, chess players ($n = 27$) and non-chess players ($n = 88$), aged 18–51 years, were administered the Berlin Intelligence Structure (BIS; Jäger, 1982), a test of general intelligence. The BIS consists of seven subscales measuring three content-related skills (verbal, numerical, and figural) and four operational skills (processing speed, memory, creativity, and information processing capacity). All the chess players in the sample were titled players (Masters, Fide Masters, and International Masters). The chess players and non-chess players were matched on age.

2.3.4. Hänggi, Brütsch, Siegel, and Jäncke (2014)

In this study, chess players ($n = 20$) and non-chess players ($n = 20$), aged 19–41 years, were administered the Raven's Advanced Progressive Matrices (Raven, 1998), a mental rotation task (Peters et al., 1995), and a block-tapping test (a measure of visuospatial short-term memory; Schellig, 1997). The chess players were all experts, and there were several international grandmasters (Elo > 2500). The two groups were matched on age and level of education.

2.3.5. Unterrainer, Kaller, Halsband, and Rahm (2006)

In this study, chess players ($n = 25$) and non-chess players ($n = 25$), mean age 29.3 years, were administered the Tower of London, the Standard Progressive Matrices (Raven, 1960), the digit span test, and the Corsi block-tapping test (Milner, 1971). Chess rating ranged from 1250 to 2100 (amateurs, intermediates, and experts). The chess players and non-chess players were matched on age and level of education.

2.3.6. Unterrainer, Kaller, Leonhart, and Rahm (2011)

In this study, chess players ($n = 30$) and non-chess players ($n = 30$), aged 20–50 years, were administered the Tower of London with a time

[3] Studies with experimental treatments were not included. See Sala and Gobet (2016) for a meta-analytic review of studies having incorporated an experimental chess treatment.

[4] The measures of cognitive ability included in the present meta-analysis are varied, but the majority correlate moderately or strongly with general intelligence. For example, the approximate correlation between full-scale IQ and Raven's Progressive Matrices is 0.80 (Jensen, 1998); between full-scale IQ and mental rotation is 0.41 (Ozer, 1987); between full-scale IQ and block-tapping is 0.38 (Orsini, 1994); between full-scale IQ and the Tower of London is 0.46 (Morice & Delahunty, 1996); between the SAT and the Cognitive Reflection Test is 0.44 (Frederick, 2005); between full-scale IQ and perseverative errors in Wisconsin Card Sorting Task is −0.30 (Ardila, Pineda, & Rosselli, 2000); and, between full-scale IQ and 2-choice processing speed tasks is 0.41 (Vernon, 1983).

[5] To account for effect sizes from dependent samples, we adjusted (i.e., lowered) the weight of such samples using the method designed by Cheung and Chan (2004). For a list of the adjusted *N*s, see the data file openly available at https://osf.io/t3uk2/.

[6] Note that the chess players received a compensation equivalent to US$12, whereas the non-chess players received unit credits valid for academic courses.

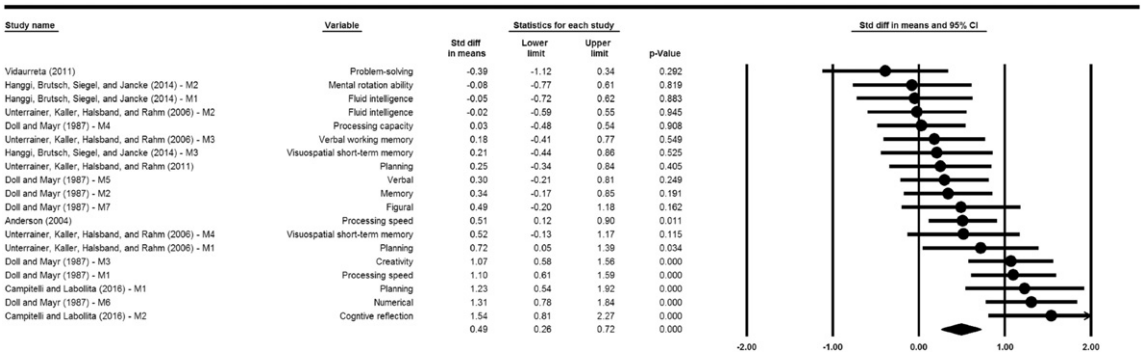

Fig. 2. Forest plot of the main model. Cohen's *d*s (circles) and 95% CIs (lines) are shown for all the effects entered into the meta-analysis. The *d*s are sorted by magnitude (from smallest to largest). The diamond at the bottom indicates the meta-analytically weighted mean $\overline{d}$. When studies used multiple outcome measures, the plot indicates the result of each measure (M1, M2, etc.) separately. These effect sizes were adjusted for dependent samples (see footnote 5).

constraint. The chess rating range was 1209–2303. The chess players and non-chess players were matched on age, level of education, and general intelligence.[7]

2.3.7. Vidaurreta (2011)

In this study, chess players ($n = 22$) and non-chess players ($n = 22$), aged 7–11 years, were administered the Wisconsin Card Sorting Task (WCST).[8] This test measures problem-solving and the ability to adapt to changing rules. Probably due to the young age of the participants, no measure of chess skill was available. The chess players were from a local chess club. The chess players and non-chess players were matched on age and years of education.

