Educational Psychology Review
https://doi.org/10.1007/s10648-022-09670-9

REVIEW ARTICLE

Chess Instruction Improves Cognitive Abilities and Academic Performance: Real Effects or Wishful Thinking?

Angel Blanch[1]

Accepted: 10 March 2022

Abstract

In accordance with the outcomes from a number of reports, there are cognitive and academic improvements derived from chess learning and chess playing. This evidence, however, endures three key limitations: (a) ignoring theoretical premises about the concept of transfer, (b) several shortcomings regarding ideal experiment guidelines, and (c) an uncritical faith in null hypothesis significance testing (NHST) statistical analyses. The present review scrutinized the NHST outcomes from 45 studies describing chess instruction interventions (n = 12,705) in nineteen countries that targeted cognitive ability (100 tests) and academic performance (108 tests), with a mean Hedge's effect size g = 572 (95% CI = [0.127, 1.062]). There was a lower average statistical power, a higher proportion of false positive outcomes, larger publication biases, and lower replication rates for the studies in the academic performance domain than in the cognitive ability domain. These findings raised reasonable concerns over the evidence about the benefits of chess instruction, which was particularly problematic regarding academic achievement outcomes. Chess should perhaps be regularly taught, however, regardless of whether it has a direct impact or not in cognitive abilities and academic performance, because these are far transfer targets. The more likely impact of chess on near transfer outcomes from higher quality studies remains at present unexplored.

Keywords Chess instruction · Cognitive ability · Academic performance

Chess requires an intensive management of cognitive abilities such as general sequential reasoning, long-term memory, and an extensive body of knowledge, and of other attributes such as will power and motivation (Blanch & Llaveria, 2021;

✉ Angel Blanch
angel.blanch@udl.cat

[1] Department of Psychology, Faculty of Education, Psychology and Social Work, University of Lleida, Avda de l'Estudi General, 4, 25001 Lleida, Catalonia, Spain

Published online: 02 April 2022

Cleveland, 1907; de Groot, 1965). Several studies highlight that learning and practising chess, particularly at early ages, has a meaningful impact on the improvement of cognitive abilities and the learning of curriculum subjects such as mathematics, science, and language in both regular teaching and special education (Costello, 2013; Storey, 2000; Thomasina & Adams, 2012).

In this view, several interesting properties of learning chess at early school periods are argued to enhance cognitive abilities, to improve learning academic subjects, and even to stimulate the development of social skills. Schoolchildren who learn or play chess appear to experience a better development of cognitive abilities or bearing a larger improvement in academic performance than schoolchildren who do not learn or play chess. Table 1 shows a sample of statements about the positive effects of chess instruction on either cognitive ability or academic achievement. In general, the accumulated evidence from these studies emphasize that the benefits of chess instruction are robust and highly replicable across different contexts and experimental conditions.

Past reviews, however, have cast a considerable degree of scepticism about the transference of skills acquired through chess to cognitive ability or academic achievement (Bart, 2014; Gobet & Campitelli, 2006; Sala & Gobet, 2016). Three main limitations undermine the reported positive impacts of chess instruction. These limitations consist in ignoring basic theoretical underpinnings about the concept of

Table 1 Claims highlighting support for the positive impact of chess instruction on cognitive abilities and academic achievement

Cognitive abilities	Academic performance
Chess not only improves cognitive capacities, it also influences socio-personal development and moulds the coping and problem-solving capacity in the children and adolescents who play chess (Aciego et al., 2012).	Taken together, these data strongly support the effectiveness of the chess program in improving student achievement and behaviour (DuCette, 2009).
This study adds evidence to the hypothesis that highly complex games like chess can favour the development of executive functions in childhood (Grau-Pérez & Moreira, 2017).	Using chess as an effective teaching method by associating it with the math lesson will contribute to the learning process of the lesson (Iskilgöz, 2016).
Chess intervention has led to significant IQ gains in both the chess-in-school program and the chess-in-academy training program (Joseph et al., 2017).	It is evident from the study that chess impacts cognitive development in children and there is a significant improvement in the academic performance of the children who underwent chess training (Joseph et al., 2016).
Chess instruction offered to children in pre-schools and kindergartens may contribute to the further development of some mental skills that are important for success in the early school years (Sigirtmac et al., 2012).	Replacing a weekly lecture of traditional mathematics with one based on chess instruction tended to increase subsequent results in math test scores (Rosholm, et al., 2017).
Training in the game of chess may play a significant role in enhancing general intelligence (Stegariu et al., 2019).	Chess seems to be an effective tool to promote mathematical problem-solving ability in primary school children, but only if the teaching includes chess problem-solving heuristics (Trinchero & Sala, 2016).

transfer, substantial deviations from an ideal experiment prototype, and a somewhat blind reliance on the null hypothesis significance testing (NHST) statistical analysis approach. For example, a meta-analysis with 24 studies and 40 effect sizes suggests a rather weak mean effect size (Hedge's g = .338, 95% CI [.242, .435]) regarding the transfer of chess instruction to academic and cognitive skills (Sala & Gobet, 2016), with larger effect sizes emerging for longer chess interventions and for studies reported in peer-reviewed journals. This latter moderator concerning the type of publication is particularly important, because it suggests that studies with significant outcomes might be selectively published while biasing the reported outcomes towards effect size inflation, unknowing statistical power, and impairing replication (Bartoš & Schimmack, 2020a; Schimmack, 2012; Simonsohn et al., 2014).

This study aims to extend and complement the meta-analysis by Sala and Gobet (2016) by conducting a quantitative analysis about the overall statistical power and replication potential of the studies addressing the impact of chess on cognitive abilities and academic performance. The present review describes in the first place the gist of each of the three aforementioned limitations with the focus on the NHST approach. This is followed by an in-depth analysis of the NHST outcomes from 45 studies that were conducted in nineteen countries (n = 12,705).

Transfer

Transfer implies that the changes in the ability to perform a task influence either the ability to perform the same task under different conditions, or the ability to perform a different task (Thorndike & Woodworth, 1901). Transfer can be formally described as $y = \varphi(x, p_x, p_y)$, where performance in the task y is a function of performance in the task x and the amount of practice (p_x, p_y) in the two tasks (Ferguson, 1954, 1956). Transfer is a central tenet of the aptitude by treatment interaction theory (ATI), which emphasizes individual differences in instructional processes with two main hypotheses (Snow & Lohman, 1984; Snow & Swanson, 1992). First, transfer effects depend on individual differences in cognitive abilities. Second, individual differences in performance depend on specific abilities and learning stages (Beier et al., 2010; Blanch, 2015; Blanch & Aluja, 2013; Deary et al., 2007).

A fundamental distinction arises between near and far transfer, depending on whether transfer occurs between tasks of similar (near) or dissimilar (far) contexts, respectively. Far transfer has attracted more research because it is more complex. A comprehensive taxonomy about far transfer organizes nine dimensions into two main factors, content (what is transferred) and context (when and where is transferred, from and to). The content factor comprises three dimensions, the specificity-generality of the learned skill, the measurement of the change in performance, and the memory demands of the task. The context factor comprises six dimensions, the knowledge domain, the physical, temporal, functional, and social contexts, and the modality of learning (Barnett & Ceci, 2002). According with this taxonomy, far transfer is a multidimensional intricate process unlikely to be grasped with effect sizes. Chess instruction studies, however, have primarily focused on far transfer outcomes (cognitive ability, academic achievement) while generally ignoring these content and context factors (Campitelli & Gobet, 2008; Gobet & Campitelli, 2006).

The ideal Experiment

Three research designs prevail in chess instruction studies. In the first design, a group of individuals is measured with a pre-test on some characteristic (i.e., cognitive ability, academic achievement), which is followed by the chess instruction intervention. Later, the same group is measured again with a post-test on the same characteristic. In the second design, there are two groups, a group undergoing a chess instruction intervention (experimental group), and a group without any exposition to the chess instruction intervention (control group). Both groups are measured on the same characteristic (i.e., cognitive ability, academic achievement) before and after the chess instruction intervention (pre and post-test). In the third design, two groups are also measured on cognitive or academic outcomes, even though one of them has some experience in chess playing, whereas the other group has none or very limited experience in chess playing. In this latter design, there is usually no chess intervention. The quantitative findings from either research design are subsequently analyzed with null hypothesis significance testing (NHST) methods, as described in the next section.

Earlier reviews express, however, several concerns about the quality and implementation of such research designs to account for the eventual benefits of chess (Bart, 2014; Gobet & Campitelli, 2006; Sala et al., 2017). These include the lack of a random assignment of participants to treatment and control groups, the omission of statistical corrections for multiple tests, the scarcity of longitudinal studies addressing long-lasting transfer effects, missing the description of chess instructor's individual characteristics, and a poor degree of replication efforts. In the light of these concerns, Gobet and Campitelli (2006) recommend undertaking several improvements, which comprehend controlling for placebo effects and chess instructor's personality, publication in peer-reviewed journals, avoiding selective and too enthusiastic conclusions, and using a methodology closer to an ideal experiment. This ideal experiment approach comprises a random allocation of participants to either experimental or control groups, more refined pre-test and post-test conditions, collecting also data from placebo and do-nothing control groups, different individuals conducting the chess teaching and the subsequent assessment, and keeping the subjects blind to the specific experiment.

