From hand to eye:**A meta-analysis of the benefit from handwriting training in visual graph recognition**

Susana Araújo^{1*}, Miguel Domingues, & Tânia Fernandes¹

Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal

Author Note

Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal. ORCID: 0000-0002-0354-3348; 0000-0001-8498-8500; 0000-0002-4453-7251

This work was supported by the Research Center for Psychological Science (CICPSI) at Universidade de Lisboa, by the Portuguese Foundation for Science and Technology, FCT (ref: PTDC/PSI-GER/3281/2020), and FEDER (POR Lisboa 2020; ref: PTDC/PSI-GER/28184/2017), and the IF-2015 Program (IF/00533/2015). Financial sources had no involvement in the research. We thank Luísa Corbal for assistance in assembling a preliminary dataset.

¹ These two authors had equal contribution to this work.

* **Corresponding author:** Susana Araújo, Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal. Fax number: +351 217 933 408; phone number: +351 217 943 871; e-mail: smaraujo@psicologia.ulisboa.pt.

Conflicts of interest/Competing interests: None.

Data Sharing: Databases used in this work are available at https://osf.io/wzdxn/?view_only=f31099e685454157b700955c4486a085

Authors' contribution: Conceptualization: T. Fernandes and S. Araújo; Methodology: S. Araújo and T. Fernandes; Formal analysis and investigation: M. Domingues, S. Araújo, T. Fernandes; Validation: S. Araújo and T. Fernandes; Writing - original draft preparation: T. Fernandes and S. Araújo; Writing - review and editing: S. Araújo, M. Domingues, T. Fernandes; Funding acquisition: T. Fernandes and S. Araújo; Supervision: T. Fernandes and S. Araújo.

Abstract

Handwriting (HW) training seems to boost recognition of visual graphs and learning to read more than other learning experiences. However, effects across studies appear to be variable and the underlying cognitive mechanism has been elusive. We thus conducted a meta-analysis on 50 independent experiments (with 1525 participants) to determine the magnitude of this HW benefit in visual graph recognition, while enlightening the underlying cognitive mechanism, by investigating four types of moderators: training program (type of control training, presence/absence of phonological training, and HW tasks adopted); set size and training regime (duration and frequency of session and total amount of training); granularity of visual discrimination and perceptual learning tasks; and age of participants. The benefit from HW training was moderate-to-large and significant (Hedge's $g = 0.58$, $SE = .09$) and was also modulated by type of control training (larger relative to motor, $g = 0.78$, than to visual control, $g = 0.37$), phonological training (larger when it was absent, $g = 0.79$, than present, $g = 0.47$), and granularity of visual discrimination (larger for fine-grained, $g = 0.93$, than coarse-grained, $g = 0.19$). These results are consistent with symbolic accounts that hold that the advantage from HW training in visual graph recognition is about perceptual learning rather than the motor act. Multiple meta-regressions also revealed that training regime moderated the HW benefit. We conclude that HW training is effective to improve visual graph recognition, and hence, is still relevant for literacy instruction in the present digital era.

Keywords: meta-analysis; sensorimotor training; handwriting; letter recognition; reading.

Introduction

Writing systems are impressive cultural inventions. Thousands of hours are devoted to reading (Pelli, Burns, Farell, & Moore-Page, 2006) given that literacy is a pillar of modern democracies. Efficient processing and recognition of *graphs*¹, i.e., characters of a written script which code linguistic units as phonemes, syllables, or morphemes (e.g., letters in the Latin alphabet, *kanas* in Japanese), is the foundation of reading from the beginning of learning until mastering it (Grainger, 2018). Reading thus implies *phonological recoding*, that is, mapping of orthographic into phonological representations (e.g., Share, 1995; Ziegler & Goswami, 2006), but before such mapping can occur, learners need to be able to discriminate visual graphs.

Reading puts heavy demands on visual processing. It is an ecological example of *perceptual learning* (Gibson, 1969; Gilbert & Li, 2012; Goldstone & Byrge, 2015): as result of experience, bottom-up visual processes and top-down knowledge assist fast graph recognition thanks to the emergence of a learning-induced reading-specialized system tuned for the properties of the orthographic code along the *ventral visual* stream, originally dedicated to visual object recognition (Dehaene-Lambertz, Monzalvo, & Dehaene, 2018; Dehaene, Cohen, Morais, & Kolinsky, 2015; Dehaene et al., 2010). The left fusiform gyrus in the *ventral occipitotemporal cortex* (vOT) responds similarly to allographs (i.e., visual forms of the same orthographic unit: e.g., A and a) because it underpins abstract orthographic representations, not determined by physical, phonological, or motor similarity (Dehaene, Cohen, Sigman, & Vinckier, 2005; Grainger, 2018; Rothlein & Rapp, 2014).