2.4. Effect size

The standardized mean difference (Cohen's *d*) in cognitive ability scores between the chess players and non-chess players was used as the effect size. The effect sizes were then corrected for attenuation due to measurement unreliability (Schmidt & Hunter, 2015) by using the two following formulas:

$$d' = \frac{d}{a} \quad (1)$$

$$Se' = \frac{Se}{a} \quad (2)$$

where d' is the corrected Cohen's *d*, Se' the corrected standard error (*Se*), and *a* the square root of the reliability coefficient (see the data file openly available at https://osf.io/t3uk2/ for a list of the coefficients and corrected effect sizes and standard errors). Finally, the Comprehensive Meta-Analysis (Version 3.0; Biostat, Englewood, NJ) and Metafor (Viechtbauer, 2010) software packages were used for conducting the meta-analytic, publication bias, and outlier analyses.

[7] While the groups were matched on general intelligence, an additional measure of cognitive ability was administered (Tower of London) and thus we included it. Given the pattern of results—chess players tending to have higher intelligence than non-chess players—it is likely that our overall effect size would be larger had the study authors not matched the chess players and non-chess players on general intelligence.

[8] Perseverative errors score is the only measure from the WCST correlating with measures of full-scale IQ (see Ardila, Pineda, & Rosselli, 2000). Thus, only the data concerning perseverative errors were used to extrapolate an effect size.

3. Results

3.1. Main model

The random-effects meta-analytic overall effect size was $\overline{d} = 0.49$, CI [0.26; 0.72], $k = 19$, $p < 0.001$. The forest plot is shown in Fig. 2. The I^2—a statistic that specifies the percentage of between-study variability in effect sizes due to heterogeneity rather than random error—was 66.56.

To evaluate whether any effect size exerted a strong effect on the meta-analytic overall effect size, we performed a one-study-removed analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009; Kepes, McDaniel, Brannick, & Banks, 2013). This analysis showed that the maximum difference between the overall effect size ($\overline{d} = 0.49$) and the several point estimates (min: $\overline{d} = 0.45$, CI [0.22; 0.67]; max: $\overline{d} = 0.53$, CI [0.31; 0.76]; Fig. 3) was small (± 0.04 standard deviations).

3.2. Outlier analysis

To test whether some effect sizes had an unusually large influence on the overall results, Viechtbauer and Cheung's (2010) outlier detection analysis was performed. No outliers were found.

3.3. Publication bias analysis

Publication bias occurs when studies with small samples and small effect sizes are systematically suppressed from the literature. To investigate whether our results were affected by publication bias, we examined a contour-enhanced funnel plot (Peters, Sutton, Jones, Abrams, & Rushton, 2008) depicting the relation between effect size and standard error, and performed Duval and Tweedie's (2000) trim-and-fill analysis. In the presence of publication bias, effect sizes are missing from the bottom left part of the funnel plot (small effect sizes with high standard error). That is, when standard error is high, larger-than-average effects sizes (those on the bottom right) are more likely to be published than smaller-than-average effect sizes (those on the bottom left). The trim-and-fill analysis estimates the number of missing studies from the funnel plot and imputes the missing effect sizes based on the observed data's asymmetry to create a more symmetrical funnel plot.

The funnel plot was approximately symmetrical around the meta-analytic mean ($\overline{d} = 0.49$; Fig. 4), suggesting no presence of publication bias. The trim-and-fill analysis estimated no missing smaller-than-average effect sizes either left or right of the mean.

Cumulative meta-analysis (Borenstein et al., 2009; Schmidt & Hunter, 2015) showed that the small-*N* effect sizes did not sensibly affect the overall effect size, suggesting no presence of publication bias (Fig. 5).