Null Hypothesis Significance Testing

Owing to the predominance of the research designs described earlier, the studies addressing the effects of chess instruction interventions or chess experiences have adopted a null hypothesis significance testing (NHST) statistical analysis framework. The positive effects of chess instruction interventions or chess playing are inferred from NHST (i.e., t-test, F test) of the target areas (cognitive, academic). In general, a positive effect is claimed when detecting statistically significant improvements in the second measurement in the first design, an improvement in the experimental group compared with the control group in the second design, and larger scores in the chess experienced group compared with the no chess experience group in the third design.

Figure 1 shows the basic decision-making schema that relates the statistical decision with the real effect in NHST. The statistical decision from a specific test contemplates either assuming that there is no statistically significant difference as stated by the null hypothesis (H_{null}), or assuming that there is a statistical significant difference represented by the alternative hypothesis ($H_{\text{alternative}}$). Eventually, two types of errors emerge within this framework, usually termed as type I (α) and type II (β) errors. The type I error (α) represents the probability of making a statistical decision that sustains a difference while there is no real difference, i.e., a false positive. The type II error (β) represents the probability of making a statistical decision that sustains no difference while there is a real difference, i.e., a false negative. Smaller type I errors (α) lead to a higher likelihood of correctly assuming no statistical significant differences, and thus correctly accepting the null hypothesis when there are indeed no real differences ($1–\alpha$).

A typical type I error used by most researchers in educational psychology is set up at $\alpha = 0.05$. This bound implies a reasonable balanced trade-off with type II errors because type I and type II errors are inversely related (Cohen, 1988; Hair et al., 2010). Lower type I errors lead to higher type II errors that decrease statistical power ($1–\beta$), a crucial point generally neglected in chess instruction studies. Statistical power is important because it actually indicates the probability of detecting a true effect (Schimmack, 2012). Apart from type

		Real effect	
		No difference	Difference
Statistical decision	H_{null}: No difference	$1 - \alpha$	Type II error β
	$H_{alternative}$: Difference	Type I error α	Statistical power $1 - \beta$

Fig. 1 Two types of errors in null hypothesis statistical testing (NHST)

I error, two additional factors govern statistical power, effect size and sample size. The effect size describes the magnitude of the studied phenomenon, which is usually gauged in standardized units, i.e., a value of 0.5 indicates an effect size of half a standard deviation. Larger sample sizes contribute in addition to increasing statistical power at any level of α.

The NHST approach suffers from several shortcomings. These include but are not limited to overlook type I error rates, describe the results in terms of "significant" versus "non significant", accept the null hypothesis with high *p* values, a limited choice between two mutually exclusive hypotheses, missing effect sizes and confidence intervals, and conflating statistical with substantial significance (Miller, 2017; Sun et al., 2010; Vacha-Haase, 2001). In addition, misusing the NHST approach may eventually lead to ignoring multiple testing, publication bias, and lessen replication (Häggström, 2017; Open Science Collaboration, 2015). It is therefore somehow paradoxical that while the findings from chess instruction studies suggest robust and highly replicable positive effects, this evidence could be indeed biased because of the NHST data analyses approach conducted with these studies.

The Present Study

There is a considerable degree of consensus in that chess instruction exerts a positive impact on cognitive abilities or school academic subjects. Several vehement claims assume that the effect is robust and replicable across different studies with varying experimental conditions (see Table 1). On empirical grounds, however, a meta-analysis about the transfer of chess instruction on cognitive and academic skills suggested mild effect sizes that could in addition be boosted by placebo effects, while being moderated by the publication type, that is, by whether the findings were reported on peer-reviewed journals (Sala & Gobet, 2016). Indeed, this meta-analysis acknowledged the higher likelihood of studies with significant outcomes to become published, a problem that might bias the reported findings by inflating the corresponding effect sizes and impairing statistical power and replication (Bartoš & Schimmack, 2020a; Schimmack, 2012; Simonsohn et al., 2014).

The robustness of the supportive evidence about the benefits of chess instruction, which is apparently consistent and replicable, was therefore evaluated here with the *z*-curve application. This review sought to extend the meta-analysis by Sala and Gobet (2016) in two main ways. First, this review included in addition new reports in peer-reviewed journals beyond 2016, and other studies such as doctoral thesis or reports that were available during the online literature search and that met the specific selection criteria. Second, rather than evaluating effect sizes and potential moderator effects, the current review examined the *p* values associated with the NHST outcomes in the reported tests of the selected studies. Several indices associated with statistical power were obtained by contrasting all the studies that were actually reported with those that reported statistically significant results (Bartoš & Schimmack, 2020b; Brunner & Schimmack, 2020).

Method

Literature search and study selection

Figure 2 shows the literature search and the study selection process (Moher et al., 2009). The relevant studies for the aims of this research were sought and identified by searching the Psych-Info, PubMed, and Google Scholar databases. The search strategy comprised key words with the terms “chess instruction”, “chess teaching”, or “chess in school”. In addition, previous reviews, meta-analyses, and tables of contents in specialized journals were also examined. This identification stage yielded 230 available records after removing duplicates.

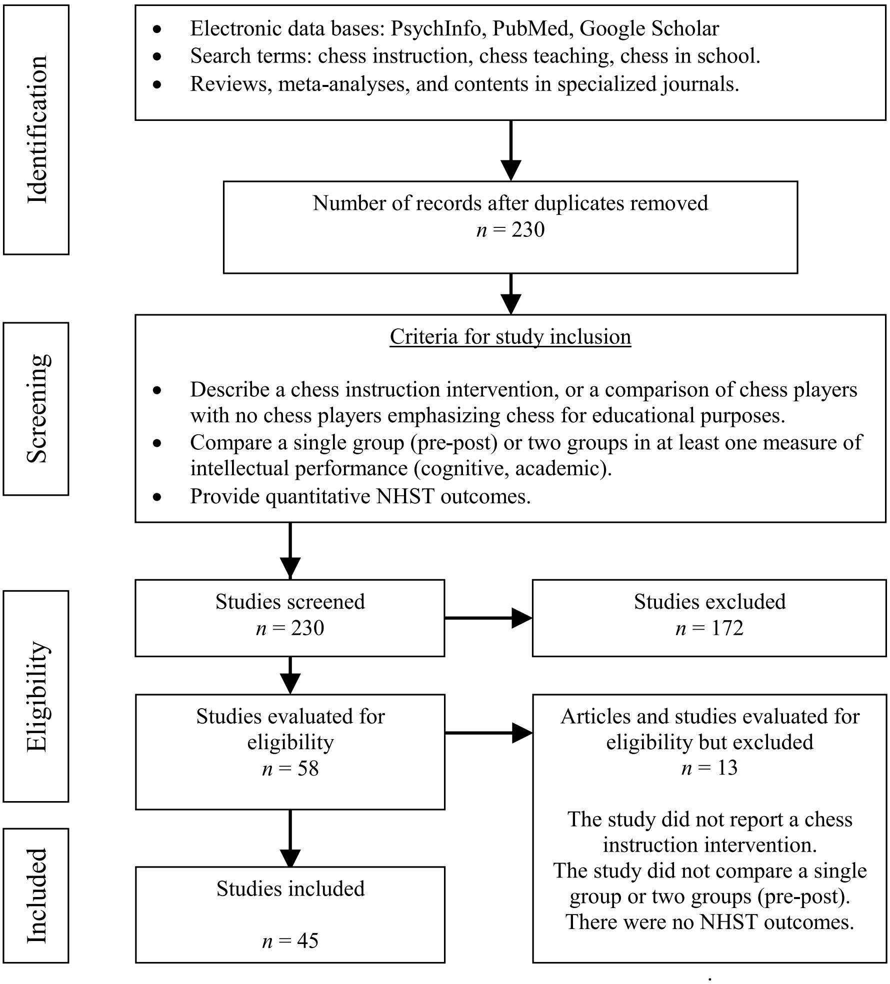

Fig. 2 Flow diagram of literature search and study selection (NHST: Null hypothesis significance testing)

In the screening stage, there were three main criteria for the inclusion of a study in the subsequent analyses. First, the study should describe a chess instruction intervention or a comparison of chess players with no chess players emphasizing the benefits of chess for educational purposes. Second, the study should compare a single group (pre-test, post-test) or two groups before and after the chess instruction intervention in at least one measure of cognitive ability or academic performance. Third, the study should provide quantitative outcomes obtained from a NHST approach. An initial selection of studies according with these inclusion criteria led to excluding 172 from the 230 records selected in the identification stage. The remaining 58 studies were additionally evaluated for eligibility. Thirteen studies were excluded at the eligibility stage because of not fulfilling one or more of the inclusion criteria. The final amount of included studies for the subsequent analyses was 45.

Table 2 summarizes these 45 selected studies, indicating the sample size, the main target areas (cognitive ability, academic subjects), the country where it was conducted, and whether it reported positive or null effects of the chess instruction intervention on the target area. Overall, there was a total sample size of n = 12,705 from nineteen countries (Argentina, Australia, Bangladesh, Belgium, Canada, Denmark, Germany, India, Iran, Italy, Malaysia, Romania, South Africa, South Korea, Spain, Turkey, UK, Uruguay, and USA). There was a considerable variability in the sample sizes ranging between 12 and 3865 individuals. The mean sample size was 282 (Sd = 627), with a median of 86 individuals, indicating that sample sizes tended to be rather low. The last column in Table 2 indicates whether the examined studies reported either positive or null effects of chess on the target area. Seven out of the 45 studies reported null effects (16%), whereas 38 out of the 45 studies (84%) reported positive effects.

Data Analyses

Every single study was screened to extract the quantitative NHST outcomes. Overall, there were 208 NHST reported outcomes, 100 tests for cognitive ability, and 108 tests for academic performance (see Appendix 1). These tests contrasted the null hypothesis of equality in the effects of chess between different groups or measurements within their respective research designs.