Noteworthy, literacy is both reading and writing, and multiple levels of representation are involved. Literacy has “(...) twin beginnings. One is its linguistic origin – it is a special form of speech; and one is its origin in writing – making marks on a piece of paper” (Gibson, 1970, p. 136). Learning to read is a multisensory experience bridging visual graphs with speech sounds and with motor gestures (Pegado, Nakamura, & Hannagan, 2014).

Letters have multiple codes, including sensorimotor and also abstract amodal representations,

¹ The term graph is preferable to letter because it applies to orthographic characters of any written system and not just to alphabets (Chang, Chen, & Perfetti, 2018). In the present work, we also used this term when referring to letter-like symbols as those adopted in studies with artificial scripts even if these symbols are not part of natural written scripts.

and activation spreads bidirectionally between them (Rothlein & Rapp, 2014). It is thus unsurprising that graph and visual word recognition activates a widespread functional neural circuitry comprising not only the vOT but also the left middle and superior temporal gyri, the left supramarginal gyrus, the left inferior frontal gyrus, and the left dorsal premotor and the posterior parietal cortices. There is thus a critical interaction between visual, linguistic, and motor brain systems while reading (James & Gauthier, 2006; Martin, Schurz, Kronbichler, & Richlan, 2015; Murphy, Jorgia, & Talcott, 2019; Nakamura et al., 2012) and while writing (e.g., Longcamp, Richards, Velay, & Berninger, 2016; Nakamura et al., 2002; Planton, Jucla, Roux, & Demonet, 2013).

What is surprising is that training in *handwriting* (HW), that is, the motor action of writing by hand, often leads to better visual graph recognition than other forms of training, regardless of presence or absence of training in phonological correspondences of graphs (i.e., *phonological training*) and of age of learners (for an overview, see, James, 2017; e.g., Bara & Gentaz, 2011; Gentaz, Colé, & Bara, 2003; Guan, Liu, Chan, Ye, & Perfetti, 2011; Guan, Perfetti, & Meng, 2015; James, 2010; Labat et al., 2020; Longcamp et al., 2008; Longcamp, Boucard, Gilhodes, & Velay, 2006; Mayer et al., 2020; Ouellette & Sénéchal, 2008; Seyll, Wyckmans, & Content, 2020; Suggate, Pufke, & Stoeger, 2016; Zemlock, Vinci-Booher, & James, 2018). The advantage from motor training when learning visual graphs was discovered ~50 years ago (e.g., Jeffrey, 1958; Williams, 1969, 1975). Abbott and Berninger (1993) have already suggested that, rather than a motor skill, HW is an orthographic act. Indeed, motor skills account for little variance in HW performance (Santangelo & Graham, 2016). Fine-motor skills differ from graphomotor skills and only the latter show a robust relation with emergent literacy skills in preliterate children (Suggate et al., 2016; Suggate, Pufke, & Stoeger, 2018).

That HW training leads to better writing skills (Abbott & Berninger, 1993; Bara, Gentaz, Colé, & Sprenger-Charolles, 2004; Kiefer et al., 2015; Santangelo & Graham, 2016; for a recent selective review, see, Lyu, Lai, Lin, & Gong, 2021) may seem trivial, but it is impressive that HW training benefits visual graph recognition. The motor act of producing letters by hand, stroke by stroke, establishes a connection between the visual percept of the graph and the motor plan for creating it, resulting in a sensorimotor experience that facilitates graph recognition and, in turn, learning to read. The underlying cognitive mechanism has, however, been elusive. Before strong conclusions can be

drawn, the generality and robustness of these findings must be examined. Therefore, we present the first meta-analysis on the benefit from HW training in visual graph recognition with a twofold aim: to determine the magnitude of this HW advantage, while enlightening the underlying cognitive mechanism(s), by investigating potential moderators of this effect.