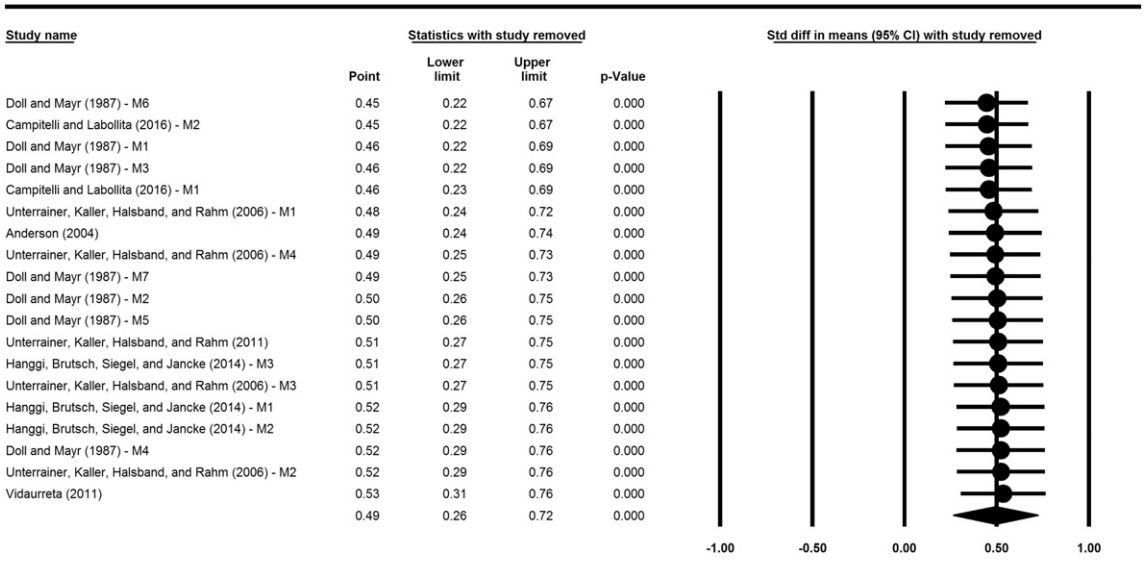

Fig. 3. Forest plot of the one-study-removed analysis. Each row reports the overall effect size (with 95% CI and *p*-value) that the model would estimate if the effect size of the corresponding study were removed from the analysis.

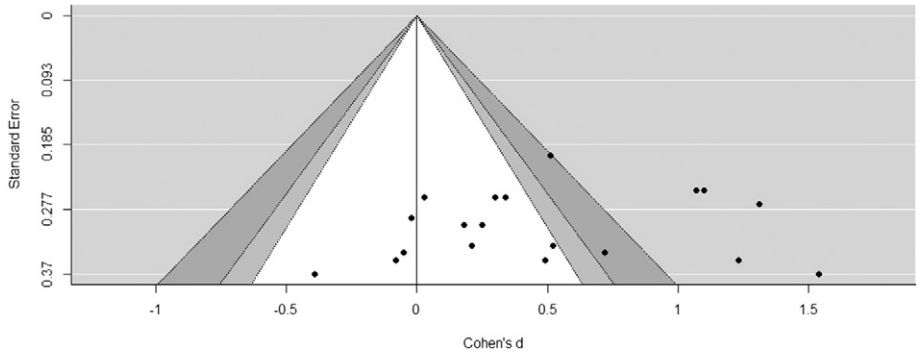

Fig. 4. Contour-enhanced funnel plot of standard errors and effect sizes (*ds*). The black circles represent the effect sizes included in the meta-analysis. Contour lines are at 1%, 5%, and 10% levels of statistical significance.

