

The impact of the SMART program on cognitive and academic skills: A systematic review and meta-analysis

Richard J. May¹ | Ian Tyndall² | Aoife McTiernan³ |
Gareth Roderique-Davies¹ | Shane McLoughlin⁴

¹School of Psychology and Therapeutic Studies, University of South Wales, Pontypridd, UK

²Department of Psychology and Counselling, University of Chichester, Chichester, UK

³School of Psychology, National University of Ireland, Galway, Ireland

⁴Jubilee Centre for Character and Virtues, University of Birmingham, Birmingham, UK

Correspondence

Richard J. May, School of Psychology and Therapeutic Studies, University of South Wales, Pontypridd CF37 1DL, UK.
Email: Richard.may@southwales.ac.uk

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Abstract

Online interventions promoted to enhance cognitive ability hold great appeal for their potential positive impact in social, employment, and educational domains. Cognitive training programs have, thus far, not been shown to influence performance on tests of general cognitive aptitude. Strengthening Mental Abilities with Relational Training (SMART) is an online program that claims to raise intelligence quotient (IQ). This systematic review and meta-analysis evaluates the effect of SMART on indices of cognitive aptitude and academic performance. The review protocol was registered at PROSPERO (CRD42019132404). A systematic literature search of bibliographic databases (ERIC, PsycINFO, PubMed, Applied Social Sciences Index and Abstracts, Scopus, Proquest Psychology) identified five studies ($N = 195$) that met the criterion for inclusion. The risk of bias was assessed using the Cochrane Collaboration Risk of Bias 'RoB 2' tool. Overall, there was a moderate impact of SMART on measures of nonverbal IQ ($g = 0.57$, 95% CI [0.24, 0.89]). There was insufficient evidence to determine the impact of SMART on any other domain. All

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *British Journal of Educational Technology* published by John Wiley & Sons Ltd on behalf of British Educational Research Association

studies included in the review were judged to be at a high risk of bias for their primary outcome. Despite the methodological limitations of published studies to date, these initial findings suggest that a large-scale study of SMART is warranted.

KEYWORDS

cognitive training, intelligence, meta-analysis, relational frame theory, SMART

INTRODUCTION

Attempts by researchers to raise cognitive performance on standardised tests have proved extremely challenging (eg, Sala & Gobet, 2019). The effects of interventions on cognitive performance are often conceptualised in the literature in terms of two categories: near transfer and far transfer (cf. Sala et al., 2019). Near transfer refers to improvements in a skillset from within the same domain as that which is targeted with the intervention (eg, training

Practitioner notes

What is already known about this topic

- SMART is a popular, commercially available online program that claims to improve cognitive skills in children.
- A number of controlled trials have investigated the efficacy of SMART and reported positive findings.
- There are no existing systematic reviews or meta-analyses of the literature for this intervention.

What this paper adds

- The present study represents the first systematic review and meta-analysis of the effect of SMART on cognitive and educational outcomes.
- We identified five trials that met the criteria for inclusion in the review. All five studies were rated as having a high risk of bias using the Cochrane Collaboration Risk of Bias tool.
- We calculated a moderate overall impact of SMART on measures of nonverbal IQ. There was insufficient evidence to determine the impact of SMART in any other cognitive or educational domain.

Implications for practice and/or policy

- Practitioners and/or teachers can use the review to inform their decisions about adopting SMART as an online educational tool.
- While the current findings are encouraging, the number of controlled trials conducted on SMART is small and the studies have a number of significant methodological limitations.
- We recommend that SMART be evaluated with larger and more robustly designed trials.

working memory to improve working memory). Far transfer refers to the generalization of acquired skills from one domain to another (eg, training working memory to improve fluid intelligence; Soveri et al., 2017). A general consensus has emerged from the literature that (1) the evidence for the benefits of cognitive training is weak, and (2) where present, these improvements are limited to near transfer effects (Sala & Gobet, 2019). Many strategies for improving *general* cognitive ability (ie, such that there is far transfer to other domains) have been championed and trialled but ultimately have not been found to be effective, eg, chess (Sala et al., 2017; Sala & Gobet, 2016, 2017a), exposure to music (Sala & Gobet, 2017b), working memory training (Melby-Lervåg & Hulme, 2013, 2016; Melby-Lervåg et al., 2016; Sala & Gobet, 2017c), playing video games (Sala et al., 2018), and compensatory education (Abenavoli, 2019; McKey, 1985). Any benefits that have been detected in these studies tend to be limited to outcomes that are closely related to the tasks that were trained. In other words, working memory training might lead to some temporary performance improvement in tasks related to working memory (Melby-Lervåg et al., 2016). Thus, given the absence of evidence for the benefits of specific cognitive training tools or programs on enhancing general cognitive abilities, it is imperative that claims to the contrary be treated with caution and given careful scrutiny.

Recently, several studies have appeared showing that a commercially available computer-based training program called Strengthening Mental Abilities with Relational Training (SMART; Roche & Cassidy, 2020) has produced general IQ rises in the region of one to two standard deviations in children and adolescents (eg, Cassidy et al., 2016; Colbert et al., 2018; J. Hayes & Stewart, 2016). SMART is a theory-driven intervention informed by a contemporary behavioural account of language and cognition called Relational Frame Theory (RFT; Hayes et al., 2001). According to RFT, a core feature of intelligence is the ability to syllogistically relate stimuli based on symbolic (ie, verbal), as opposed to physical, properties. For example, when provided with specified relationships between arbitrary stimuli A:B and B:C (eg, A is more than B, B is more than C), language-able humans will readily *derive* novel relations among stimulus combinations. In this example, $B < A$, $C < B$, $A > C$ and $C < A$ relations will be derived without further training or feedback (Dymond et al., 2010). Importantly, these inferences are made in the absence of any supplementary information about the perceptual properties (eg, physical size) of the relations, showing that this process is symbolic and not due to the physical nature of the stimuli. In other words, the inferences are not attributable to a process of similarity-based stimulus generalization. Advocates of RFT suggest that relating stimuli based on their symbolic properties is central to general cognitive ability (McLoughlin et al., 2020a).