These data were analyzed with the z-curve package from the R software (Bartoš & Schimmack, 2020b; R Core Team, 2019). The procedure implemented by the z-curve package allows for the evaluation of the mean statistical power of heterogeneous significance tests outcomes, and the degree of discovery and replicability of a group of studies through a few indices. The observed discovery rate (ODR) is the proportion of reported studies that produced significant results. The expected discovery rate (EDR) estimates the unconditional mean power of all conducted studies, regardless of whether produced statistical significant outcomes. Higher discrepancies between the ODR and EDR suggest higher publication biases. The false

Table 2 Studies describing the effects of a chess instructional intervention on cognitive abilities and academic subjects in nineteen countries with 12,705 schoolchildren

Study	*N*	Main target areas	Country	Effects
Aciego et al., 2012	210	Cognitive ability	Spain	+
Atashafrouz, 2019	45	Cognitive ability	Iran	+
Aydin, 2015	26	Mathematics	Turkey	+
Barrett & Fish, 2011	31	Mathematics	USA	+
Basson, 2015	64	Cognitive ability	South Africa	+
Berkley, 2012	12	Critical thinking, mathematics	USA	+
Buyk & Ho, 2008	119	Mathematics	USA	+
Christiaen et al., 1981	40	Cognitive ability	Belgium	+
DuCette, 2009	151	Mathematics, reading	USA	+
Eberhard, 2003	153	Cognitive ability	USA	+
Fernández-Amigo, 2008	144	Mathematics	Spain	+
Forrest et al., 2005	54	Reading, mathematics, social adjustment	UK	+
Garcia, 2008	54	Mathematics, reading	USA	Null
Gliga & Flesner, 2014	38	Mathematics, language	Romania	+
Grau-Pérez & Moreira, 2017	28	Cognitive ability	Uruguay	+
Hong & Bart, 2007	38	Mathematics, reading, writing	South Korea	Null
Isikgöz, 2016	274	Mathematics	Turkey	+
Islam et al., 2021	569	Mathematics, science, language, social science	Bangladesh	+
Jerrim et al. 2018	3865	Mathematics, science, reading	UK	Null
Joseph et al., 2016	100	Mathematics, language, science	India	+
Joseph, et al., 2017	86	Cognitive ability	India	+
Joseph et al., 2018	151	Cognitive ability	India	+
Kazemi et al., 2012	180	Meta-cognitive abilities, mathematics	Iran	+
Khosrorad et al., 2014	20	Mathematics, cognitive ability	Iran	+
Liptrap, 1998	571	Reading, Mathematics	USA	+
Margulies, 1992	53	Reading	USA	+
Mel, 2021	53	Mathematics	Malaysia	+
Meloni & Fanari, 2019	85	Meta-cognitive abilities, mathematics	Italy	+
Ramos et al., 2018	65	Cognitive ability	Argentina	+
Rifner, 1992	18	Problem solving	USA	+
Romano, 2011	1756	Mathematics	Italy	+
Rosholm et al., 2017	482	Mathematics	Denmark	+
Sala & Gobet, 2017	233	Mathematics	Italy	Null
Sala et al., 2015	560	Mathematics	Italy	+
Sala et al., 2016	52	Mathematics, metacognitive skills	Italy	Null
Scholz et al., 2008	53	Mathematics, concentration	Germany	Null
Sigirtmac, 2012	100	Cognitive ability	Turkey	+
Sigirtmac, 2016	87	Creativity, theory of mind	Turkey	+
Smith, 1998	39	Mathematics	USA	+
Stegariu et al., 2019	67	Cognitive ability	Romania	+
Thompson, 2003	508	Scholastic performance	Australia	Null

Table 2 (continued)

Study	N	Main target areas	Country	Effects
Trinchero & Sala, 2016	931	Mathematics	Italy	+
Velea & Cojocaru, 2019	34	Cognitive ability	Romania	+
Voyer et al., 2018	185	Mathematics	Canada	+
Yap, 2006	321	Mathematics, reading	USA	+

discovery rate (FDR) estimates the maximum proportion of false positive results, i.e., rejecting the null hypothesis when it is true (type I error, α). The file drawer index estimates the proportion of unreported studies that produced non-significant results for every significant outcome. Finally, the expected replication rate (ERR) estimates the amount of studies with statistical significant outcomes that would replicate in an independent experiment. This indicator provides the mean power of the conducted studies with statistical significant outcomes. Compared with other methods, the z-curve method appears to yield accurate estimates with realistic circumstances such as heterogeneity in effect sizes across different studies and unknown population effect sizes (Brunner & Schimmack, 2020).

The input data were the p values corresponding to both cognitive ability and academic performance domains (see Appendix 1). First, z-values were obtained from the p values. After this transformation, there were 7 and 16 extreme z values ($z > 6$) for cognitive ability and academic performance, respectively. These extreme values were removed from the data set because z-scores above 6 correspond to $p = 0.000000002$, which if maintained in the analyses produce extreme long right tails of very rare results. Moreover, because z-curve assumes a perfect statistical power of 1 for z-values above 6, there is no need in estimating power for these values in accordance with the z-curve approach (Bartoš & Schimmack, 2020a). Second, two estimation methods were used to analyze these z values, expected maximization (EM) and kernel density (KD2). Moreover, there were 500 bootstrap replications to obtain 95% conservative confidence intervals (Bartoš & Schimmack, 2020a). Appendix 2 shows the R code used to analyze the input data.

Results

Descriptive Outcomes

Table 3 summarizes the included studies concerning the age range of the participants, the duration of the chess intervention in hours, whether the intervention was implemented by a chess instructor, the type of research design, whether there was a random assignment to either treatment or control groups, and the Hedge's g effect size with 95% confidence intervals.

Table 3 Age range, average duration of the chess intervention, research design, and effect sizes of the included studies. Hedge's *g* effect sizes are reported according with the larger effects reported in Sala & Gobet (2016) or calculated for the new studies in this review. Rand = random assignment to treatment and control groups

Study	Age range	Duration (hours)	Chess Instructor	Research design	Rand	Hedge's *g*	95% CI
Aciego et al., 2012	6–16	96	Yes	Treatment - control (pre, post)	No	0.118	– 0.185, 0.422
Atashafrouz, 2019	15–16	15*	Yes	Treatment - control (pre, post)	Yes	0.184	– 0.437, 0.805
Aydin, 2015	11–18	48	Unreported	Treatment - control (pre, post)	No	1.657	0.836, 2.478
Barrett & Fish, 2011	11–14	25	Teacher	Treatment - control (pre, post)	No	0.232	0.457, 2.007
Basson, 2015	4–6	5	Researcher	Treatment - control (pre, post)	No	0.074	– 0.417, 0.565
Berkley, 2012	18–54	10*	Researcher	Treatment - control (pre, post)	No	1.151	0.288, 2.015
Buyk & Ho, 2008	6–14	60	Researcher	Treatment (pre, post)	No	1.071	0.798, 1.342
Christiaen et al., 1981	10–11	42	Unreported	Treatment - control (pre, post)	Yes	0.410	– 0.186, 1.006
DuCette, 2009	6–15	Unknown	Unreported	Treatment - control (pre, post)	No	0.373	0.138, 0.608
Eberhard, 2003	12–14	60	Teacher	Treatment - control (pre, post)	No	– 0.028	– 0.382, 0.326
Fernández-Amigo, 2008	6–8	62	Researcher	Treatment - control (pre, post)	No	0.379	0.049, 0.709
Forrest et al., 2005	7–8	37	Yes	Treatment - control (pre, post)	No	0.236	– 0.368, 0.841
Garcia, 2008	10–11	90	Yes	Treatment - control (pre, post)	Yes	0.364	– 0.221, 0.948
Gliga & Flesner, 2014	9–10	10	Yes	Treatment - control (pre, post)	Yes	0.563	– 0.072, 1.198
Grau-Pérez & Moreira, 2017	7–12	Unknown	Yes	Chess - no chess	No	0.841	0.068, 1.614
Hong & Bart, 2007	8–12	20	Yes	Treatment - control (pre, post)	Yes	0.152	– 0.473, 0.777
Isikgöz, 2016	10–14	Unknown	Yes	Chess - no chess	Yes	1.607	1.335, 1.880
Islam et al., 2021	9–11	30	Yes	Treatment - control (pre, post) *R*	Yes	0.553	0.362, 0.745
Jerrim et al. 2018	9–10	30	Yes	Treatment - control (pre, post) *R*	Yes	0.010	– 0.150, 0.160
Joseph et al., 2016	11–12	40*	Unreported	Treatment - control (pre, post)	Yes	0.453	0.048, 0.858
Joseph, et al., 2017	4–15	40*	Unreported	Treatment - treatment (pre, post)	No	0.715	0.349, 1.081
Joseph et al., 2018	6–14	40*	Unreported	Treatment - treatment (pre, post)	No	0.444	0.120, 0.768
Kazemi et al., 2012	10–14	96	Unreported	Treatment - control	Yes	1.193	0.607, 1.779
Khosrorad et al., 2014	9–12	80	Unreported	Treatment - treatment (pre, post)	Yes	1.076	0.138, 2.013
Liptrap, 1998	8–10	Unknown	Yes	Chess - No Chess	No	0.680	0.420, 0.940

Table 3 (continued)