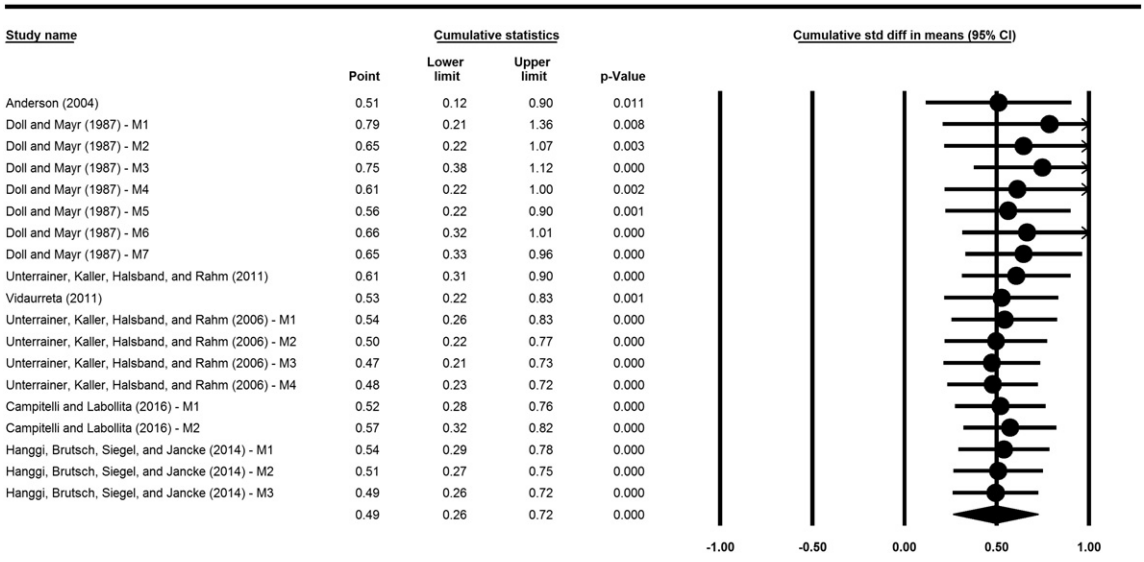

Fig. 5. Forest plot of the cumulative meta-analysis. The effect sizes are sorted by adjusted *N*s (from largest to smallest).

Table 2
Meta-analytic and publication bias outcomes.

Meta-analysis									Publication bias analyses				
N	*k*	$\overline{d}$	*95% CI*	*80% Cr. I*	*Q*	I^2	τ	*osr*	*t&f*	sm_m	sm_s	*PET*	*PEESE*
485	19	0.49	0.26, 0.72	0.00, 1.09	53.82	66.56	0.17	0.45, 0.53, 0.50	0.49	0.47	0.40	0.91 (0.28)	0.65 (0.35)

Note. N = number of the participants; k = number of the effect sizes; $\overline{d}$ = random effects meta-analytic mean; *95% CI* = 95% confidence intervals; *80% Cr. I* = 80% credibility intervals; Q = weighted sum of square deviations from the mean; I^2 = ratio of true heterogeneity; τ = between-sample standard deviation; osr = one-sample removed: minimum, maximum, and median overall effect sizes; *t&f* = trim-and-fill; sm_m = two-tailed moderate selection model; sm_s = two-tailed moderate selection model; *PET* = precision effect test (one-tailed *p*-value in brackets); *PEESE* = precision effect estimate with standard error (one-tailed *p*-value in brackets).

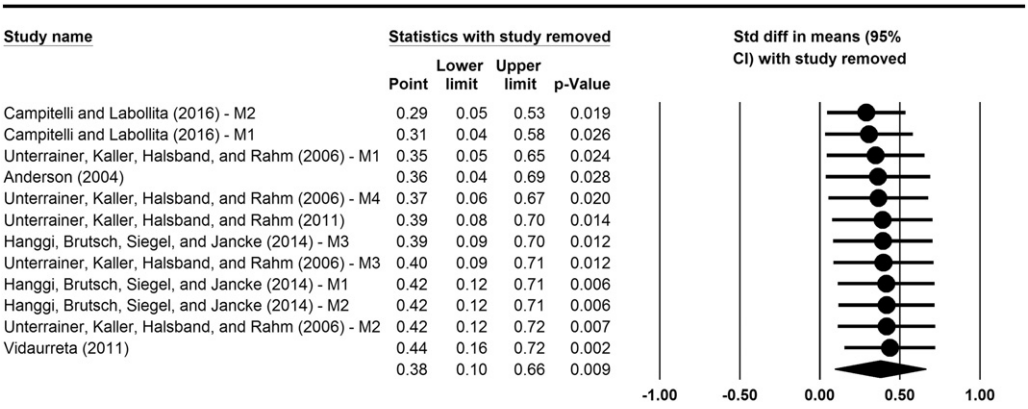

Fig. 6. Forest plot of the one-study-removed analysis without Doll and Mayr (1987). Each row reports the overall effect size (with 95% CI and *p*-value) that the model would estimate if the effect size of the corresponding study were removed from the analysis. The *d*s are sorted by magnitude (from smallest to largest).

Table 3
Meta-analytic and publication bias outcomes (without Doll & Mayr, 1987).

Meta-analysis									Publication bias analyses				
N	*k*	$\overline{d}$	*95% CI*	*80% Cr. I*	*Q*	I^2	τ	*osr*	*t&f*	sm_m	sm_s	*PET*	*PEESE*
370	12	0.38	0.10, 0.66	−0.18, 0.93	27.79	60.41	0.15	0.29, 0.44, 0.39	0.44[a]	0.35	0.30	0.35 (0.33)	0.08 (0.34)

Note. N = number of the participants; k = number of the effect sizes; $\overline{d}$ = random effects meta-analytic mean; *95% CI* = 95% confidence intervals; *80% Cr. I* = 80% credibility intervals; Q = weighted sum of square deviations from the mean; I^2 = ratio of true heterogeneity; τ = between-sample standard deviation; osr = one-sample removed: minimum, maximum, and median overall effect sizes; *t&f* = trim-and-fill; sm_m = two-tailed moderate selection model; sm_s = two-tailed moderate selection model; *PET* = precision effect test (its one-tailed *p*-value in brackets); *PEESE* = precision effect estimate with standard error (its one-tailed *p*-value in brackets).