The proposal that relational reasoning is a core feature of cognition is consistent with recent findings from within cognitive psychology (Goldwater et al., 2018), linguistics (Everaert et al., 2015; Goldwater, 2017) and education (Alexander, 2019; Goldwater & Schalk, 2016). Alexander (2019) outlined four categories of relational reasoning that appear key in approaching common tasks and tests in educational settings (analogy, anomaly, antimony and antithesis). In mathematics, Farrington-Flint et al. (2007) found that tests of relational reasoning correlated with young children's mathematics ability; the authors reported that changes from domain-specific to domain-general relational reasoning ability predicted success in solving addition problems. Goldwater and Schalk (2016) also highlighted the link between relational reasoning and cognition as a potential pathway to enhance the efficacy of educational curricula. Taken together, these studies are indicative of an association between relational reasoning and performance in various educational domains. Accordingly, SMART explicitly targets domain-general relational reasoning as a means of improving global cognitive ability. The intervention does not use any material or stimuli that appear in school or university tests of English, mathematics, science and so on. It aims to establish fluent relational reasoning skills that can be applied to any educational domain or subject matter. Thus, in a

very real sense, SMART was developed with the aim of far transfer to the fore, where the benefits of the training would be seen in educational contexts and tests that do not appear, on the surface, to look in any way similar to SMART training in form or function.

During SMART training, participants are presented with relational tasks involving sets of nonsense words across 55 levels of increasing difficulty. Initial levels of the program include simple one-premise relations (eg, A is the same as B. Is B the same as A?), whereas later levels include a higher degree of relational complexity (A is more than B. B is more than C. Is C more than A?). For example, a learner will be presented with a statement such as: 'JUP is the same as HET, HET is opposite ORP', and then be asked to select 'Yes' or 'No' to a series of related questions (eg, 'Is ORP the same as JUP?') Specific stimulus combinations are not repeated across training levels so that the act of relational reasoning itself is targeted (and thus generalized). In simple terms, SMART trains the relations of same and opposite and more than/less than. These basic relational skills can then be applied to educational tasks in mathematics, science and vocabulary development. For example, school-based assessments often include test items such as 'What does irate mean' or 'think of another word for condolence' which, it is argued, consist of 'same as' relations (eg, 'irate' same as 'angry', or 'condolence' same as 'sympathy'). While SMART is conceptually consistent with and draws on findings from, a rich literature on relational reasoning (McLoughlin et al., 2020a; O'Connor et al., 2017), the extent to which it produces meaningful gains has not, as yet, been evaluated using meta-analytic techniques.

To date, there has been no systematic attempt to critically review or appraise the available evidence for SMART. Given the burgeoning empirical research base (eg, Cassidy et al., 2016; Colbert et al., 2018; McLoughlin et al., 2020b, 2021) and increased interest in SMART, a review and meta-analysis of the existing literature is clearly warranted at this juncture. In this systematic review and random effects meta-analysis, we appraise the existing literature on SMART. The purpose of the current review was to assess the impact of SMART on cognitive and academic outcomes (RQ1) and evaluate the content and methodological quality of the identified studies (RQ2).

METHOD

Protocol and registration

This systematic review protocol followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2015). The protocol was registered with PROSPERO (PROSPERO 2019: CRD42019132404).

Eligibility criteria

Studies were included if they met the following inclusion criteria:

- a. Full or in-part implementation of the SMART intervention.
- b. Studies reported original data with at least one outcome measure of IQ or academic performance using standardised instruments.
- c. Participants included children or adults without neurological, psychiatric and/or sensory impairments.
- d. Studies were available in English.
- e. Studies incorporated an experimental design: Randomised Control Trials (RCT), quasi-RCTs (where the quasi-random method of allocation was used), non-randomised designs

involving a control group, or single-case experimental designs (SCED). All single group pretest posttest designs, qualitative case studies, non-experimental case studies, theoretical and discussion papers were excluded. This was a deviation from our preregistered protocol which had initially specified the inclusion of single group pretest posttest designs. The inclusion of studies without a control group is likely to inflate the risk of test-retest effects or placebo effects (eg, Foroughi et al., 2016). The approach we took is more consistent with other recent reviews in the cognitive training literature (Sala et al., 2019; Sala & Gobet, 2017b, 2017c). No restrictions on publication date were specified.

Database searches

Searches were first undertaken in July 2019 using the following electronic databases: ERIC, PsycINFO, PubMed, Applied Social Sciences Index and Abstracts (ASSIA), Scopus and Proquest Psychology. Search strings were developed using the words: 'relational frame theory', 'smart' and 'iq'. For example, '*relational frame theory*' AND ('*strengthening mental abilities with relational training*' OR *smart*) AND (*iq* OR *i.q.*). The term, 'Relational Frame Theory' was included given that the theory is foundational to the development of SMART. We conducted forward and backwards searches of all included studies. We identified unpublished or ongoing trials by contacting all corresponding authors of included studies. To maximise the currency of the review updated searches were performed in February 2021 using the last year of the original search as the beginning date for the update.

Review strategy

The initial electronic database search produced 155 records. This list was screened for duplications resulting in 136 unique records. Two graduate-level reviewers independently reviewed the titles and abstracts in accordance with the study inclusion criteria specified above. This resulted in 133 agreements (97.8% agreement, $\kappa = 0.85$). Disagreements were resolved by reviewing the full text of the paper in consultation with a third member of the research team. Following an initial screening, nine studies were selected for full-text review. All papers were read by two members of the research team. A third member of the research team was consulted for one of the papers in which the eligibility was unclear. Three papers were selected for inclusion in the review. Two additional studies were identified following contact with the corresponding authors of the nine studies selected for full-text review. The updated database searches produced nine additional records. This list was screened for duplication resulting in four unique records. Two reviewers independently reviewed the titles and abstracts in accordance with study inclusion criteria. This resulted in four agreements (100% agreement). No additional papers were deemed eligible for the review.

At the end of the screening process, five studies in total had met the criteria for inclusion (the process is summarised in Figure 1 using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart). The summary characteristics of these studies are presented in Table 1.