Study	Age range	Duration (hours)	Chess Instructor	Research design	Rand	Hedge's g	95% CI
Margulies, 1992	5–11	Unknown	Yes	Treatment - control (pre, post)	No	0.275	0.077, 0.474
Mel, 2021	18–19	Unknown	Researcher	Treatment - control (pre, post)	No	1.360	0.750, 1.980
Meloni & Fanari, 2019	8–10	30	Yes	Treatment - control	Yes	0.750	0.300, 1.190
Ramos et al., 2018	8–12	Unknown	Unreported	Chess - no chess	No	0.780	0.270, 1.290
Rifner, 1992	11–12	30	Researcher	Treatment - control (pre, post)	Yes	0.173	– 0.628, 0.975
Romano, 2011	8–9	30	Yes	Treatment - control (pre, post)	Yes	0.026	– 0.078, 0.130
Rosholm et al., 2017	6–8	27*	Teacher	Treatment - control (pre, post)	No	0.200	0.010, 0.390
Sala & Gobet, 2017	8–9	25	Yes	Treatment - control (pre, post) *P*	No	0.151	– 0.195, 0.497
Sala et al., 2015	8–10	18	Yes	Treatment - control (pre, post)	Yes	0.454	0.227, 0.681
Sala et al., 2016	8–9	15	Yes	Treatment - control (pre, post) *P*	Yes	0.333	– 0.304, 0.969
Scholz et al., 2008	9–11	24	Teacher	Treatment - control (pre, post)	Yes	0.122	– 0.396, 0.641
Sigirtmac, 2012	5–6	50	Unreported	Chess - no chess	No	1.060	0.540, 1.581
Sigirtmac, 2016	5–6	56	Teacher	Treatment - control	No	0.940	0.490, 1.390
Smith, 1998	16–17	105	Yes	Treatment - control (pre, post)	No	1.180	0.489, 1.870
Stegariu et al., 2019	7–10	Unknown	Unreported	Chess - no chess	No	1.375	0.697, 2.052
Thompson, 2003	11– 17	Unknown	Unreported	Chess - no chess	No	0.177	– 0.086, 0.439
Trinchero & Sala, 2016	7–12	19	Yes	Treatment - control (pre, post) *B*	Yes	0.344	0.168, 0.520
Velea & Cojocaru, 2019	6–11	20	Unreported	Treatment (pre, post)	No	0.600	0.114, 1.086
Voyer et al., 2018	8–9	60	Yes	Treatment - control (pre, post) *R*	No	0.339	0.009, 0.670
Yap, 2006	8–10	50	Yes	Treatment - control (pre, post)	No	0.273	0.025, 0.520
					Mean	0.572	0.127, 1.062

*Assuming one session of one hour per week. Chess - no chess = this design contrasts individuals regularly involved in chess playing with individuals without any chess experience. *B* = the research design includes teaching by chess instructor and teacher. *P* = the research design includes a placebo control group. *R* = the research design included repeated measures over at least two time points or evaluated long-lasting effects

The age range of the participants varied between 4 and 54 years old. Most studies were conducted with children or youngsters in either primary or secondary education, with two studies conducted with adults between 18 and 54 years old (Berkley, 2012) and between 18 and 19 years old (Mel, 2021). Concerning the duration of the intervention, there were 9 studies (20%) that unreported the duration of the chess intervention. In addition, the duration from 6 studies (13%) was estimated from the limited information provided. The mean duration of the chess interventions was 42 h (SD = 26). It was in addition reported that a chess instructor delivered the chess intervention in 21 studies (47%), a teacher in 5 studies (11%), the researcher in 6 studies (13%), and it was unreported in 13 studies (29%).

There were five main kinds of research designs. Thirty studies (67%) applied a treatment and control with a pre-test and a post-test design. Within this research design, there were only 3 studies (7%) with repeated measures over time or addressing long-lasting effects, two studies (4%) that used a placebo group, and a single study (2%) that contrasted teaching by a chess instructor with teaching by a professional teacher. Moreover, two studies (4%) used a treatment group with pre-test and post-test, 7 studies (16%) contrasted individuals regularly involved in chess playing (chess) with individuals unacquainted with chess (no chess), 3 studies (7%) contrasted two treatments with pre-test and post-test measures, and 3 studies (7%) compared a treatment with a control group in a single measure without post-test. Of the studies using some kind of control group (40), 17 studies (38%) reported a random assignment to groups, whereas 23 studies (51%) reported no random assignment.

The last two columns in Table 3 contain the Hedge's *g* effect size, which corrects for small samples, and its corresponding 95% confidence intervals. A confidence interval that excludes zero indicates a statistically significant effect size. These values correspond to the larger effects reported in the earlier meta-analysis about the impact of chess on cognitive and academic outcomes (Sala & Gobet, 2016), and also as obtained for each new study added to the current review. Hence, when considering the larger effects for each study, there were 23 effect sizes (51%) falling above 0.4, which was reported in the earlier meta-analyses as the median point of intervention effectiveness (Sala & Gobet, 2016).

NHST outcomes

Figure 3 shows the histograms for the statistically significant p values < 0.05 corresponding to 57 tests for cognitive ability (a), and 69 tests for academic performance (b), before removing the extreme z-values described earlier. These plots highlight a right-skewed pattern with most values falling within the lowest range (0 to 0.01), suggesting that there was evidential value in the data as long as selective reporting could be dismissed as the sole explanation of the findings (Simonsohn et al., 2014). This pattern was even more pronounced for the academic performance domain, with lower p values compared with the cognitive ability domain.

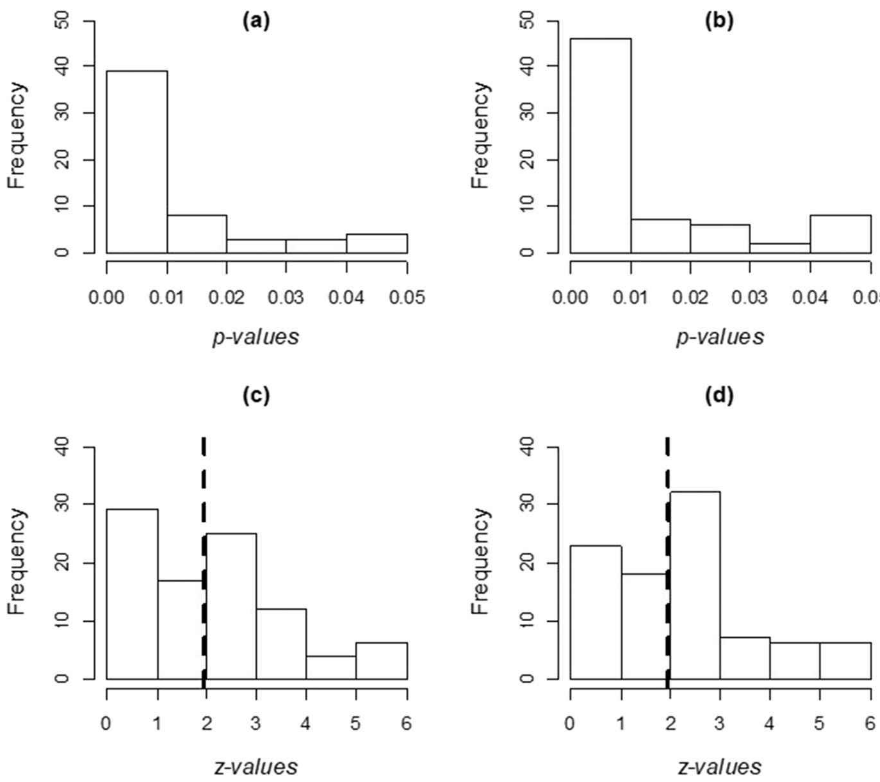

Fig. 3 The two top plots show the frequency distributions of the statistically significant findings ($p < 0.05$) in 57 tests for cognitive ability (**a**) and 69 tests for academic performance (**b**). The two bottom plots show the frequency distributions of the z-values in 93 tests for cognitive ability (**c**) and 92 tests in academic performance (**d**) after removing extreme values with $z > 6$. The discontinuous line indicates $z = 1.96$ corresponding to the $p = 0.05$, with non-significant values to the left of the line and significant values to the right of the line

In addition, Fig. 3 shows the histograms for the actual z-values < 6 used in the subsequent analyses, 93 tests for cognitive ability (c), and 92 tests for academic performance (d). In both of these plots, the values to the left of the $z = 1.96$ point ($p = 0.05$) indicated by the discontinuous line were statistically non-significant, whereas the values to the right of this point were statistically significant. The medians of both z-scores distributions were 2.03 for the cognitive ability domain (range from 0.01 to 5.79) and 2.05 for the academic performance domain (range from 0.02 to 5.93), suggesting stronger evidence against the null hypothesis ($1 - \alpha$) in the tests about academic performance than in the tests about cognitive ability.