[a] One effect size filled right of the mean. No missing effect sizes left of the mean.

Finally, using multiple methods for the detection of publication bias is important to test the robustness of results (Kepes & McDaniel, 2015). We thus ran two additional publication bias analyses: (a) selection models (moderate two-tailed selection and severe two-tailed selection;[9] Vevea & Woods, 2005); and (b) PET-PEESE (Stanley & Doucouliagos, 2014). The results of the two analyses showed either minimal differences between the overall effect size ($\overline{d} = 0.49$) and the point estimates, or non-significant point estimate (PET-PEESE). All the results are summarized in Table 2.

3.4. Additional meta-analytic models

Doll and Mayr (1987) was the only study not controlling for the level of education of the participants. Thus, we ran a random-effects meta-analytical model without the effect sizes extracted from that study. The random-effects meta-analytic overall effect size was $\overline{d} = 0.38$, CI [0.10; 0.66], $k = 12$, $p = 0.009$. The degree of heterogeneity was $I^2 = 60.41$. One-study-removed analysis showed that the maximum difference between the overall effect size ($\overline{d} = 0.38$) and the several point estimates (min: $\overline{d} = 0.29$, CI [0.05; 0.53]; max: $\overline{d} = 0.44$, CI [0.16; 0.72]; Fig. 6) was small (−0.09 standard deviations). Consistent with the one-study-removed analysis, Viechtbauer and Cheung's (2010) outlier detection analysis found no outliers.

With regard to publication bias analyses, the pattern of results was similar to the main model, that is small differences between the overall effect sizes and the point estimate (PET-PEESE point estimates were still not significant). All the results are summarized in Table 3, while cumulative meta-analysis is shown in Fig. 7.

Finally, four studies (i.e., Doll & Mayr, 1987; Campitelli & Labollita, 2016; Hänggi et al., 2014; Unterrainer et al., 2006) reported statistically dependent effect sizes ($k = 16$). To evaluate whether those effect sizes meaningfully affected the overall effect size, we calculated the weighted average effect sizes for three of these studies, while, for Doll and Mayr (1987), we used the reported full-scale measure of the BIS (see Section 2.3 and Table 1). We thus inserted the merged effect sizes in a new meta-analytic model. The random-effects meta-analytic overall

[9] The two-tailed selections were preferred over the one-tailed selections because of the particular features of the studies included in this meta-analysis. In fact, several studies used the data analyzed here as control variables (e.g., Hänggi et al., 2014), or tried to support the null hypothesis (i.e., no significant difference between chess players and non-chess players; Unterrainer et al., 2011). Moreover, some studies reported more than one variable regardless of the significance of the outcomes (e.g., Doll & Mayr, 1987). Thus, there is no reason to think that high *p*-values were more likely to be suppressed from the literature (as assumed by one-tailed selections).

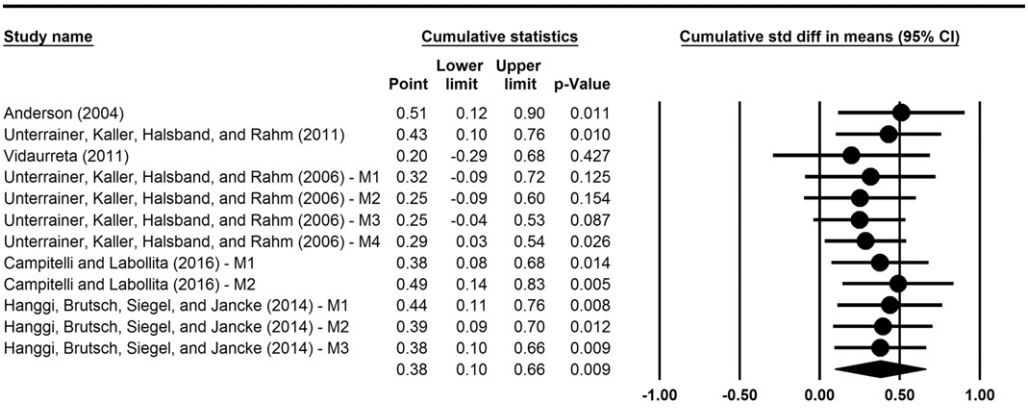

Fig. 7. Forest plot of the cumulative meta-analysis without Doll and Mayr (1987). The effect sizes are sorted by adjusted *Ns* (from largest to smallest).

effect size was slightly smaller but still significant, $\overline{d} = 0.39$, CI [0.06; 0.73], $k = 7$, $p = 0.022$. Due to scarcity of the effect sizes, no additional analysis was performed for this model.