Data extraction and classification

The following data were extracted for each article where available: (a) Author and year of publication); (b) study design and study characteristics (sample size, recruitment method, type of control group, outcome measure(s), time of assessments, number of levels completed,

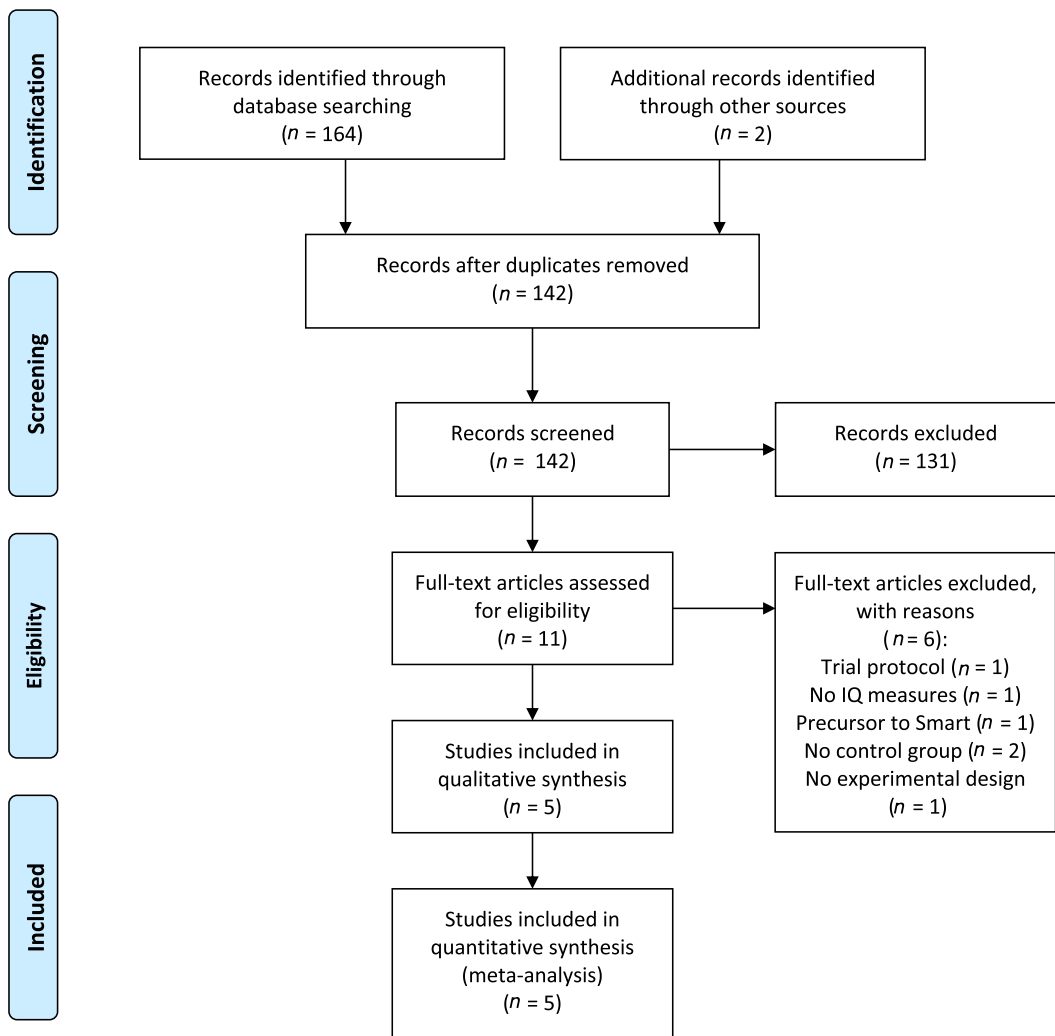


FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-analyses flow diagram of the decision process for included studies

and duration of the intervention), and sample characteristics (mean age and gender); (c) setting (country and context); (d) participants dropout rate and missing data handling; (e) attrition rate; and (f) data needed to calculate effect sizes (means and standard deviations). Where information was missing from the paper, corresponding authors were contacted to obtain the information. Two members of the research team independently extracted the data to ensure the accuracy of data summaries and provide a measure of inter-coder agreement. The initial agreement was obtained on 132 (94%) out of 140 items. When there was a disagreement on information extracted, the authors reviewed and discussed the studies until agreement was reached.

Quality assessment

We assessed the validity of the included trials using the Cochrane Collaboration Risk of Bias 'RoB 2' tool (Sterne et al., 2019). Two researchers independently scored all studies. When

TABLE 1 Descriptive characteristics of the studies (k = 5)

Study	Demographics			Comparison	Instrument	Randomization		Intervention		Adherence		Dropout Rate
	Country	Age	Sex			Exp.	Control	Stages	Intensity	With.	Exc.	
Colbert et al. (2018)	Ireland	M = 16.5 (SD = 0.67)	15 = f 11 = m	Regular classroom activities	WASI	12	14	M = 41.6 (SD = NR)	45 minutes per session	0	0	0%
Hayes and Stewart (2016)	Ireland	M = 10.34 (SD = 0.45)	13 = f 15 = m	Scratch coding	WASI; WIAT-II (3 × subtests); WISC-IV (2 × subtests); DPRT; DPMT; RAI	14	14	NR	1-hour session x 29	0	0	0%
Thirus et al. (2016)	Sweden	M = 17.3 (SD = 0.67)	27 = f 8 = m	Regular classroom activities	RAI; RPM; UMT	Original: 18 N = 10	Original: 17 N = 11	M = 39.2 (SD = 16.86)	N/A	9	5	40%
McLoughlin et al. (2021)	Ireland	M = 8.67 (SD = 0.91)	33 = f 16 = m	Chess condition (Dora Logic Ltd., 2018)	KBIT-2 (sub non-verbal); DPRT; DPMT	Original: NR N = 30	Original: NR N = 19	M = 14.5 (SD = 11.20)	240 minutes	5	1 MD	11%
McLoughlin et al. (2020b)	Ireland	M = 13.47 (SD = 0.62)	NR	Scratch coding	KBIT-2 (sub non-verbal); DSST	Original: 93 N = 43	Original: 82 N = 27	M = 16.4 (SD = 8.40)	240 minutes	57	45	60%

Note: All studies included were Parallel Randomised Control Trials. Abbreviations: DPMT, Drumcondra Primary Mathematics Test—Revised; DPRT, Drumcondra Primary Reading Test—Revised; DSST, Digital Symbol Substitution Test; Exc., Exclusion; Exp., Experiment; KBIT-2, Kaufman Brief Intelligence Test Second Edition; MD, Missing Data; NR, Not Reported; RAI, Relational Abilities Index; RPM, Ravens Progressive Matrices; UMT, Unstandardised mathematics test; WASI, Wechsler Abbreviated Scale of Intelligence; WIAT-II, Wechsler Individual Achievement Test Second Edition; WISC-IV, Wechsler Intelligent Scale for Children Fifth Edition; With., Withdrawal.

there was a disagreement with respect to scoring, the scorers discussed the studies until agreement was reached. Each study was evaluated according to six criteria: (a) random sequence generation (selection bias); (b) allocation concealment (selection bias); (c) blinding of participants (performance bias); (d) blinding of outcome assessment (detection bias); (e) incomplete outcome data (attrition bias); and (f) selective outcome reporting (reporting bias). Studies were rated as having 'low', 'high' or 'some' risk of bias in each of the domains specified above. Studies were rated as having an overall low risk of bias if they scored a 'low' in all domains.