Table 4 shows the ODR, EDR, FDR, file drawer, and ERR indices for the cognitive ability and academic performance domains estimated with two different methods, the expectancy-maximization (EM) method and the kernel-density

Table 4 Observed discovery rate (ODR), expected replicability rate (ERR), expected discovery rate (EDR), false discovery rate (FDR), and File Drawer. These estimates of 93 tests for cognitive ability and 92 tests for academic performance were obtained with expectancy-maximization (EM) and kernel-density (KD2) estimation methods, 500 bootstrap replications, and 95% conservative confidence intervals (CCI)

	Cognitive ability		Academic performance	
Indicator	Estimate	95% CCI	Estimate	95% CCI
	Expectancy-maximization (EM)			
ODR	0.540	[0.430, 0.640]	0.576	[0.470, 0.680]
EDR	0.531	[0.084, 0.692]	0.247	[0.050, 0.535]
FDR	0.046	[0.023, 0.575]	0.160	[0.046, 1.000]
FDR_Lower	0.574	---	1	---
File Drawer	.883	[0.446, 10.924]	3.043	[0.870, 19.000]
ERR	0.605	[0.392, 0.729]	0.422	[0.256, 0.632]
	Kernel-density (KD2)			
ODR	0.540	[0.430, 0.640]	0.576	[0.470, 0.680]
EDR	0.573	[0.081, 0.738]	0.321	[0.050, 0.628]
FDR	0.039	[0.019, 0.600]	0.111	[0.031, 1.000]
FDR_Lower	0.597	---	1	---
File Drawer	0.744	[0.354, 11.406]	2.115	[0.593, 19.000]
ERR	0.614	[0.427, 0.770]	0.486	[0.295, 0.663]

FDR_Lower = (1/ x – 1) * (0.05 * 0.95), where x is the lower bound of the 95% CCI for the EDR

(KD2) method. The ODRs obtained with both methods were the same in each respective domain. For cognitive ability, 50 out of the 93 tests were statistically significant (54%). For academic performance, 53 out of the 92 tests were statistically significant (58%). In both estimation methods (EM, KD2), the discrepancy between the ODR and EDR was lower for the studies about cognitive ability (EDR = 0.531 and 0.573) than for the studies about academic performance (EDR = 0.247 and 0.321). These discrepancies were not statistically significant because of the overlap in the respective ODR and EDR 95% CCI.

Because the EDR provides the unconditional mean power of the studies producing either non-significant or significant results, however, these estimates suggest that the studies targeting the effects of chess on academic subjects were more underpowered in contrast with the studies concerned with cognitive ability. The unconditional mean power of the studies targeting cognitive ability was around 53–57%, while being substantially lower (25–32%) for the studies targeting academic performance.

There was also a notable discrepancy in the maximum FDR across cognitive ability (FDR = 0.046 and 0.039) and academic performance (FDR = 0.160 and 0.111). Moreover, this discrepancy remained when using the lower bound of the EDR confidence interval to estimate the maximum FDR. These latter figures indicated a maximum of 57% and 60% of false positive outcomes for cognitive ability according with EMD and KD2 estimation methods, respectively. On the other hand, the findings for the academic performance domain suggested a maximum of 100% of false positives regardless of the estimation method. Regarding the file drawer estimates, the findings for cognitive ability indicated that there was at least one unreported non-significant outcome for every reported significant result. For academic performance, these figures were also higher than for the cognitive ability domain, indicating that, depending on the estimation method (EMD, KD2), there were 3 or 2 unreported non-significant results for every statistically significant result, respectively.

The expected replication rates were higher for the studies about cognitive ability (ERR = 0.605 and 0.614) than for the studies about academic performance (ERR = 0.422 and 0.486). These findings suggest that the amount of studies with statistically significant outcomes that would replicate in an independent study would be larger when targeting cognitive ability than academic performance. In any case, the estimated replication rate was substantially lower than suggested by just looking at the NHST main outcomes of the analyzed studies.

These findings highlight that chess instruction studies might be affected by a rather low statistical power, an excess of eventually false positive outcomes, publication biases derived from unreported non-significant results, and a low replicability of the reported outcomes. These outcomes, however, appeared to be even more remarkable for the academic performance domain than for the cognitive ability domain.

Discussion

Chess and chess teaching are argued to improve cognitive abilities and academic performance (Costello, 2013; Storey, 2000; Thomasina & Adams, 2012). The current study examined 208 outcomes obtained from null hypothesis significance testing (NHST) conducted in 45 studies from 19 countries. In the light of this review, such NHST outcomes entail notable drawbacks that undermine the alleged impact of chess on cognitive and academic development. An innovative application based on the *z*-curve approach was used to examine this assortment of heterogeneous findings (Bartoš & Schimmack, 2020a, b; Brunner & Schimmack, 2020). More specifically, the present review evaluated the statistical power, false positive rates, publication biases, and replication prospects of the available evidence about the benefits of chess instruction on cognitive abilities and academic performance.

Overview of Main Outcomes

The studies dealing with academic performance reported in the first place a higher proportion of statistically significant tests outcomes than the studies about cognitive ability (Fig. 3). Moreover, academic performance studies showed a larger discrepancy between observed and expected discovery rates, which suggests a larger publication bias compared with the studies targeting cognitive ability. These findings suggest in addition a low statistical power of the studies in either domain, underlining a low probability of detecting true effects in contrast with the minimum 80% acceptable statistical power rule of thumb (Bartoš & Schimmack, 2020b; Cohen, 1988). Furthermore, the tests from the academic performance domain yielded larger false discovery rates and file drawer estimates than the tests about the cognitive ability domain. Hence, these estimates suggest that there were large proportions of findings that could be false positives (i.e., rejecting the null hypotheses when being true), whereas studies producing non-significant results could have been generally unreported.

The expected replication rate (ERR) that estimates the proportion of studies supposed to replicate in independent experiments, was also lower for the academic performance tests than for the cognitive ability tests. That the estimated replication rate was rather low in both studied domains is somehow worrisome bearing in mind the importance of replication for educational psychology (Plucker & Makel, 2021). Such a low estimated replication ability of the available evidence undermines the overall quality of chess instruction research. On the other hand, this evidence might be generally uninformative and misguiding for practitioners, policy makers, and the public, regarding the magnitude and generalization of the claimed effects of chess instruction on cognitive and academic development (see Table 1).

According with the outcomes from the current review, the NHST findings about the impact of chess on the improvement of cognitive ability and academic performance undergo substantial drawbacks associated with statistical power, excess of false positive outcomes, publication biases, and low replication rates. Compared with cognitive ability, however, these drawbacks are even larger for the studies about academic performance.

Drawbacks of Current Studies

Most of the reports describing chess education interventions highlight the positive effects of practising chess to improve cognitive abilities and academic subjects such as mathematics, science, or language. It is remarkable, however, that the few studies (16%) with more robust research designs, larger sample sizes, or more sophisticated data analysis approaches, tend to report lower or null effects (Islam et al., 2021; Jerrim et al., 2018; Sala & Gobet, 2017; Thompson, 2003).

Theoretical and methodological drawbacks limit the arguments about the benefits of chess instruction. In the first place, chess instruction studies tend to ignore the theoretical basis underlying far transfer effects (Barnett & Ceci, 2002; Campitelli & Gobet, 2008; Gobet & Campitelli, 2006; Sala et al., 2017). Similarly, most of the chess instruction interventions conducted so far lack consistent hypotheses, neglecting for instance individual differences in cognitive ability as the nuclear factor involved in far transfer processes (Ferguson, 1954, 1956; Snow & Lohman, 1984; Snow & Swanson, 1992). On the methodological side, several drawbacks identified in the past are also corroborated in this review. These include failing to adopt basic ideal experiment guidelines and a preponderance of cross-sectional studies that prevent the evaluation of far transfer over time. Moreover, there are inconsistencies in the reporting of key elements such as the duration of the interventions, or the individuals involved at different intervention stages, i.e., teaching and data collection (Gobet & Campitelli, 2006; Jerrim et al., 2018; Sala et al., 2017; Sala & Gobet, 2016).

Nonetheless, the specific major drawback highlighted in the current review consists in the unduly reliance on NHST findings, while making strong claims about the benefits of chess. Why is this so? Three interrelated factors could account for the excess of confidence in such methods. In the first place, several of these studies are not published in peer-reviewed journals, a point that was also raised in the earlier critical review about the benefits of chess instruction (Gobet & Campitelli, 2006). Some of the evaluated studies are unpublished reports or doctoral dissertations that might have not been sufficiently shaped and polished to meet the minimum scientific requirements posed by peer-reviewed journals. Peer-reviewed journals may bear more stringent criteria regarding theoretical motivation and hypotheses building, research designs meeting ideal experiment guidelines, and requiring a more nuanced reporting of outcomes by including information about effect sizes, confidence intervals, and statistical power (Sun et al., 2010). Secondly, the authors of these studies may be engaged as well in the chess teaching process. Henceforth, they may feel quite devoted to teaching chess with a genuine and honest intention to contribute to child development and adjustment. This, however, entails the peril of becoming emotionally involved in both, the teaching and the subsequent evaluation of the intervention, which might contribute to introducing subtle experimenter bias. Indeed, one of the guidelines of the ideal experiment comprises different persons conducting both activities teaching and evaluation. On the other hand, whether teaching quality relates with a larger achievement remains unclear (Jerrim et al., 2018). A third related issue with the lack of publication in peer-reviewed journals and experimenter bias, is the emergence of conflicting interests because teachers could be largely interest in obtaining positive outcomes derived from the chess interventions. Potential conflicts of interest are for instance asked to be disclosed in peer-reviewed journals.

These drawbacks are somehow similar when compared with studies evaluating the quality of other theories usually applied in education, such as multiple intelligences, the Mozart effect, or emotional intelligence (Ferrero et al., 2021; Pietschnig et al., 2010; Waterhouse, 2006). For example, the extensive review about multiple intelligences identified comparable pitfalls regarding experimental research designs, the need of larger samples to increase statistical power, the inclusion of active or placebo groups, or improvements in the reporting of research outcomes. Moreover, a substantial point also highlighted in the review about multiple intelligences, is that teachers and policy-makers may strive for alternative teaching strategies in an attempt to remedy academic failure of individuals involved in more traditional teaching (Ferrero et al., 2021). Seeking for alternative paths to increase academic achievement, might be congruent in turn with a test-based accountability approach, which implies large scale testing to account for school achievement, eventually implying either school, teacher, and student rewards or sanctions (Hamilton et al., 2002). From this viewpoint, however, it has been called into question whether a single measure is useful to adequately represent individual performance.