4. Discussion

Ericsson (2014) argued that observed differences in cognitive ability between experts and non-experts are likely due to academic selection processes. That is, those who have higher cognitive ability are more likely to become experts because admission tests limit opportunities for those with lower cognitive ability to enter the field and receive training. Chess does not use academic selection mechanisms, which makes it an ideal domain to test whether cognitive ability differences emerge without the presence of confounding access limitations. We analyzed the performance of chess players compared to non-chess players in cognitive abilities related to general intelligence. The results demonstrated that chess players' overall cognitive ability is higher than age-matched comparison groups ($\overline{d} = 0.49$, $p < 0.001$). Importantly, the difference in cognitive ability between chess players and non-chess players does not seem to be due to level of education, because this variable was controlled for in almost all the reviewed studies. When the only study not controlling for level of education (i.e., Doll & Mayr, 1987) was excluded from the analysis, the overall effect size remained significant ($\overline{d} = 0.38$, $p = 0.009$). Thus, we found no evidence for Ericsson's (2014) claim that expert vs. non-expert differences in cognitive ability only reflect ability-related differences in access to training opportunities.

This finding, combined with the results of Burgoyne et al. (2016), a meta-analysis that found a significant positive correlation between chess skill and cognitive ability, provides evidence in favour of the idea that cognitive ability, to some extent, accounts for the acquisition of chess skill. Given that deliberate practice appears to be necessary, but not sufficient, to achieve high levels of expert performance in chess (Campitelli & Gobet, 2011; Hambrick et al., 2014), it is important to identify what other factors influence expertise acquisition in chess.

4.1. What is the size of the "True Effect?"

The advantage of chess players in overall cognitive ability remains statistically significant in all the meta-analytic models. However, the estimation of the effect size ranges between $\overline{d} = 0.49$ (i.e., the effect size of the main model) and $\overline{d} = 0.30$ (i.e., the point estimate from the severe selection model without Doll & Mayr, 1987): a not negligible difference of 0.19 standard deviations.

Evaluating the most likely estimate of the effect size depends mainly on the reliability of the results of Doll and Mayr (1987). Unfortunately, not enough information was provided to verify whether the educational level of the participants affected the results. If we assume that Doll and Mayr's (1987) results are trustworthy, then the probable mean estimate lies between 0.49 and 0.40 (the effect size of the main model and its smallest point estimate from publication bias analysis, respectively). If we assume that these results are unreliable, then the mean estimate is reasonably between 0.44 (i.e., the estimation of the trim-and-fill analysis) and 0.35 (i.e., the estimation of the two-tailed moderate selection model).[10] Either way, the size of the effect representing the superiority of chess players over non-chess players in overall cognitive ability appears to be approximatively medium (Cohen, 1988).

4.2. Limitations of the study

The total numbers of effect sizes ($k = 19$) and studies ($n = 7$; $n = 6$ without Doll & Mayr, 1987) are relatively small. For this reason, the outcome of the study, although statistically significant, must be interpreted with caution. In addition, although chess players outperformed non-chess players in almost all the reviewed cases, regardless of the participants' age and type of cognitive ability measured, the limited number of studies and effect sizes prevented us from testing for moderating effects of age and type of ability. For example, the difference between chess players and non-chess players may be more pronounced in youth than in adulthood (or vice versa). Moreover, this meta-analysis estimates an overall effect size from indicators of different cognitive abilities. In fact, the moderate degree of heterogeneity between effect sizes ($I^2 = 66.56$) indicates that the distribution is sufficiently homogenous to be aggregated in the same model (see footnote 4). Thus, the results support the hypothesis that chess players tend to have higher overall cognitive ability, but little can be inferred about the difference between chess players and non-chess players in particular cognitive abilities. For example, chess players may excel only in specific cognitive abilities (e.g., visuospatial ability, working memory) but not in measures of verbal reasoning. Additionally, it is yet to be determined whether the difference between chess players' and non-chess players' cognitive ability is stronger when non-chess players are compared to chess experts (Elo > 2000) and masters (Elo > 2200), or to amateurs and intermediate players.

Finally, our results specifically speak to the domain of chess. It is possible that the academic selection hypothesis holds in other intellectual domains. However, our results indicate that, at a minimum, the academic selection hypothesis does not hold for chess and therefore cannot be generalized to all intellectual domains. Future research is needed to test

[10] The value provided by the severe two-tailed selection model (i.e., 0.30) is very likely to be an underestimation. First, the trim-and-fill analysis found a missing study right of the mean (point estimate 0.44), which means the overall meta-analytic mean ($\overline{d} = 0.38$) may be an underestimation rather than an overestimation. Second, as mentioned in footnote 9, the assumption that non-significant effect sizes––especially the ones with *p*-values between 0.10 and 0.90––have been systematically suppressed from the literature is probably wrong. For these reasons, we consider the moderate selection model point estimate (i.e., 0.35) more reliable than the severe selection model one and sufficiently conservative.

whether the academic selection hypothesis accurately reflects expertise in any domain.