Effect-size calculation

Effect sizes were grouped into five categories based on the Cattell-Horn-Carroll taxonomy (McGrew, 2009): *non-verbal ability* (eg, fluid intelligence, spatial reasoning), *verbal ability* (eg, vocabulary, reading, spelling), *memory* (eg, working memory tasks), *speed* (eg, processing speed) and *mathematics*. Effect sizes were calculated from the data reported in each article. Two authors independently extracted all means and standard deviations from the relevant papers. If the relevant means and standard deviations were not available from the data reported in the papers, the authors were contacted for additional details. Where this occurred, data was provided in all instances. We calculated the effect sizes as the standardised mean difference of the pretest and posttest outcome between experimental and control groups. As the majority of studies included in the meta-analyses comprised small numbers of participants, we calculated Hedges's *g* based on Morris (2008)—an adjusted standardised mean difference—which adjusts for small sample bias. The summary statistics required for each outcome were the number of participants in the experimental and control groups, the mean outcome in each group, and the pooled pretest standard deviations. We assumed a pre–posttest correlation of $r = 0.6$ (Morris, 2008). Where two or more outcomes (eg, several subscales) were reported for the same domain (eg, block design and matrices as a measure of non-verbal ability), we calculated a mean effect size and the corresponding variance (Borenstein et al., 2010). All statistical analyses were performed using R Statistical Software (R Core Team [2020]; R Foundation for Statistical Computing, Vienna, Austria).

Meta-analytic procedure

All meta-analyses were performed using the *dmetar* (Harrer et al., 2019), *Metafor* (Viechtbauer, 2010), and *meta* (Balduzzi et al., 2019) R packages. All other statistical procedures were conducted in R. We used a Random Effects Meta-analysis. Outcomes for which fewer than three studies were available were not synthesised. An alpha of $\alpha = 0.05$ was employed for all analyses, with the size of the effect interpreted as small, moderate, and large when at 0.2, 0.5, and 0.8 respectively (Cohen, 1988). To estimate the distribution of all calculated effect sizes, we calculated prediction intervals (95%; Borenstein et al., 2011). Heterogeneity for each meta-analysis was assessed using the I^2 statistic (Higgins & Thompson, 2002) and was considered low, moderate, and substantial when at 25%, 50%, or 75%, respectively (Higgins et al., 2003). In addition, we examined estimates of the between-studies variance using the τ^2 statistic. As we identified only five studies that met the inclusion criterion for effect size pooling, we did not undertake funnel plots to assess reporting bias. With fewer than 10 studies, the statistical power of the tests is too low to detect meaningful asymmetry (Higgins et al., 2011). In our systematic review protocol, we had planned to perform a regression analysis using the number of completed stages of

SMART as a predictor variable; however, given the small number of identified studies any meta-regression would be minimally informative (Higgins & Thompson, 2004).

We undertook additional exploratory analysis in the form of influence analyses. Specifically, we conducted ‘leave-one-out’ analyses (Viechtbauer & Cheung, 2010). This procedure involved recalculating the pooled effect size while leaving out one study at a time. This allowed us to estimate the influence of individual studies on the overall effect. Following the influence analyses, we performed the meta-analysis again whilst omitting the study exerting the greatest influence on the pooled effect. Finally, we undertook a subgroup analysis in which we calculated the standardised mean difference for those studies in which SMART was compared to an active control group.

RESULTS

The characteristics of the included studies are presented in Table 1. In total 319 participants were randomised in the studies; however, after withdrawals and exclusions, data from 195 participants were included in the analyses across the studies. The sample size within individual studies ranged from $N = 21$ to 175 at randomization, and from $N = 21$ to 70 at analyses. The mean age of participants ranged from 8.67 (SD = 0.91) to 16.5 (SD = 0.67) years old. All of the trials were conducted with typically developing participants in school settings with the exception of one study in which the setting was unspecified (Thirus et al., 2016).

The outcome measures reported in the studies are presented in Table 1. The most commonly employed instruments were the Wechsler Abbreviated Scale of Intelligence (WASI; $k = 2$), and the Kaufman Brief Intelligence Test (KBIT-2; $k = 2$). The Wechsler Individual Achievement Test—Second UK Edition (WIAT-II), Wechsler Intelligence Scale for Children Fourth UK Edition (WISC-IV), Ravens Progressive Matrices (RPM), and the Digit Symbol Substitution Test (DSST) were all used once. In addition to the standardised IQ instruments, other measures included the Drumcondra Primary Reading Test (DPRT; $k = 2$), Drumcondra Primary Mathematics Test (DPMT; $k = 1$), which are national standardised tests of the Irish primary school curriculum. A measure employed within the SMART program called the Relational Abilities Index (RAI), was also reported by two studies. In two of the studies, SMART was compared with a passive control group in which participants engaged in regular classroom activities (Colbert et al., 2018; Thirus et al., 2016). In three studies, SMART was compared to an active control. Two studies compared SMART to a coding training program called Scratch (Hayes & Stewart, 2016; McLoughlin et al., 2020b), and one involved a control condition in which participants played chess (McLoughlin et al., 2021).

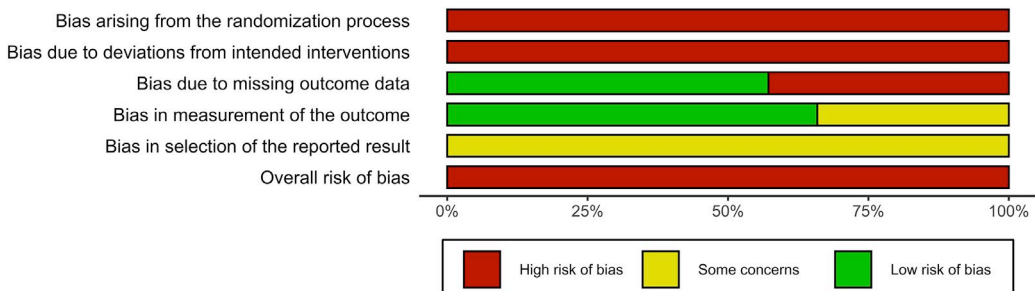


FIGURE 2 Risk of bias summary

Risk of bias

All five of the studies received an overall high risk of bias rating according to the Cochrane Collaboration 'RoB 2' tool (Sterne et al., 2019). Figure 2 shows the overall percentages of studies with high, low or some concerns risk of bias in each of the six domains.