One factor that grounds our lack of understanding of these cognitive mechanisms is the observation of discrepant results. Many studies found an advantage from HW training on graph recognition and on the rapid emergence of a reading-specialized brain system (e.g., Cao, Rickles, et al., 2013; Cao, Vu, et al., 2013; James, 2010; James & Atwood, 2009; James & Engelhardt, 2012; Longcamp et al., 2008; Vinci-Booher, James, & James, 2016). However, some studies reported effects that only tended towards significance (i.e., $.05 < p\text{-value} < .10$: Bara et al., 2004; Longcamp, Zerbato-Poudou, & Velay, 2005; Suggate et al., 2016) or significant differences were exclusive of brain activity (e.g., Cao, Rickles, et al., 2013; James, 2010; James & Gauthier, 2009; Kersey & James, 2013; Zhao, Li, Zhao, Gaspar, & Weng, 2015; Zhao, Zhao, Weng, & Li, 2018). Other studies found mixed results (e.g., an advantage from HW training in one particular task of graph recognition but not in others: Guan et al., 2015; Hulme, Monk, & Ives, 1987; Labat et al., 2020; Longcamp et al., 2006; Xu, Chang, Zhang, & Perfetti, 2013; Zemlock, Vinci-Booher, & James, 2018; e.g., positive effects only for the older preliterate children: Labat, Ecalle, & Magnan, 2010; Longcamp et al., 2005) or did not find any advantage from HW relative to control training (e.g., Jensen & King, 1970; Kiefer et al., 2015; Koenigsberg, 1973; Naka & Naoi, 1995; Williams, 1975). Beyond possible arguments of lack of statistical power and/or sensitivity of behavioral tasks, studies also differ on other factors that may potentially affect the benefit from training, including the methodological parameters (e.g., studies with adults vs. with children; training in a single session vs. in multiple sessions for several weeks; training on real letters without control for exposure in-between sessions vs. on artificial scripts). It is thus timely to conduct a synthesis that establishes the size of this effect and whether and how it is modulated by other factors (e.g., age of participants; amount of training: Kiefer et al., 2015; Longcamp et al., 2005), hence, enlightening the nature of these cognitive mechanisms.

Three hypotheses have been proposed: the *sensorimotor* hypothesis (e.g., Longcamp et al., 2005, 2006, 2008; Yin & Zhang, 2020), the *perceptual variability* hypothesis (e.g., James &

Engelhardt, 2012; Li & James, 2016), and the *visual segmental* hypothesis (e.g., Courrieu & de Falco, 1989; Koenigsberg, 1973; Seyll et al., 2020).

The sensorimotor hypothesis and embodied cognition

The sensorimotor hypothesis (e.g., Longcamp et al., 2005, 2006, 2008; Yin & Zhang, 2021) is framed by the proposal of *embodied* cognition (also called grounded or situated cognition: e.g., Allport, 1985; Barsalou, 2008, 2016; Kiefer & Pulvermuller, 2012; for a critical discussion, see, e.g., Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016; Machery, 2016; Mahon, 2015; Mahon & Hickok, 2016), which holds that the content and the format of cognitive representations are isomorphic: sensory concepts have a sensorial format and action concepts have a motor format.

In this vein, cognitive representations of graphs and written words are not symbolic but rather sensorimotor, and hence, the graphic motor programs acquired via HW would be reactivated during visual graph recognition because they would be an inner part of graph representations (Bara et al., 2004; Labat et al., 2020; Labat, Vallet, Magnan, & Ecalte, 2015; Longcamp et al., 2016). Therefore, due to such reactivation, HW would assist visual graph recognition more than other learning experiences where no such detailed motor reproduction of the shape of graphs is trained (Longcamp et al., 2016; Z. Xu, Liu, & Joshi, 2020). It would also contribute to the development of the reading-specialized neural network, establishing functional connectivity between the left vOT and frontal and parietal regions (Longcamp et al., 2016; Vinci-Booher et al., 2016). Indeed, immediately after training, learners show an enhancement in functional activity of the left vOT for visual graphs trained via HW relative to those trained either by typing or mere visual exposure (James, 2010; James & Atwood, 2009; James & Engelhardt, 2012). Nonetheless, this result seems to be transient rather than reflecting long-term changes (Vinci-Booher & James, 2020; Vinci-Booher, James, & James, 2021).

Abstractionist symbolic accounts and perceptual learning

The perceptual variability and the visual analysis hypotheses are framed by *abstractionist*, *symbolic* accounts (e.g., Goldinger et al., 2016; Machery, 2016; Mahon & Hickok, 2016) which do not deny the role of sensorimotor experiences when learning to read but rather consider that abstract representations are not reducible to, and hence differ from, modality-dependent ones. Therefore, orthographic representations are connected to the input (visual) and output (motor) systems and

activation cascades automatically from them; hence, motor activation is subsequent to and contingent upon access to abstract orthographic representations (Rothlein & Rapp, 2014).