Chess, Cognitive Ability, and Academic Achievement

The current review identified one hundred tests addressing the impact of chess on cognitive abilities. The higher scores in cognitive abilities and academic performance observed for those individuals involved in chess may bear alternative more plausible explanations to account for the observed findings, when compared with a purported and intriguing far transfer impact of chess instruction (Barnett & Ceci, 2002).

A more straightforward account could be perhaps that children who are more cognitively able might be more attracted and motivated to an intellectually demanding activity such as chess. Individuals tend to seek environments that are more suitable to their abilities and predispositions, i.e., active gene-environment correlation (Petrill & Wilkerson, 2000; Scarr & McCartney, 1983). For example in professional chess, changes in performance over time may depend on factors related with individual abilities and predispositions and factors related with intensive practice and learning (Blanch, 2018). Other evidence indicates that apart from practice, spatial and logical abilities (Horgan & Morgan, 1990) and numerical abilities (Ferreira & Palhares, 2008) predispose to chess performance and chess involvement at early ages. Moreover, a meta-analysis about the association of chess performance with cognitive ability suggests that numerical or spatial abilities might be more important for chess performance at early ages (Burgoyne et al., 2016). Another study contrasting chess playing with cognitive ability highlights the latter as the stronger covariate of academic achievement (Thompson, 2003). From this viewpoint, children with more outstanding cognitive abilities would be also more predisposed and attracted to the chess environment. Hence, children with a greater involvement in chess playing might be more likely to score higher in specific cognitive abilities, and consequently, obtain higher scores in academic subjects.

Since cognitive abilities relate robustly with academic performance (Beier et al., 2010; Blanch & Aluja, 2013; Deary et al., 2007; Laidra et al., 2007), the link between chess instruction with the reported improvements in academic achievement could be largely confounded because of being moderated or mediated by cognitive abilities. Cognitive ability could then be alternatively specified as a moderator or a mediator variable, with chess instruction or chess proficiency as the predictor variable and academic performance as the outcome variable. With cognitive ability as a moderator variable, the impact of chess on academic performance should vary at different levels of cognitive ability. With cognitive ability as a mediator variable, the impact of chess on academic performance would depend on the indirect effect through cognitive ability rather than on the direct effect from chess to academic performance (Baron & Kenny, 1986). From this point of view, it would be informative to examine chess instruction effects through the lens of more complex mediator or moderator models.

As far as it is known, only a single study has examined a model of this kind. With a sizeable sample of primary schoolchildren (n = 255), the impact of chess skill on language and mathematics achievement appears to be actually mediated by fluid intelligence (Gao et al., 2021). Conventional tests of intelligence, however, might yield biased information to characterizing actual cognitive abilities in real-life settings such as chess. Analytical, creative, and practical abilities measures have been suggested as better choices for teaching, instruction, and assessment purposes (Sternberg, 2002; Sternberg & Grigorenko, 2000).

Suggestions and Recommendations

As described earlier, past reviews recommend that publishing chess studies in peer-reviewed journals and improving the quality of experimental research designs could increase considerably the quality of the evidence about the benefits of chess (Bart, 2014; Gobet & Campitelli, 2006; Sala et al., 2017; Sala & Gobet, 2016). The outcomes from the current review are useful to delineate a few additional recommendations regarding the implementation and reporting of chess instruction interventions.

A key recommendation to improve statistical power consists in increasing the sample size (Schimmack, 2012). The sample sizes of the analyzed studies were rather low, which was a serious constraint of statistical power. One possibility to achieve larger sample sizes could be multi-site collaborative studies with the same experimental research design. Furthermore, more complex randomized experimental research designs could also be used in order to conduct more accurate analyses of individual differences in the response to chess interventions (Tucker-Drob, 2011). There are alternative useful research designs that could be applied to reduce selection bias in quasi-experiments addressing the impact of chess interventions, such as parallelization, matching, or a waiting control group (Cook & Steiner, 2010). Moreover, there are several evaluation methods that could be used in chess interventions to provide causal based evidence about the effectiveness of chess interventions on cognitive and academic areas (Schlotter et al., 2011). Beyond randomized chess interventions, it is suggested in addition to increase the amount of correlational studies with both cross-sectional and longitudinal designs, and to conduct data analyses based on regression and structural equation modelling methods (Bollen, 1989). These methods could be used to specify more complex causal models by including either moderator or mediator variables such as cognitive ability as discussed earlier.

More efforts should be also devoted to the longitudinal assessment of chess interventions, as most studies to date focus on the immediate assessment shortly after the intervention. Similarly to educational interventions in other areas, long-term effects of chess instruction are rarely evaluated, with only three studies (7%) in the present review. Comprehensive evaluations of chess interventions over larger time span appears to yield null effects (Jerrim et al., 2018). The generalized lack of long-term follow-up studies about educational interventions has been recently highlighted regarding several implications for funding bodies and researchers (Watts et al., 2019). In this view, long-term research designs require perhaps more rigorous planning, execution, and reporting of outcomes, which would be beneficial for advancing the field. Furthermore, strong claims such as those usually vested to the chess interventions examined here should probably require a more comprehensive and detailed framework to account for the chess impact on relevant outcomes as proposed by theories of action. Some key features to be included in such assessment systems should include the intended effects (short and long-term), the components of the assessment system with their corresponding rationales, the expected interpretive claims, the mechanisms designed to cause intended effects, and potential unexpected negative effects (Bennett, 2010).

Mastering chess requires an intensive dedication and deep study that may preclude the learning of other curricular subjects during the schooling period (Campitelli & Gobet, 2008; Tricot & Sweller, 2014). On the other hand and as suggested elsewhere, the benefits of chess could range from low-level skills such as concentration or assuming playing mistakes and defeats, to higher-level skills such as intelligence or school performance, which fit well with the concepts of near and far transfer, respectively. Most studies to date, however, have placed the focus on high-level far transfer outcomes, which may take longer to be transferred to other domains, involve more complex processes, and bear a considerable larger individual variability (de Groot, 1965; Gobet & Campitelli, 2006).

Another unexplored area of research that should be addressed within this field relates with sex differences. On the one hand, meaningful sex differences are usually reported in cognitive abilities (Haier et al., 2005; Halpern & LaMay, 2000) and academic achievement (Blanch & Aluja, 2013; Reilly et al., 2015; Voyer & Voyer, 2014). On the other hand, chess appears as more appealing for males than for females, whereas males tend to outperform females in competitive chess (Bilalic et al., 2009; Blanch, 2016, 2021; Blanch et al., 2015; Howard, 2014). Because girls tend to perform better than boys in academic subjects, and boys tend to be more prone to become involved in chess, perform better than girls, and persist longer in this domain, examining sex differences in the purported benefits of chess instruction on cognitive abilities and academic achievement emerges as a topic of considerable theoretical interest.

Conclusion

The popularity of chess instruction has somewhat risen considerably in the past few years with many reports highlighting several individual benefits of chess. This evidence, however, appears somehow flawed in the light of the current review. Nonetheless, chess seems as an interesting activity to teach in school contexts. Chess is an ancient universal game, an intellectually related activity that is part of human cultural heritage, and that comprises a considerable body of specialized knowledge. It is easy to implement, it is relatively cheap, and it is appealing to both children and adults because of its playing and competitive nature. That the focus of chess interventions has been too much on far transfer to cognitive and academic outcomes, with little or no attention to near transfer skills such as concentration, motivation, or accepting responsibilities is not new, albeit evident from the outcomes in this review (Gobet & Campitelli, 2006). Similarly, the largest and most comprehensive study to date suggests to focus on children's well-being and non-cognitive related skills (Jerrim et al., 2018). Perhaps it is not necessary to justify the need of teaching chess in the classrooms because of being beneficial for intelligence and academic performance. What is needed ultimately is the undertaking of higher quality studies about the benefits of chess for other kinds of simpler outcomes that may be relevant for education.

Appendix A

Tables 5, 6

Table 5 Input data corresponding to the p values of 100 tests of cognitive ability, and 108 tests of academic performance.