Main analyses

Five studies reported outcome measures that were categorised as assessing *non-verbal ability* according to the Cattell-Horn-Carroll taxonomy (McGrew, 2009). We did not pool any further outcome measures for the purpose of undertaking a meta-analysis, as no more than two studies reported standardised measures from within the same Cattell-Horn-Carroll taxonomy category: *verbal ability* ($k = 2$), *mathematics* ($k = 2$), *memory* ($k = 1$) and *speed* ($k = 1$).

Non-verbal ability

We compared the effects of SMART on non-verbal ability where passive and active control groups were combined into a single comparator group. As part of the effect size calculations, we used single outcome measures reported in the original studies with the exception of Hayes and Stewart (2016) where the authors reported multiple outcome measures from the same domain. We computed a composite effect size and corresponding variance for the Matrices and Block Design outcomes (cf. Borenstein et al., 2011). The overall effect size was $g = 0.57$ (95% CI [0.24, 0.89], $p = 0.001$) with low-to-moderate heterogeneity ($I^2 = 34\%$; 95% CI [0%; 75.3%]). The prediction interval ranged from $g = -0.30$ to 1.44. Figure 3 shows the forest plot for the individual and overall effect size estimates. Similar effects were obtained in a sensitivity analysis in which the pretest and posttest correlations were adjusted from 0.6 to 0.4 (Borenstein et al., 2011). A significant effect was found when leaving one study out (McLoughlin et al., 2021) in the influence analysis ($g = 0.43$, 95% CI [0.14; 0.72], $p = 0.01$) with heterogeneity remaining low ($I^2 = 0\%$; 95% CI [0%; 84.5%]). The prediction interval ranged from $g = -0.21$ to 1.07. Table 2 shows a summary of the pooled effect sizes for each meta-analysis.

We also compared the effects of SMART in studies in which the intervention was compared to an active control group ($k = 3$). The analysis revealed a significant effect size of $g = 0.53$ (95% CI [0.11, 0.96], $p = 0.02$) with moderate heterogeneity ($I^2 = 50.3\%$; 95% CI

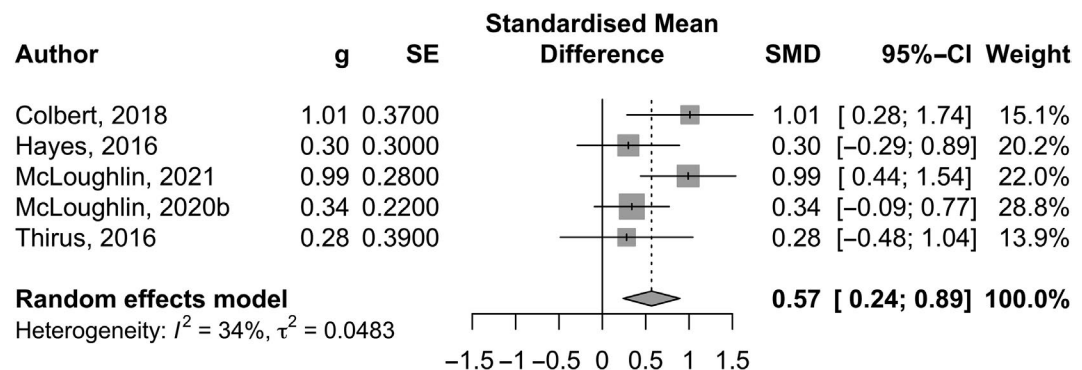


FIGURE 3 Forest plot for the effects of SMART on Non-Verbal IQ measures

TABLE 2 Pooled effects of ($k = 5$)

Outcome	n_c	Effect size			Heterogeneity			
		g	95% CI	p	I^2	95% CI	τ^2	95% PI
Non-verbal ability	5	0.57	[0.24,0.89]	0.001	34	[0,75]	0.048	[-0.30, 1.44]
Influence analysis ^a	4	0.43	[0.14,0.72]	0.004	0	[0,85]	0.001	[-0.21, 1.07]
Only active control	3	0.53	[0.11, 0.96]	0.014	50	[0,85]	0.071	[-3.83, 4.90]

Note: n_c : number of comparisons.

Abbreviation: CI, confidence intervals; PI, prediction intervals.

^aLeave-one-out analysis—removed McLoughlin et al. (2021).

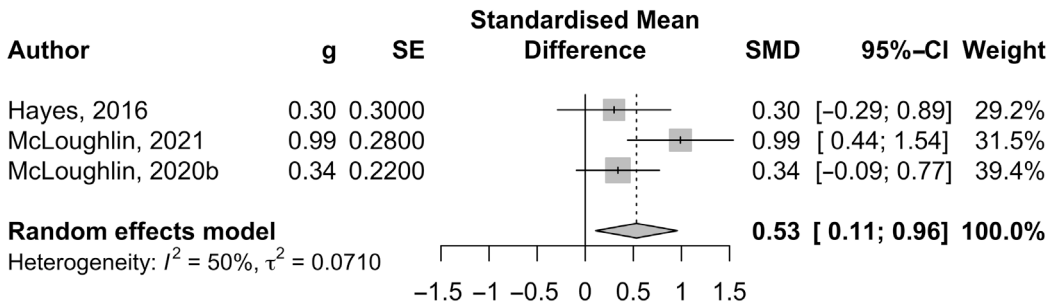


FIGURE 4 Forest plot for the effects of SMART on Non-Verbal IQ only including studies with an active control

[0.0%; 85.6%]. The prediction interval ranged from $g = -3.83$ to 4.90. Figure 4 shows the forest plot for the individual and overall effect size estimates.