HW is a visuomotor skill but is also a perceptual learning skill (Bara & Gentaz, 2011; Kirk, 1980; Merritt, Swain, Vinci-Booher, & James, 2020; Z. Xu et al., 2020). The cognitive mechanism underpinning the HW benefit could thus be perceptual learning because HW training would highlight visual features of graphs: “it might be a function of the amount of attention that is focused on the stimuli rather than the motor activity per se” (Williams, 1975, p. 790).

Both the perceptual variability and the visual segmental hypotheses propose that perceptual learning is the cognitive mechanism underpinning the benefit from HW training. Although not mutually exclusive, these two hypotheses focus on different operations of perceptual learning.

Indeed, two perceptual challenges must be tackled when learning to read (Goldstone & Byrge, 2015; Prettyman, 2019). Learners must become able to group together percepts that were previously distinguishable (e.g., a and A), giving rise to the formation of abstract graph identities (Dehaene et al., 2005; Grainger, 2018) by a process of unitization and category learning. Second, learners become able to isolate diagnostic perceptual features via differentiation, which allows discrimination of graphs that were at first indistinguishable. Contrary to what happens for other visual objects, orientation is a diagnostic feature of graph identity (Fernandes, Leite, & Kolinsky, 2016; Gibson, 1969; Gibson, Pick, Osser, & Gibson, 1962). For example, graphs with the same shape and visual features which are mirror images (e.g., b and d) are originally processed as equivalent by the ventral system due to *mirror invariance* (Bornstein, Gross, & Wolf, 1978; Dehaene et al., 2015; Kolinsky et al., 2011; Logothetis, Pauls, & Poggio, 1995; Pegado et al., 2014). However, such graphs must rapidly be discriminated so that fluent readers can distinguish words as big and dig in a single fixation.

The perceptual variability hypothesis

According to the perceptual variability hypothesis (James, 2017; James & Engelhardt, 2012; Li & James, 2016) the benefit from HW training when learning visual graphs is because learning to generate motor commands is contingent to visual pattern recognition. The spatiotemporal contingency between motor and visual experiences would lead to gains in visual graph recognition because during HW learners are actively engaged in extracting and developing the rules necessary for production of

graphs, and hence, in extracting perceptual invariants (Kirk, 1980). Handwritten instances of graphs are more variable and do not conform to rigid prototypes, which in turn, would affect graph categorization by broadening perceptual categories at the left vOT (James & Engelhardt, 2012; Li & James, 2016). In this vein, after learning letters via copying, tracing, or typing, only preliterate children trained via copying (and hence, exposed to more variable percepts) showed greater activation at the left posterior fusiform region when viewing letters than geometric shapes (James & Engelhardt, 2012). When presented with Latin-alphabet letters, functional activation at the bilateral temporal cortex was also modulated by type of letter (handwritten vs. printed) in beginning readers (Vinci-Booher & James, 2020).

According to this hypothesis, the key component underpinning the advantage from HW training is category learning, regardless of the motor act. Therefore, if training provides experience with variable forms, no additional benefit from HW would be found. Indeed, post-training gains in visual graph categorization were larger for children exposed to variable instances of graphs independent of training in HW or not (Li & James, 2016). In the same vein, exposure to variable forms of letters at school (i.e., cursive and printed: e.g., *a* and a) boosts letter category learning, facilitating reading acquisition. Beginning readers who learned to read with printed letters and to write with cursive letters showed better reading skills than those trained with (less variable) printed letters only (Bara, Morin, Alamargot, & Bosse, 2016).

The visual analysis hypothesis

The visual analysis hypothesis (Courrieu & de Falco, 1989; Koenigsberg, 1973) was the first proposed (for an early review, see, Guralnick, 1972) and has recently seen a revival (Seyll et al., 2020). HW training would enhance feature-based processing, which in turn grounds efficient visual graph recognition (Gibson et al., 1962; Grainger, 2018; Pelli et al., 2006). Indeed, HW requires visual analysis of the component parts of the visual graph, awareness of size, orientation, and spatial relations between graph parts and the whole shape. In natural settings, HW might be the most obvious and spontaneous experience promoting such detailed visual analysis (Seyll et al., 2020).

Based on this hypothesis, the construction of detailed graphic motor programs would strongly

rely on the visual system, especially in the earliest stages of learning. This prediction coheres with the transient enhancement in functional activity at vOT found immediately after training for graphs learned via HW than via control training (James, 2010; James & Engelhardt, 2012; Vinci-Booher & James, 2020). This hypothesis also agrees with behavioral evidence showing no added value of graphomotor action (copying and tracing, proxies of HW) when visual training was fully focused on diagnostic features of graphs, as orientation-differences of mirrored pairs (e.g., d and b: Koenigsberg, 1973; Williams, 1975), or when visual segmental analysis was explicitly trained (Courrieu & de Falco, 1989; Seyll et al., 2020). Such non-motor training conditions led to robust post-training gains in visual graph recognition and no significant extra gain was found from graphomotor (HW) training.