#Cognitive ability					
0.010900742	0.000719177	0.001297801	0.000000200	0.001201156	0.090970561
0.000000000	0.000000000	0.000000000	0.000000017	0.000000001	0.000000000
0.091457814	0.005528507	0.000187122	0.719657560	0.539860144	0.120391707
0.820709104	0.066473081	0.487181747	0.446505846	0.050923527	0.446505846
0.012071947	0.511822237	0.000418364	0.001594659	0.012646398	0.394419544
0.008157904	0.000000007	0.016620365	0.797950837	0.948970227	0.029705135
0.012245139	0.023129370	0.088115407	0.504745012	0.589129850	0.097033619
0.630624452	0.003852418	0.003852418	0.006933948	0.035291308	0.382118254
0.042356539	0.357572759	0.171625306	0.130082634	0.683439073	0.000014342
0.000000000	0.000000666	0.000000078	0.000000047	0.000000014	0.365395781
0.000000000	0.000682575	0.094848136	0.000001621	0.001610045	0.000000947
0.005072763	0.006045584	0.032953899	0.541911717	0.130227249	0.795044429
0.414584776	0.276857426	0.010280088	0.008849073	0.007551199	0.002276986
0.024089327	0.538848877	0.692687101	0.694736765	0.402609608	0.491322699
0.221792516	0.991043244	0.000134679	0.047602498	0.018842627	0.000573217
0.036118528	0.046875315	0.000232907	0.008011058	0.811140548	0.536194933
0.047127187	0.073139891	0.014835781	0.388030547		
#Academic performance					
0.000000160	0.000000000	0.000000145	0.011660760	0.000000000	0.007260339
0.007960659	0.000000000	0.000000004	0.000000000	0.000000000	0.000000000
0.000000000	0.000000000	0.000000000	0.022836558	0.000000003	0.000000000
0.000003821	0.000013049	0.000000001	0.000006669	0.000959981	0.905275356
0.002191301	0.000821787	0.111535712	0.433674019	0.002956335	0.251762253
0.002667802	0.029461226	0.651168879	0.017671906	0.000000000	0.923239164
0.319060196	0.133566524	0.025377604	0.000844933	0.004189616	0.091217419
0.018020104	0.001495225	0.012670468	0.059704214	0.049504607	0.040030364
0.050242189	0.088048467	0.077515772	0.665776736	0.156804168	0.140795206
0.565092142	0.427790553	0.000000000	0.180411323	0.030237646	0.426072711
0.338238333	0.434210141	0.084383727	0.086066952	0.909556943	0.960789837
0.940824146	0.436127397	0.820378178	0.008717625	0.300545651	0.041004140
0.006132099	0.000013971	0.004331681	0.006315937	0.692548155	0.041710921
0.005140023	0.000000236	0.023112897	0.000000000	0.118648136	0.963489902
0.000008929	0.006912951	0.932448805	0.023834645	0.000028357	0.981183628
0.981183628	0.020976266	0.000000031	0.040022045	0.677507665	0.042704600
0.010310578	0.000000000	0.042667171	0.001349002	0.048902068	0.000000000
0.007990795	0.039869706	0.014957593	0.010445017	0.155231868	0.539835624

Table 6 R code used to estimate the *z*-curves in cognitive ability and academic performance tests.

```
library(zcurve)

#COGNITIVE ABILITY (100 TESTS)
cog <- c(0.010900742, 0.000719177, … , 0.014835781, 0.388030547)

#ACADEMIC PERFORMANCE (108 TESTS)
aca <- c(0.000000160, 0.000000000, … , 0.155231868, 0.539835624)

#COMPUTATION OF z-VALUES AND SELECTION OF z < 6
#z values for cognitive ability
z<--qnorm(log(cog/2), log.p=TRUE)
zcog<-subset(z, z !=Inf & z < 6)

#z values for academic performance
z<--qnorm(log(aca/2), log.p=TRUE)
zaca<-subset(z, z !=Inf & z < 6)


#EM ESTIMATION
set.seed(100)
co <- zcurve(zcog, method = "EM",
           control = list(sig_level = 0.051),
           bootstrap = 500)

ac <- zcurve(zaca, method = "EM",
           control = list(sig_level = 0.051),
           bootstrap = 500)


#KD2 ESTIMATION
set.seed(100)
co <- zcurve(zcog,   method = "density",
           control = list(model="KD2", sig_level = 0.051),
           bootstrap = 500)

ac <- zcurve(zaca,   method = "density",
           control = list(model="KD2", sig_level = 0.051),
           bootstrap = 500)


#RESULTS
summary(co, all=T, ERR.adj=0.03, EDR.adj=0.05)
summary(ac, all=T, ERR.adj=0.03, EDR.adj=0.05)
```

Acknowledgements This research was done with the support of the County Council of Lleida (Diputació de Lleida), Grant ID: 202100063). This research was performed within the Catalonian Consolidated Research Group SGR 00008. 2017–2019 period. The comments and suggestions from four anonymous reviewers were very much appreciated.

Author Contribution Not applicable

Funding Diputació de Lleida (Grant ID: 202100063).

Data Availability Data is available in the supplemental material.

Code Availability Data and software code are available in Appendix 1 and 2.

Declarations

Ethics Approval The study did not require approval from ethics comittee.
Consent to Participate Not applicable.
Consent for Publication Not applicable.

Conflicts of Interest There are no conflicting/competing interests.

References

Aciego, R., García, L., & Betancort, M. (2012). The benefits of chess for the intellectual and social-emotional enrichment in schoolchildren. *The Spanish Journal of Psychology, 15*(2), 551–559.

Atashafrouz, A. (2019). The effectiveness of chess on problem-solving, working memory, and concentration of male high school students. *Iranian Evolutionary and Educational Psychology Journal*, 249-258. https://doi.org/10.29252/ieepj.1.4.2

Aydin, M. (2015). Examining the impact of chess instruction for the visual impairment on mathematics. *Educational Research and Reviews, 10*(7), 907–911.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Barrett, D. C., & Fish, W. W. (2011). Our move: Using chess to improve math achievement for students who receive special education services. *International Journal of Special Education, 26*(3), 181–193.

Bart, W. M. (2014). On the effect of chess training on scholastic achievement. *Frontiers in Psychology, 5*, 762.

Bartoš, F., & Schimmack, U. (2020a). Z-curve.2.0: Estimating Replication rates and discovery rates. https://doi.org/10.31234/osf.io/urgtn.

Bartoš, F., & Schimmack, U. (2020b). zcurve: An R package for fitting z-curves. R package version 1.0.0, URL: https://github.com/FBartos/zcurve. Accessed 22 Feb 2020.

Basson, M. R. (2015). *The effects of chess instruction on the intellectual development of grade R leaners*. University of South Africa.

Beier, M. E., Campbell, M., & Crook, A. E. (2010). Developing and demonstrating knowledge: Ability and non-ability determinants of learning and performance. *Intelligence, 38*(1), 179–186.

Bennett, R. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*(2-3), 70–91.

Berkley, D. K. (2012). *The impact of chess instruction on the critical thinking ability and mathematical achievement of developmental mathematics students*. ProQuest LLC.

Bilalic, M., Smallbone, K., McLeod, P., & Gobet, F. (2009). Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains. *Proceedings of the Royal Society B, 276*, 1161–1165.

Blanch, A. (2015). Evaluating fluid and crystallized abilities in the performance of an educational process. *Instructional Science, 43*, 427–442.

Blanch, A. (2016). Expert performance of men and women: A cross-cultural study in the chess domain. *Personality and Individual Differences, 101*, 90–97.

Blanch, A. (2018). Top hundred chess experts: A cross-domain analysis of change over time. *Intelligence, 71*, 76–84.

Blanch, A. (2021). *Chess and individual differences*. Cambridge University Press.

Blanch, A., & Aluja, A. (2013). A regression tree of the aptitudes, personality, and academic performance relationship. *Personality and Individual Differences, 54*(6), 703–708.

Blanch, A., Aluja, A., & Cornadó, M. P. (2015). Sex differences in chess performance: Analyzing participation rates, age, and practice in chess tournaments. *Personality and Individual Differences, 86*, 117–121.

Blanch, A., & Llaveria, A. (2021). Ability and non-ability traits in chess skill. *Personality and Individual Differences, 179*, 110909.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Brunner, J., & Schimmack, U. (2020). Estimating population mean power under conditions of heterogeneity and selection for significance. *Meta-Psychology, 4*, https://doi.org/10.15626/MP.1012018.1015874.

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.