Verbal ability

Two studies evaluated the effects of SMART on verbal abilities. Colbert et al. (2018) reported a single outcome measure for verbal IQ; however, Hayes and Stewart (2016) reported multiple outcome measures. As before, we computed a composite effect size and corresponding variance. This composite score was obtained by combining the outcomes from the Similarities, Vocabulary, Spelling and English subtests from the WASI and WIAT and the DPRT scores. The calculated effects were $g = 1.43$ (95% CI [0.65, 2.21]; Colbert et al., 2018), and $g = 0.48$ (95% CI [-0.07, 1.03]; Hayes & Stewart, 2016).

Mathematics

Two studies evaluated the effects of SMART on mathematics performance. Both studies reported a single outcome measure in this domain. The calculated effects were $g = -0.09$ (95% CI [-0.85, 0.67]; Thirus et al., 2016), and $g = 0.24$ (95% CI [-0.44, 0.92]; Hayes & Stewart, 2016).

DISCUSSION

We identified five studies that met the inclusion criteria; three that compared the effects of SMART to an active control group, and a further two that compared SMART to

treatment as usual (TAU). All five studies reported at least one standardised measure of nonverbal IQ (eg, matrices, block design). When compared to a combination of passive and active controls ($k = 5$), we estimate a moderate overall impact of SMART on non-verbal IQ ($g = 0.57$, 95% CI [0.24, 0.89]), with moderate heterogeneity ($I^2 = 34\%$). When compared in studies involving only active controls ($k = 3$), we also found a statistically significant effect of SMART on standardised measures of non-verbal IQ ($g = 0.53$, 95% CI [0.11, 0.96]). In both analyses, however, the prediction intervals indicate substantial uncertainty as to range of effect sizes that might be expected in future trials. We were unable to ascertain the impact of SMART on any other cognitive or academic domain. In summary, while these findings provide some tentative support for the benefits of SMART training on measures of nonverbal intelligence, further studies are needed to capture the true impact of the intervention.

Our results suggest that there is preliminary evidence that SMART training produces some cross-domain transfer. The tasks employed in SMART training are formally dissimilar from the tasks utilised in the nonverbal IQ subtests. SMART consists of a series of verbal reasoning tasks, whereas nonverbal tests of intelligence such as Ravens Matrices require participants to complete a sequence of geometrical shapes. This finding is interesting given that, as reviewed earlier, existing syntheses of the evidence on cognitive or so-called 'brain-training' interventions have indicated that there is little evidence for this type of 'far transfer' (Melby-Lervåg et al., 2016; Sala et al., 2019).

Limitations in the studies

We identified a number of methodological limitations in the literature. According to Gobet and Campitelli (2006), one of the features of an 'ideal design' in a randomised trial of educational interventions is the use of both an active and passive (a do-nothing) control group. Consistent with the overwhelming majority of studies in the education literature, none of the studies in the present review met this criterion. Two of the five studies included in our review did not incorporate an active control group. The absence of an active control presents a challenge in determining whether improvements arose from the application of the intervention or because of placebo effects (cf. Colbert et al., 2018). We found a high degree of uncertainty in the effect of SMART training, particularly when we synthesised studies involving active controls, which adds further weight to this concern. In echoing calls by Simons et al. (2016) and Melby-Lervåg et al. (2016), we strongly recommend that future research on cognitive training interventions (including SMART) involve 'treated' control groups.

Risk of bias was formally evaluated using Cochrane RoB2 assessment, which revealed that all of the included studies carried an overall rating of a 'high risk' of bias.

One consistent weakness across all five of the studies was the absence of double-blinding. In all trials, participants were aware of the intervention they were assigned to. This further increases the risk of expectancy effects discussed earlier. Indeed, the issue is particularly salient given that the purpose of SMART is made explicit in the presentation format; participants accessed the intervention via the domain name 'www.raiseyouriq.com'. In this case, those in the SMART training conditions might expect to perform better after training compared to before (making the training appear to work even if it does not), and to have higher expectancy effects the more times they log in (making it look like the number of stages completed matters even if it does not). There are challenges inherent in arranging double-blinding in trials of educational interventions (Hutchinson & Styles, 2010), as once participants begin the intervention it is usually clear to which intervention they have been assigned; however, researchers might adopt creative means of minimising the impact of such expectancy effects. As the theory behind SMART posits a very particular type of training as

critical (ie, relational reasoning), researchers could incorporate a 'sham' control condition in which the key features of the activities are altered while other aspects of the training environment remain identical (eg, similar domain name, same format, etc). A further significant concern was the relatively large number of participants who dropped out or whose data were not analysed for other reasons following randomisation. An intention-to-treat analysis was not used in any of the studies which is a significant limitation. Specifically, there may have been undetected differences between those that completed SMART and those in the control groups that were a function of the relatively high rate of attrition.

All of the studies limited the posttest measures to assessments that were conducted immediately following the intervention. As a result, we were not able to determine the longevity of any improvements following SMART training. This is an important limitation as previous findings have highlighted that benefits derived from cognitive training can be short-lived. Melby-Lervåg et al. (2016) reported that improvements on verbal tasks following working memory training disappeared following a period of just a few months, similar to the 'wash-out' effects seen in compensatory education programs (McKey, 1985). In addition to these issues, only two studies included broader educational outcome measures. These findings highlight the need for future studies to incorporate assessment methods to determine whether SMART training achieves sustained, far transfer of training effects (ie, does SMART improve real-world outcomes?). The studies involved relatively small numbers of participants. The average number of participants in the SMART and control conditions was 21.8 (SD = 14.3) and 14.3 (SD = 6.3) respectively, with three of the five trials involving less than 15 participants in each of the trial arms. Underpowered trials are more likely to inflate the effect size estimates ('Winner's Curse'; Button et al., 2013; Gelman & Carlin, 2014; Ioannidis, 2008). We recommend that future research on SMART aims to undertake adequately-powered trials in order to generate more precise effect size estimates.

There were large discrepancies in the number of SMART stages completed and participant drop-out across studies, making it difficult to compare effect sizes directly. Regarding the results of the present study, the wide confidence and prediction intervals should perhaps be interpreted with these factors in mind. In the two largest SMART studies to date (McLoughlin et al., 2020b, 2021), the authors found that attrition rates were related to baseline levels of trait negative emotion and trait agreeableness. It is possible that future large-scale SMART studies might benefit from providing additional supports for participants found to be relatively low in agreeableness or high in negative emotion to ensure that they stay motivated to take part. On the other hand, one-to-one support may not be a realistic, affordable, or scalable solution to the attrition problem. It may have been the case in the smaller scale SMART studies that the availability of one-to-one support accounted for their low attrition. Perhaps broader group-level schedules of reinforcement for taking part could be applied; it is of course an empirical question as to which methods work best. In considering these factors, however, we must separate the question of how effective SMART is when it is completed in full, from how effective SMART is in global terms as an overall package (the motivational contingencies inherent in the training, the difficulty of the task etc.) when implemented on a larger scale.