A synthesis of the available primary studies can only consider the parameters that were already examined. Therefore, we next define the four types of moderators investigated here and how their analysis allowed shedding some light upon the underlying cognitive mechanisms.

Training program

It is now established that the specific training experience when learning visual graphs influences the processing strategies and neural substrates recruited in subsequent visual graph recognition (James, 2010; James & Engelhardt, 2012; Kersey & James, 2013; Longcamp et al., 2008; Vinci-Booher et al., 2016, 2021). We thus considered three types of moderators of training program: type of control training, presence/absence of phonological training, and the HW tasks used.

Studies on the effect of HW training have usually adopted either a *visual control* training, where no motor action is involved and visual graphs are presented either with or without phonological correspondences (Bara et al., 2004; e.g., look and say: Li & James, 2016; or just see: Naka & Naoi, 1995), or a *motor control* training, where a motor action is required but without a graphomotor component, such as in typewriting (Guan & Wang, 2017; Longcamp et al., 2005, 2006; Mayer et al., 2020), pointing (e.g., Suggate et al., 2016), circling (Naka & Naoi, 1995), or drawing pictures (Ouellette & Sénéchal, 2008). The embodied and symbolic accounts differ on their predictions about the size of the HW advantage relative to these two types of control training.

According to the embodied account, sensorimotor learning experiences emerge from brain-body-environment interactions, which become part of graph representations (Bara, Gentaz, & Colé,

2007; Labat et al., 2020; Longcamp et al., 2016; Z. Xu et al., 2020). Therefore, the advantage from HW training would be smaller relative to motor control than to visual control training given that in the latter no such sensorimotor integration is experienced. In both HW and motor control training, kinesthetic information would be coupled with the visual form of graphs via hand movements associated with those visual images, but only HW creates a graphomotor representation of graphs.

Alternatively, if the underlying mechanism regards perceptual learning (e.g., Li & James, 2016; Seyll et al., 2020), then the more the control learning experience diverts attention from the visual graphs, the larger the advantage from HW training. In other words, if the mechanism underpinning the HW benefit is about visual segmental analysis (e.g., Courrieu & de Falco, 1989; Seyll et al., 2020), then the advantage from HW training would be larger relative to motor than to visual control training because in the latter attention is fully focused on graphs (albeit without explicit visual segmental analysis). Note that in motor control training, the learner needs not only to attend to the graph but also to focus on the motor action towards it, which would divert attention from graphs.

In the same vein, the two accounts diverge on the moderator role of *phonological training*.

The sensorimotor hypothesis predicts that learning effects would be amplified by rich sensorimotor experiences. Thus, HW training combined with phonological training would boost motor and audiovisual integration and the emergence of graph representations (relative to control training) more than when HW training occurs solely (e.g., Kiefer et al., 2015; Labat et al., 2015).

In contrast, according to perceptual learning accounts, graph knowledge includes knowledge of the visual form of graphs and of their phonological correspondences because phonological recoding is the essence of reading (Araujo, Faisca, Bramaio, Reis, & Petersson, 2015; Brem et al., 2010; Brem et al., 2018; Share, 1995; Ziegler & Goswami, 2006). Therefore, unitization and differentiation of graphs would be assisted by phonological training, resulting in a smaller benefit from HW training (relative to the control) when phonological training is present than absent.

Note that, regarding the perceptual learning account, the two predictions outlined are mainly based on the visual analysis hypothesis (Courrieu & de Falco, 1989; Seyll et al., 2020; Williams, 1975). According to the perceptual variability hypothesis (James & Engelhardt, 2012; Kersey & James, 2013; Li & James, 2016) the key aspect of HW training regards exposure to variable exemplars

(e.g., Merritt et al., 2020). Therefore, differences in the magnitude of the HW advantage would also be related to the *HW tasks* adopted (James & Engelhardt, 2012).