Buyk, J., & Ho, F. (2008). *The effect of math and chess integrated instruction on math scores*. The Chess Academy.
Campitelli, G., & Gobet, F. (2008). The role of practice in chess: A longitudinal study. *Learning and Individual Differences, 18*(4), 446–458.
Christiaen, J., & Verhofstadt-Denève, L. (1981). Schaken en cognitieve ontwikkeling. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 36*, 561–582.
Cleveland, A. (1907). The psychology of chess and of learning to play it. *American Journal of Psychology, 18*(3), 269–308.
Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Lawerence Erlbaum Associates Inc.
Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 15*(1), 56–68.
Costello, P. J. M. (2013). The gymnasium of the mind: Teaching chess in early childhood. *Early Child Development and Care, 183*(8), 1133–1146.
de Groot, A. D. (1965). *Thought and choice in chess* (2nd ed.). Mouton.
Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21.
DuCette, J. (2009). *An evaluation of the Chess Challenge Program of ASAP/after school activities partnerships*. Philadelphia, PA. https://www.saintlouischessclub.org/sites/default/files/DUCETTE_2009_EVALUATION.pdf. Accessed 28 Nov 2018.
Eberhard, J. W. (2003). *The relationship between classroom chess instruction and verbal, quantitative, and nonverbal reasoning abilities of economically disadvantaged students.* Unpublished Doctoral Dissertation, Texas A&M University, College Station.
Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8*, 95–112.
Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology, 10*, 121–131.
Fernández-Amigo, J. (2008). *Utilización de material didáctico con recursos de ajedrez para la enseñanza de las matemáticas. Estudio de sus efectos sobre una muestra de alumnos de segundo de primaria.* Unpublished Doctoral Dissertation, Universitat Autònoma de Barcelona, Barcelona.
Ferreira, D., & Palhares, P. (2008). Chess and problem solving involving patterns. *The Mathematics Enthusiast, 5*(2), 249–256.
Ferrero, M., Vadillo, M. A., & León, S. P. (2021). A valid evaluation of the theory of multiple intelligences is not yet possible: Problems of methodological quality for intervention studies. *Intelligence, 88*, 101566.
Forrest, D., Davidson, D., Shucksmith, J., & Glendinning, T. (2005). *Chess development in Aberdeen's primary schools: A study of literacy and social capital.* Aberdeen City Council. University of Aberdeen. Scottish Executive Education Department.
Gao, Q., Feng, Y., Chen, W., & Ping, X. (2021). Does perceived chess skills mediate the relationship between fluid intelligence and academic performance? *Journal of Psychology in Africa, 31*(1), 56–60.
Garcia, N. V. (2008). *Scholastic chess club participation and the academic achievement of hispanic fifth grade students in south Texas.* Unpublished Doctoral Dissertation, University of Houston, Houston.
Gliga, F., & Flesner, P. I. (2014). Cognitive benefits of chess training in novice children. *Procedia - Social and Behavioral Sciences, 116*, 962–967.
Gobet, F., & Campitelli, G. (2006). Educational benefits of chess instruction: A critical review. In T. Redman (Ed.), *Chess and education: Selected essays from the Koltanowski conference*. Chess Program at the University of Texas at Dallas.
Grau-Pérez, G., & Moreira, K. (2017). A study of the influence of chess on the executive functions in school-aged children. *Estudios de Psicología, 38*(1), 1–22.
Häggström, O. (2017). The need for nuance in the null hypothesis significance testing debate. *Educational and Psychological Measurement, 77*(4), 616–630.
Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *Neuroimage, 25*, 320–327.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Prentice Hall.
Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*(2), 229–246.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education*. RAND.
Hong, S., & Bart, W. M. (2007). Cognitive effects of chess instruction on students at risk for academic failure. *International Journal of Special Education, 22*(3), 89–96.
Horgan, D. D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology, 4*(1), 109–128.
Howard, R. W. (2014). Explaining male predominance at the apex of intellectual achievement. *Personality and Individual Differences, 68*, 217–220.
Isikgöz, E. (2016). Analysis of math success of secondary school students playing and not playing chess (Sakarya province sample). *International Journal of Human Sciences, 13*(1), 1689–1699.
Islam, A., Lee, W. S., & Nicholas, A. (2021). The effects of chess instruction on academic and non-cognitive outcomes: Field experimental evidence from a developing country. *Journal of Development Economics, 150*, 102615.
Jerrim, J., Macmillan, L., Micklewright, J., Sawtell, M., & Wiggins, M. (2018). Does teaching children how to play cognitively demanding games improve their educational attainment? Evidence from a Randomised Controlled Trial of chess instruction in England. *Journal of Human Resources, 53*(4), 993–1021.
Joseph, E., Easvaradoss, V., Abraham, S., & Chan, M. A. (2017). Mentoring children through chess training enhances cognitive functions. *International Journal of Information and Education Technology, 7*(9), 669–672.
Joseph, E., Easvaradoss, V., Abraham, S., & Jain, S. (2018). Enhancing verbal reasoning of school children through chess learning. *International Journal of Scientific Engineering and Science, 2*, 7–13.
Joseph, E., Easvaradoss, V., & Solomon, N. J. (2016). Impact of chess training on academic performance of rural Indian school children. *Open Journal of Social Sciences, 4*, 20–24.
Kazemi, F., Yektayar, M., & Abad, A. M. B. (2012). Investigation the impact of chess play on developing meta-cognitive ability and math problem-solving power of students at different levels of education. *Procedia - Social and Behavioral Sciences, 32*, 372–379.
Khosrorad, R., Kouhbanani, S. S., & Sani, A. R. (2014). Effectiveness of chess training for improving executive functions and mathematics performance of students with mathematics disorders. *International Journal of Educational Investigations, 1*(1), 283–295.
Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*(3), 441–451.
Liptrap, J. M. (1998). Chess and standard test scores. *Chess Life, 3*, 41–43.
Margulies, S. (1992). *The effect of chess on reading scores: District nine chess program second year report*. The American Chess Foundation.
Mel, B. N. (2021). Does playing chess improve mathematics scores? An experimental study among co-curricular chess students in Politeknic Kuching Sarawak. *International Journal of Advanced Research in Education and Society, 13*(2), 9–17.
Meloni, C., & Fanari, R. (2019). *Chess training effect on meta-cognitive processes and academic performance.* Paper presented at the 16th international Conference on Cognition and Exploratory Learning in Digital Age. https://doi.org/10.33965/celda2019_201911L048.
Miller, J. (2017). Hypothesis testing in the real world. *Educational and Psychological Measurement, 77*(4), 663–672.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.
Petrill, S. A., & Wilkerson, B. (2000). Intelligence and achievement: A behavioral genetic perspective. *Educational Psychology Review, 12*(2), 185–199.
Pietschnig, J., Voracek, M., & Forman, A. K. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence, 38*(3), 314–323.
Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist, 56*(2), 90–100.
R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
Ramos, L., Arán, V., & Krumm, G. (2018). Funciones ejecutivas y práctica de ajedrez: Un estudio en niños escolarizados. *Psicogente, 21*(39), 25–34.

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107*(3), 645–662.

Rifner, P. J. (1992). *Playing chess: A study of the transfer of problem-solving skills in students with average and above average intelligence.* Unpublished Doctoral Dissertation, Purdue University, West Lafayette.

Romano, B. (2011). *Does playing chess improve math learning? Promising (and inexpensive) results from Italy*. University of Pennsylvania.

Rosholm, M., Mikkelsen, M. B., & Gumede, K. (2017). Your move: The effect of chess on mathematic test scores. *PLoS ONE, 12*(5), e0177257.

Sala, G., Foley, J. P., & Gobet, F. (2017). The effects of chess instruction on pupils' cognitive and academic skills: State of the art and theoretical challenges. *Frontiers in Psychology, 8*, 238.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review, 18*, 16–57.

Sala, G., & Gobet, F. (2017). Does chess instruction improve mathematical problem-solving ability? Two experimental studies with an active control group. *Learning & Behavior, 45*, 414–421.

Sala, G., Gobet, F., Trinchero, R., & Ventura, S. (2016). *Does chess instruction enhance mathematical ability in children? A three-group design to control for placebo effects.* Paper presented at the 38th Annual Meeting of the Cognitive Science Society, Philadelphia, PA.

Sala, G., Gorini, A., & Pravettoni, G. (2015). Mathematical problem-solving abilities and chess: An experimental study on young pupils. *Sage Open, 5*, 1–9.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of gentoype®environment effects. *Child Development, 54*(2), 424–435.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551–566.

Schlotter, M., Schwerdt, G., & Woessmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: A non-technical guide. *Education Economics, 19*(2), 109–137.

Scholz, M., Niesch, H., Steffen, O., Ernst, B., Loeffler, M., Witruk, E., et al. (2008). Impact of chess training on mathematics performance and concentration ability of children with learning disabilities. *International Journal of Special Education, 23*(3), 138–148.

Sigirtmac, A. D. (2012). Does chess training affect conceptual development of six-year-old children in Turkey? *Early Child Development and Care, 182*, 797–806.

Sigirtmac, A. D. (2016). An investigation on the effectiveness of chess training on creativity and theory of mind development at early childhood. *Educational Research and Reviews, 11*(11), 1056–1063.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547.

Smith, J. P. (1998). *A quantitative analysis of the effects of chess instruction on the mathematics achievement of southern, rural, black secondary students.* Unpublished Doctoral Dissertation, Louisiana Tech University.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*(3), 347–376.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: aptitude, adaptation, and assessment. *Annual Review of Psychology, 43*, 583–626.

Stegariu, V., Popusoi, S. A., Abalasei, B., Voinea, N. L., Stelescu, I., Trofin, F. P., et al. (2019). The role of chess in the intellectual development of childrens from primary school. *Sport and Society, 19*(2), 58–62.

Sternberg, R. J. (2002). Raising the achievement of all students: Teaching for successful intelligence. *Educational Psychology Review, 14*(4), 383–393.

Sternberg, R. J., & Grigorenko, E. L. (2000). Theme-park psychology: A case study regarding human intelligence and its implications for education. *Educational Psychology Review, 12*(2), 247–268.

Storey, K. (2000). Teaching beginning chess skills to students with disabilities. *Preventing School Failure, 44*(2), 45–49.

Sun, S., Pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology, 102*(4), 989–1004.

Thomasina, C. P., & Adams, M. S. (2012). Chess from square a1: Incorporating chess into the gifted class. *Gifted Child Today, 35*(4), 243–251.

Thompson, M. (2003). Does the playing of chess lead to improved scholastic achievement. *Issues in Educational Research, 13*(2), 13–26.
Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions (I.). *Psychological Review, 9*, 247–261.
Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*, 265–283.
Trinchero, R., & Sala, G. (2016). Chess training and mathematical problem solving: The role of teaching heuristics in transfer of learning. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(3), 655–668.
Tucker-Drob, E. M. (2011). Individual differences methods for randomized experiments. *Psychological Methods, 16*(3), 298–318.
Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. *Educational and Psychological Measurement, 61*(2), 219–224.
Velea, T., & Cojocaru, V. (2019). The effect of playing chess on focused attention. *European Proceedings of Social and Behavioural Sciences*. https://doi.org/10.15405/epsbs.12019.15402.15484
Voyer, D., Rousseau, M., Rajotte, T., Freiman, V., & Cabot-Thibault, J. (2018). Le jeu d'échecs dans les classes du primaire :un moyen ciblé pour développer les habiletés des garçons en résolution de problèmes. *Formation et Profession, 26*(3), 94–108.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204.
Waterhouse, L. (2006). Inadequate evidence for multiple intelligences, mozart effect, and emotional intelligence theories. *Educational Psychologist, 41*(4), 247–255.
Watts, T., Bailey, D. H., & Li, C. (2019). Aiming further: Addressing the need for high-quality longitudinal research in education. *Journal of Research on Educational Effectiveness, 12*(4), 648–658.
Yap, K. O. (2006). *Chess for success evaluation*. Northwest Regional Educational Laboratory.