The putative effects of SMART have been independently replicated. Nonetheless, the largest of these independent replications (McLoughlin et al., 2020b, 2021) suffer from a number of important limitations that other studies do not, including higher attrition and lower training completion. Although these authors made a number of methodological improvements over previous studies and did not have a financial interest in SMART, their publication history also shows evidence of prior allegiance to a behaviour-analytic worldview (indeed, this cautionary note applies to all SMART studies to-date except Thirus et al. [2016]). For these reasons, future large-scale studies might have greater face validity by employing the services of an independent clinical trials unit with no vested interests whatsoever.

Strengths

There were a number of key strengths in our methodology which help to ensure that the review and meta-analyses make a meaningful contribution to the literature. First, we employed relatively stringent inclusion criteria in order to overcome some of the existing methodological limitations in the literature on SMART. For example, we sought to include only those studies that had utilised the SMART programme in its commercially available incarnation. Doing so ensured that we could be confident that there was a high degree of consistency in the application of the intervention across the studies that were synthesised. Second, we excluded studies that involved either correlational or single-group pretest–posttest designs. Studies lacking a control group are more likely to lead to spurious conclusions with respect to the efficacy of an intervention (eg, test-retest effects or placebo effects; Foroughi et al., 2016). This approach is consistent with other recent reviews in the cognitive training literature (Sala et al., 2019; Sala & Gobet, 2017b, 2017c). In practice, this meant that a number of studies that have been conducted on SMART were ineligible for inclusion in the review (Amd & Roche, 2018; Cassidy et al., 2016). The findings from these evaluations are broadly consistent with the present synthesis. Cassidy et al., reported full-scale IQ increases of around one standard deviation in adolescents who had undertaken SMART for a period of around three months. Similarly, Amd and Roche (2018) found pre to posttest increases in fluid intelligence following SMART training in a cohort of 7- to 13-year olds.

Finally, a review protocol setting out the main design features, search strategy and planned analysis was preregistered on the PROSPERO database (albeit with some deviations that we have reported here). Publicly available preregistration helps guard against biases in the review process by making the planned methodology as transparent as possible.

Limitations

One key issue was the number of studies that met the inclusion criteria. The fact that we identified five studies comprising a cumulative total of 194 participants represents a weakness of the current literature. With respect to the meta-analysis, in particular, a greater number of studies would have improved the precision of the effect size estimates. The uncertainty in the variation in effects is reflected in the intervals reported in our measures of heterogeneity. Prediction intervals in particular can be helpful in capturing uncertainty in the estimation of an effect (IntHout et al., 2016). While confidence intervals summarise information about the mean effect and related precision, prediction intervals should be interpreted as an estimate of the distribution of effect sizes across settings. Accordingly, prediction intervals can be useful in estimating the range of effect sizes that might be expected in future settings. In our analysis, the calculated prediction intervals for all three effect size estimates included both null and deleterious effects of SMART. These estimates reflect high uncertainty as to the distribution of the true effects of SMART, and conclusions need to be interpreted with this in mind.

The small number of available studies also precluded analyses of several potentially important variables. For example, we were unable to systematically evaluate the impact of factors such as the intensity and frequency (ie, dose) as potential mediators of the intervention by using a meta-regression. The extent to which dose is a useful predictor of improvement is an important future research question. For example, it is possible that the effect of SMART training might be related to the number of training stages completed overall, with higher training completion producing larger increases in IQ than those with lower training completion. In the studies reviewed here, those with larger samples had lower training completion and retention, highlighting the difficulty of recruiting for larger-scale trials in this area. Given

the small number of studies, we were also not able to ascertain the presence or absence of publication bias in the literature. In addition, the review was limited in that we focused our analysis on a single intervention rather than examining the benefits of relational reasoning interventions on cognitive outcomes more generally. While this strategy necessarily makes the findings narrower in scope, it meant that we had greater confidence in the characteristics of the intervention. Finally, while evaluating sources of bias in systematic reviews is considered best practice, it should be acknowledged that the Cochrane Risk of Bias tool is typically used for assessing bias in trials undertaken in clinical settings rather than in cognitive or educational domains, and thus, the rating of risk might be interpreted as overly conservative here.

Implications and conclusions

Overall, there is a clear trajectory of methodological improvement in the SMART literature, with consistently positive results even as tighter controls have been implemented. While the data are promising, it is arguably not justifiable to conduct more studies that do not address some of the limitations that we have outlined here, as for consumers who might not be well trained in the evaluation of evidence (eg, teachers), 'more studies' can be seen as meaning 'better evidence'. In this way, publishing additional uncontrolled studies lacking the appropriate power could serve to obscure the literature in a misleading way, rather than to provide a better estimate of the true efficacy of the intervention, whatever that may be. Nonetheless, SMART is a promising, theoretically-plausible and empirically-grounded approach to increasing general cognitive ability (Dymond & Roche, 2013; McLoughlin et al., 2020a). At this juncture, we believe that the existing literature provides sufficient justification for further investment in large-scale trials.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare that are relevant to the content of this article.

ETHICS STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. An ethics statement is not applicable as all data and analyses was based on the published literature.

DATA AVAILABILITY STATEMENT

The data and data analysis scripts are available at: https://osf.io/d94bj/?view_only=645483b3de1b46f5a6fb8830fe1e53d6 and the systematic review was preregistered with PROSPERO (PROSPERO 2019: CRD42019132404).