Three types of HW tasks have been used: *copying* a graph from a model (Longcamp et al., 2006; Longcamp et al., 2005; Williams, 1969), *tracing* over dashes or dots comprising the outline of the graph (e.g., Bara et al., 2004; Hulme et al., 1987; James & Engelhardt, 2012; Jensen & King, 1970; Labat et al., 2015), or *writing from memory*, that is, delayed copying after presentation of the graph (e.g., Chang, Xu, Perfetti, Zhang, & Chen, 2014; Guan & Wang, 2017; Mayer et al., 2020; Ouellette & Sénéchal, 2008; Y. Xu, Chang, Zhang, & Perfetti, 2013). According to the perceptual variability hypothesis, only after copying (but not tracing) the visuomotor brain network gets recruited in subsequent letter perception (James & Engelhardt, 2012), even if such recruitment is transient (Vinci-Booher et al., 2021). The more variable the perceptual output produced by HW, the larger the benefit to be found in subsequent visual graph recognition.

Set size and training regime

To our knowledge, the potential role of *set size* (the number of symbols comprising the script to be learned) and *training regime*, that is the *total amount of training* and the *duration* and *frequency* of each training session, have never been tested in HW studies.

Natural writing systems vary in set size (e.g., Daniels & Share, 2018), which accounts for unique variance in visual recognition of graphs, beyond visual complexity (Chang et al., 2018; Pelli et al., 2006). The studies synthesized in the present work ranged from set sizes that were small (from three graphs trained in Suggate et al., 2016; Williams, 1969) to large (of 30 graphs: Guan et al., 2011, Experiment 2), and hence, we explored the potential role of this moderator. Due to the underestimation of training regime, these studies might present a trade-off between set size and amount of training.

Laboratory studies might not produce the effects found along real-life literacy instruction. Yet, hallmarks of reading expertise do not require years to emerge (Brem et al., 2010; Dehaene-Lambertz et al., 2018; Hervais-Adelman et al., 2019). Regarding the HW benefit, studies with a single session (from 1-3 min in Naka & Naoi, 1995; Suggate et al., 2016, to 30 min in Guan et al., 2011, Experiment 1) and multiple sessions (from two sessions of 14 min each: James & Atwood, 2009; to 28 sessions of 25 min each: Mayer et al., 2020) have both found larger gains in visual graph recognition. Although

acknowledged (Jeffrey, 1958; Longcamp et al., 2005), amount of training has barely been considered.

Related parameters are duration and frequency of training, that is, to what extent learning depends on how training is parsed across sessions. In a recent study on motor learning (where visual recognition was not tested), Ghanamah, Eghbaria-Ghanamah, Karni, and Adi-Japha (2020) showed that practice does not always makes perfect. Increasing training duration within a single session had immediate but transient benefits, quickly lost after training. Spacing practice in several sessions provided a better opportunity for encoding and integration into long-term memory, which was visible in follow-up assessment. Such distinction between transient and permanent learning effects is important. In a recent study, Vinci-Booher et al. (2021) showed that the immediate post-training changes in visual-motor brain functional connectivity induced by HW training are already gone one week after training, revealing the transitory nature of this effect. Note that, to establish that HW training is effective when learning visual graphs, and hence, education-relevant, its effects must endure. Although emphasized by Longcamp et al. (2005, 2008); Vinci-Booher and James (2020), follow-up assessment was rarely conducted and with a disperse interval: from one week (e.g., Longcamp et al., 2005), three weeks (e.g., Seyll et al., 2020, Experiment 1), four months (Labat et al., 2015), to one year (Bara et al., 2007). Therefore, although surely relevant, we could not investigate long-term post-training effects in the present meta-analysis.

Granularity of visual discrimination and perceptual learning tasks

The impressive aspect of the benefit from HW training regards the fact that it corresponds to a *transfer* effect (Gilbert & Li, 2012; Prettyman, 2019), where performance in an untrained non-motor visual graph recognition task is enhanced afterwards. Such result indicates how ready the graph knowledge acquired during specific training can be applied across different perceptual tasks tapping into visual recognition of graphs. Transfer effects depend on the degree by which learning and testing tasks have elements in common (Gilbert & Li, 2012) and both the sensorimotor and the symbolic accounts predict the benefit from HW training to be modulated by *granularity* of discrimination.

Graphs can differ along two perceptual dimensions: shape and orientation. They can differ both on features and whole shape (e.g., M and S), where *coarse-grained* discrimination is enough, or they can differ only by orientation, sharing visual features and even the whole shape (e.g., N and Z; b

and d). In this latter case, graphs are highly confusable and *fine-grained discrimination* is required, as happens for mirror images due to mirror invariance (Bornstein et al., 1978; Dehaene et al., 2015; Fernandes et al., 2016). Both theoretical accounts predict that the advantage from HW training would be larger for fine-grained than coarse-grained discrimination, either because discrimination of highly-confusable graphs depends on multi-system interplay (Pegado et al., 2014), assisted by automatic (re)activation of distinct graphic motor programs (Longcamp et al., 2016), or because HW training forces attention to the distinctive features of highly-confusable pairs (Samuels, 1973; Tawney, 1972).