4.3. Recommendations for future research and conclusions

As mentioned above, the number of studies comparing chess players to non-chess players on cognitive skills is quite small. Additional studies are needed that compare chess players of different ages and ratings to matched non-chess players on a variety of cognitive tests. A larger number of studies examining variables such as level of skill, the age of players, and a broad range of cognitive abilities would allow us to test more sophisticated models and hypotheses.

Based on the available evidence, we demonstrated that chess players tend to perform better than non-chess players on measures of cognitive ability. Consequently, cognitive ability is a strong candidate to play an important role, together with deliberate practice, in skill acquisition in chess. Contrary to Ericsson's (2014) hypothesis, this difference in cognitive ability cannot be explained by academic selection processes, because no such processes operate in chess. As previously mentioned, deliberate practice alone is insufficient to account for differences in chess skill among chess players (Campitelli & Gobet, 2011; Hambrick et al., 2014). In line with that finding, this study suggests that cognitive ability may also play an important role. To determine whether the effect found in this study goes above and beyond the effect of deliberate practice on chess skill, studies should follow Bilalić et al.'s (2007) recommendation of measuring both deliberate practice and general abilities.

Acknowledgements

We thank Josef Unterrainer for providing unpublished data, and Sven Kepes and Tom Stanley for their assistance with the publication bias analyses.

References[11]

*Anderson, T. L. (2004). *The relation between gender, age, giftedness, and chess activity and attention in middle school student.* TX: Houston: University of Houston (Doctoral dissertation).

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology*, *15*, 31–36. http://dx.doi.org/10.1016/S0887-6177(98)00159-0.

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

Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence? A study with young chess players. *Intelligence*, *35*, 457–470. http://dx.doi.org/10.1016/j.intell.2006.09.005.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* London: Wiley.

Brickenkamp, R., & Zillmer, E. (1998). *The d2 Test of Attention* (1st US ed.). Seattle, WA: Hogrefe & Huber Publishers.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*, *59*, 72–83. http://dx.doi.org/10.1016/j.intell.2016.08.002.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*, 434–447. http://dx.doi.org/10.3758/s13421-013-0367-9.

Campitelli, G., & Gobet, F. (2011). Deliberate practice: Necessary but not sufficient. *Current Directions in Psychological Science*, *20*, 280–285. http://dx.doi.org/10.1177/0963721411421922.

*Campitelli, G., & Labollita, M. (2016). *Campitelli-Labollita_Datset_Chessplayers&Students.* figshare. http://dx.doi.org/10.6084/m9.figshare.3172813.v1.

Cheung, S. F., & Chan, D. K. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*, 780–791. http://dx.doi.org/10.1037/0021-9010.89.5.780.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale; NJ: Lawrence Erlbaum.

de Bruin, A. B., Kok, E. M., Leppink, J., & Camp, G. (2014). Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence*, *45*, 18–25. http://dx.doi.org/10.1016/j.intell.2013.07.004.

*Doll, J., & Mayr, U. (1987). Intelligenz und Schachleistung—Eine Untersuchung an Schachexperten [Intelligence and performance in chess—A study of chess experts]. *Psychologische Beiträge*, *29*, 270–289.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 276–284. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x.

Elo, A. (1978). *The rating of chessplayers, past and present.* New York: Arco.

Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in general population: A response to criticisms. *Intelligence*, *45*, 81–103. http://dx.doi.org/10.1016/j.intell.2013.12.001.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*, 25–42. http://dx.doi.org/10.1257/089533005775196732.

Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology*, *83*, 233–235. http://dx.doi.org/10.1111/j.2044-8295.1992.tb02437.x.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* London: MacMillan.

Gobet, F. (2016). *Understanding expertise: A multi-disciplinary approach.* London: Palgrave.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *24*, 651–682. http://dx.doi.org/10.1016/S0364-0213(00)00031-8.

Gobet, F., de Voogt, A., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games.* Hove, UK: Psychology Press.

Grabner, R. H. (2014). The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*, *45*, 26–33. http://dx.doi.org/10.1016/j.intell.2013.07.023.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, *45*, 34–45. http://dx.doi.org/10.1016/j.intell.2013.04.001.