REFERENCES

- Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, 145(12), 1103–1127. <https://doi.org/10.1037/bul0000212>
- Alexander, P. A. (2019). Individual differences in college-age learners: The importance of relational reasoning for learning and assessment in higher education. *British Journal of Educational Psychology*, 89(3), 416–428. <https://doi.org/10.1111/bjep.12264>
- Amd, M., & Roche, B. (2018). Assessing the effects of a relational training intervention on fluid intelligence among a sample of socially disadvantaged children in Bangladesh. *The Psychological Record*, 68(2), 141–149. <https://doi.org/10.1007/s40732-018-0273-4>
- Balduzzi, S., Rucker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, 22, 153–160. <https://doi.org/10.1136/ebmental-2019-300117>

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470743386>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376. <https://doi.org/10.1038/nrn3475>
- Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences, 47*, 222–235. <https://doi.org/10.1016/j.lindif.2016.03.001>
- Cohen, J. (1988). *Statistical power for the social sciences*. Laurence Erlbaum and Associates.
- Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education, 27*(4), 509–531. <https://doi.org/10.1007/s10864-018-9302-2>
- Dymond, S., May, R. J., Munnely, A., & Hoon, A. E. (2010). Evaluating the evidence base for relational frame theory: A citation analysis. *Behavior Analyst, 33*(1), 97–117. <https://doi.org/10.1007/BF03392206>
- Dymond, S., & Roche, B. (2013). *Advances in relational frame theory: Research and application*. Context Press.
- Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., & Bolhuis, J. J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences, 19*(12), 729–743. <https://doi.org/10.1016/j.tics.2015.09.008>
- Farrington-Flint, L., Canobi, K. H., Wood, C., & Faulkner, D. (2007). The role of relational reasoning in children's addition concepts. *British Journal of Developmental Psychology, 25*(2), 227–246. <https://doi.org/10.1348/026151006X108406>
- Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7470–7474. <https://doi.org/10.1073/pnas.1601243113>
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651. <https://doi.org/10.1177/1745691614551642>
- Gobet, F., & Campitelli, G. (2006). Educational benefits of chess instruction. A critical review. In T. Redman (Ed.), *Chess and education. Selected essays from the Kolanowski Conference* (pp. 124–143). University of Texas at Dallas.
- Goldwater, M. B. (2017). Grammatical constructions as relational categories. *Topics in Cognitive Science, 9*(3), 776–799. <https://doi.org/10.1111/tops.12272>
- Goldwater, M. B., Don, H. J., Krusche, M. J. F., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General, 147*(1), 1–35. <https://doi.org/10.1037/xge0000387>
- Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin, 142*(7), 729–757. <https://doi.org/10.1037/bul0000043>
- Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmeter: Companion R Package for the guide "doing meta-analysis in R"* (R package version 0.0, 9000). <http://dmeter.protectlab.org/>
- Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *The British Journal of Educational Psychology, 86*(3), 397–411. <https://doi.org/10.1111/bjep.12114>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. Plenum Press.
- Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Online), 343*, d5928. <https://doi.org/10.1136/bmj.d5928>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine, 23*(11), 1663–1682. <https://doi.org/10.1002/sim.1752>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hutchinson, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. NFER.
- IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7), e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640–648. <https://doi.org/10.1097/EDE.0b013e31818131e7>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>

- McKey, R. H. (1985). *The Impact of Head Start on Children, Families and Communities. Final Report of the Head Start Evaluation, Synthesis and Utilization Project*. <https://files.eric.ed.gov/fulltext/ED263984.pdf>
- McLoughlin, S., Tyndall, I., & Pereira, A. (2020a). Convergence of multiple fields on a relational reasoning approach to cognition. *Intelligence*, 83(Nov–Dec), 101491. <https://doi.org/10.1016/j.intell.2020.101491>
- McLoughlin, S., Tyndall, I., & Pereira, A. (2020b). Relational operant skills training increases standardized matrices scores in adolescents: A stratified active-controlled trial. *Journal of Behavioral Education*. <https://doi.org/10.1007/s10864-020-09399-x>
- McLoughlin, S., Tyndall, I., Pereira, A., & Mulhern, T. (2021). Non-verbal IQ gains from relational operant training explain variance in educational attainment: An active-controlled feasibility study. *Journal of Cognitive Enhancement*, 5(1), 35–50. <https://doi.org/10.1007/s41465-020-00187-z>
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270–291. <https://doi.org/10.1037/a0028228>
- Melby-Lervåg, M., & Hulme, C. (2016). There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin & Review*, 23(1), 324–330. <https://doi.org/10.3758/s13423-015-0862-z>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer” evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Estarli, M., Barrera, E. S. A., Martínez-Rodríguez, R., Baladia, E., Agüero, S. D., Camacho, S., Buhring, K., Herrero-López, A., Gil-González, D. M., Altman, D. G., Booth, A., ... Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Revista Espanola de Nutricion Humana y Dietetica*, 4(1), 1. <https://doi.org/10.1186/2046-4053-4-1>
- Morris, S. B. (2008). Estimating effect sizes from Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>
- O'Connor, M., Farrell, L., Munnely, A., & McHugh, L. (2017). Citation analysis of relational frame theory: 2009–2016. *Journal of Contextual Behavioral Science*, 6(2), 152–158. <https://doi.org/10.1016/j.jcbs.2017.04.009>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Roche, B., & Cassidy, S. (2020). *RaiseYourIQ*. www.raiseyouriq.com
- Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, 5(1), 18. <https://doi.org/10.1525/collabra.203>
- 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, 1–4. <https://doi.org/10.3389/fpsyg.2017.00238>
- 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. <https://doi.org/10.1016/j.edurev.2016.02.002>
- Sala, G., & Gobet, F. (2017a). Does chess instruction improve mathematical problem-solving ability? Two experimental studies with an active control group. *Learning and Behavior*, 45(4), 414–421. <https://doi.org/10.3758/s13420-017-0280-3>
- Sala, G., & Gobet, F. (2017b). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, 20, 55–67. <https://doi.org/10.1016/j.edurev.2016.11.005>
- Sala, G., & Gobet, F. (2017c). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, 53(4), 671–685. <https://doi.org/10.1037/dev0000265>
- Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences*, 23(1), 9–20. <https://doi.org/10.1016/j.tics.2018.10.004>
- Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, 144(2), 111–139. <https://doi.org/10.1037/bul0000139>
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest*, 17(3), 103–186. <https://doi.org/10.1177/1529100616661983>
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, 24(4), 1077–1096. <https://doi.org/10.3758/s13423-016-1217-0>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *The BMJ*, 366, i4898. <https://doi.org/10.1136/bmj.i4898>
- Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology and Psychological Therapy*, 16(2), 141–155.

- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., & Cheung, M.-W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>

How to cite this article: May, R. J., Tyndall, I., McTiernan, A., Roderique-Davies, G., & McLoughlin, S. (2022). The impact of the SMART program on cognitive and academic skills: A systematic review and meta-analysis. *British Journal of Educational Technology*, 00, 1–18. <https://doi.org/10.1111/bjet.13192>