Several tasks have been adopted to examine post-training gains in visual graph recognition and reading and discrepant results have been reported (e.g., larger gains in visual recognition of graphs trained via HW than typing in old/new judgment but not in orientation-based same-different matching tasks: Longcamp et al., 2006). However, no primary study has directly investigated the potential role of perceptual learning tasks, which we also explored here.

Age of participants

The earliest studies on the role of motor action in visual graph recognition were conducted with preliterate children (e.g., Jeffrey, 1958; Jensen & King, 1970; Williams, 1969, 1975) but since 1995 studies with adults have flourished (Naka & Naoi, 1995). Studies with children and with adults have adopted artificial scripts to control for exposure outside training (e.g., Seyll et al., 2020; Williams, 1969) or natural scripts (e.g., Bhide, 2018; Cao, Rickles, et al., 2013; Y. Xu et al., 2013; Z. Xu et al., 2020), and a positive effect of HW has often been reported. However, this is not always the case, and some authors argued that mixed results could be due to age effects. For example, Kiefer et al. (2015) found no advantage from HW training in letter recognition by 5-year-olds, and Labat et al. (2010) only found an advantage from HW training in 5-year-olds but not in 3-year-olds, similarly to what was found by Longcamp et al. (2005). We thus also explored the potential role of age.

This meta-analysis traces the profile of the advantage from HW training when learning to read, hopefully highlighting the nature of the underlying cognitive mechanisms, which could ultimately reveal the limits of experience-dependent plasticity. The present results can also translate into pedagogical practices by informing the best strategies for teaching literacy, taking into account antagonist forces: HW is at first a laborious graphomotor activity (Santangelo & Graham, 2016) and

pen(cil)s have been replaced by technological-based tools (e.g., keyboards and touchscreens), leading Heuer (2016) to announce the “death of handwriting” (p. 123).

Method

Literature search and inclusion criteria

The literature search covered the title, abstract, and keywords of empirical (experimental and quasi-experimental) studies published in English, French, or Spanish, in four databases (up until May 2021): Web of Science, PubMed, PsycINFO, and ERIC. We used combinations of keywords related to training (*training* OR *longitudinal* OR *learning* OR *intervention*) and type of training (*handwriting* OR *writing action* OR *motor action* OR *sensorimotor* OR *visuomotor* OR *graphomotor*) crossed with terms about graph identification (*letter* OR *reading*). This search returned 3024 references. Review articles, dissertations, or studies with participants with sensory and/or cognitive impairments were not included. The reference lists of eligible articles were also hand-searched, returning 50 additional studies (see Figure 1).

Studies had to meet five inclusion criteria: (1) to include a group/condition with HW training; (2) to include a control training group/condition on the same graphs but without HW; (3) to include at least one measure of visual graph recognition, and hence, a post-training perceptual-learning outcome expressed in accuracy and/or response times (RT); (4) to report enough data to compute effect sizes; and (5) to test healthy, typically-developing participants, novice on the graphs to-be-learned.

For preventing the violation of the independence of observations, we inspected studies from the same authors, checking for possible duplicate samples (e.g., Zhao et al., 2015, 2018). Studies with between- or within- participants design were included as well as studies with pre-/post- or with only post- training assessment, given that effect sizes do not depend on these methodological aspects (Borenstein, Hedges, Higgins, & Rothstein, 2009). Note that for studies with multiple testing sessions along training (e.g., Cao et al., 2013; Vinci-Booher et al., 2021), only the last testing in graph recognition, after the end of training, was considered. In Guan et al. (2017), given that all children received HW training in Session 2, we considered the post-training scores after the first training session, on which one group received HW training and the other received motor control training.

Seven studies with beginning readers were also included because they were trained on unfamiliar graphs: two studies with school children (Guan & Wang, 2017; Kersey & James, 2013) and five with adults (Chang et al., 2014; Guan et al., 2011, 2015; Y. Xu et al., 2013). In Guan and Wang (2017), Chinese children were trained on unfamiliar hanzis (their pre-training accuracy on these graphs was < 15%). In Kersey and James (2013), American children were trained on cursive letters (pre-training knowledge was ~ 56%). The studies with adults were all with foreigners learning Chinese as second language who knew much less hanzis than fluent readers (~3000 characters): < 180 hanzis (Chang et al., 2014; Guan et al., 2011; Guan et al., 2015) and ~ 450 hanzis (Y. Xu et al., 2013).