*Hänggi, J., Brütsch, K., Siegel, A. M., & Jäncke, L. (2014). The architecture of the chess player's brain. *Neuropsychologia*, *62*, 152–162. http://dx.doi.org/10.1016/j.neuropsychologia.2014.07.019.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98. http://dx.doi.org/10.1037/0033-2909.96.1.72.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*, 594–612. http://dx.doi.org/10.1037/0021-9010.91.3.594.

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterenwicklung eines deskriptiven Intelligenzstrukturmodells. [Multimodal classification of intelligence achievement: Experimentally controlled further development of a descriptive structure model of intelligence]. *Diagnostica*, *28*, 195–225.

Jastrzembski, T. S., Charness, N., & Vasyukova, C. (2006). Expertise and age effects on knowledge activation in chess. *Psychology and Aging*, *21*, 401. http://dx.doi.org/10.1037/0882-7974.21.2.401.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kepes, S., & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS ONE*, *10*, e0141468. http://dx.doi.org/10.1371/journal.pone.0141468.

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, *28*, 123–143. http://dx.doi.org/10.1007/s10869-013-9300-2.

Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavior Genetics*, *39*, 350–358. http://dx.doi.org/10.1007/s10519-009-9273-0.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269. http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135.

Morice, R., & Delahunty, A. (1996). Frontal/executive impairments in schizophrenia. *Schizophrenia Bulletin*, *22*, 125–137. http://dx.doi.org/10.1093/schbul/22.1.125.

Orsini, A. (1994). Corsi's block-tapping test: Standardization and concurrent validity with WISC-R for children aged 11 to 16. *Perceptual and Motor Skills*, *79*, 1547–1554. http://dx.doi.org/10.2466/pms.1994.79.3f.1547.

Ozer, D. J. (1987). Personality, intelligence, and spatial visualization: Correlates of mental rotations test performance. *Journal of Personality and Social Psychology*, *53*, 129–134. http://dx.doi.org/10.1037/0022-3514.53.1.129.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test – Different versions and factors that affect performance. *Brain and Cognition*, *28*, 39–58. http://dx.doi.org/10.1006/brcg.1995.1032.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*, 991–996. http://dx.doi.org/10.1016/j.jclinepi.2007.11.010.

Raven, J. (1960). *Guide to the standard progressive matrices.* London: H. K. Lewis.

Raven, J. (1998). *Advanced progressive matrices.* Oxford: Oxford Psychology Press.

Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis. *Current Directions in Psychological Science*, *19*, 346–351. http://dx.doi.org/10.1177/0963721410391442.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46–57. http://dx.doi.org/10.1016/j.edurev.2016.02.002.

Schellig, D. (1997). *Block-tapping-test.* Frankfurt: Swets Test Services.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. http://dx.doi.org/10.1037/0033-2909.124.2.262.

[11] References marked with an asterisk indicate studies included in the meta-analysis.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Newbury Park, CA: Sage.

Schmidt, F. L., Shaffer, J. A., & Oh, I. S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, *61*, 827–868. http://dx.doi.org/10.1111/j.1744-6570.2008.00132.x.

Schneider, W., Gruber, H., Gold, A., & Opwis, K. (1993). Chess expertise and memory for chess positions in children and adults. *Journal of Experimental Child Psychology*, *56*, 328–349. http://dx.doi.org/10.1006/jecp.1993.1038.

Shallice, T. (1982). *The Tower of London test.* Colorado, US: Colorado Assessment Tests (CAT).

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 393–403.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60–78. http://dx.doi.org/10.1002/jrsm.1095.

*Unterrainer, J. M., Kaller, C. P., Halsband, U., & Rahm, B. (2006). Planning abilities and chess: A comparison of chess and non-chess players on the Tower of London task. *British Journal of Psychology*, *97*, 299–311. http://dx.doi.org/10.1348/000712605X71407.

*Unterrainer, J. M., Kaller, C. P., Leonhart, R., & Rahm, B. (2011). Revising superior planning performance in chess players: The impact of time restriction and motivation aspects. *American Journal of Psychology*, *124*, 213–225. http://dx.doi.org/10.5406/amerjpsyc.124.2.0213.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*, 53–70. http://dx.doi.org/10.1016/0160-2896(83)90006-5.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*, 428–443. http://dx.doi.org/10.1037/1082-989X.10.4.428.

*Vidaurreta, L. R. (2011). Aproximación al estudio de la flexibilidad cognitiva en niños ajedrecistas [Approach to the study of cognitive flexibility in children chess players]. *Rev. Cub. Med. Dep. & Cul. Fís.*, *6*.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48 Retrieved from http://brieger.esalq.usp.br/CRAN/web/packages/metafor/vignettes/metafor.pdf

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*, 112–125. http://dx.doi.org/10.1002/jrsm.11.

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities in chess players. *British Journal of Psychology*, *30*, 303–311. http://dx.doi.org/10.1348/000712602761381402.