As shown in Figure 1, 33 articles comprising 50 independent studies with 1525 participants met all inclusion criteria, and hence entered, this meta-analysis.

-----Figure 1 about here -----

Coding procedure and moderator variables

For each study, we extracted sample size, mean, and standard deviation of each group/condition on visual graph recognition at pre- (whenever tested) and at post- training.

Coding was performed by two independent raters (second and last authors) on 85% of the studies. Inter-rater reliability was estimated with Pearson correlation for continuous variables (such as for effect size) and Cohen's kappa for categorical variables (such as for type of control training). Any between-raters disagreement was resolved by consulting the article and by discussion (e.g., checking for transcription errors, revising the coding criteria for classification of moderators). Inter-rater correlation coefficients ranged from .94 to 1, and kappa coefficients ranged from .93 to 1.

Training program

Type of control training. This moderator was coded into two categories: *visual*, when participants were presented with the graphs and were required to pay attention to them without any motor task; *motor*, when participants performed motor actions but not graphomotor ones (e.g., typing, pointing, circling).

Phonological training. Studies were coded for *presence* or *absence* of phonological training, i.e., *Ph+* or *Ph-*, depending on whether phonological correspondence of graphs were also trained.

HW task. The HW training group/condition was coded into three categories, depending on

task: copying, tracing, and writing from memory. Whenever training included multiple HW tasks (e.g., copying and tracing: e.g., Williams, 1969; Zhao et al., 2018; e.g., copying, tracing, and writing from memory: Mayer et al., 2020), these studies were excluded from the analysis of this moderator ($k = 5$).

Set size. The number of graphs comprising the script (integral numbers).

Training regime. Three parameters were considered: amount of training (i.e., total duration of the program in minutes); session duration (in minutes) and frequency (i.e., number of sessions per week, excluding studies with a single training session).

Perceptual learning tasks

The perceptual learning tasks were coded into three categories: *categorization* (i.e., symbol/non-symbol judgment), which can be accomplished based solely on graph familiarity (e.g., Guan et al., 2011; Guan et al., 2015; Y. Xu et al., 2013), *visual discrimination*, which requires discrimination between visual graphs differing in shape and/or orientation (e.g., Li & James, 2016; Longcamp et al., 2008; Zemlock et al., 2018), and *phonological mapping*, when phonological processing was also involved, either of access to (e.g., audio-visual matching, where the name of the graph is presented and participants need to choose the corresponding visual graph: Bara, Fredembach, & Gentaz, 2010; Li & James, 2016) or retrieval of (e.g., graph naming: Hulme et al., 1987; Kiefer et al., 2015; graph-string reading: Mayer et al., 2020; Ouellette & Sénéchal, 2008) phonological representations.

Note that, in some studies, a letter categorization task involved visual discrimination, and hence, was coded accordingly. For example, Zemlock et al. (2018) used a letter sorting task, where children were asked to sort all handwritten instances of one letter in the same pile, with different piles for different letters, and hence, it required letter identification and visual discrimination. Furthermore, phonological mapping could only be considered in Ph+ studies but not all of these studies assessed phonological mapping (e.g., Hulme et al., 1987, Experiment 4; Naka & Naoi, 1995, Experiment 4).

Granularity of visual discrimination. The granularity of visual discrimination was coded into two categories (only for studies for which granularity could be deduced): *coarse-grained*, when graph discrimination depended only on visual shape (e.g., d and v); *fine-grained*, when orientation processing was also necessary besides shape (e.g., d and b). Note that some phonological mapping tasks involved visual discrimination and granularity could also be considered (e.g., four-alternative

choice of the graph auditorily presented, where visual distractors included a mirror image of the target: Li & James, 2016); therefore, such studies were also coded for this moderator.

Age of participants

Age was coded into two categories: *children* vs. *adults*. It was not treated as a continuous variable because participants were either preliterate children or adults.

Meta-analytic procedures

Effect size (ES) estimates

Data were analysed with *Comprehensive Meta-Analysis* (v.3.3.070; Borenstein, Hedges, Higgins, & Rothstein, 2014). First, for each comparison between HW vs. control training, *Cohen's d* (Cohen, 1988) was calculated as the standardized difference between these two conditions/groups in the perceptual learning outcomes (measured at pre-/post-training or at post-training only), divided by their pooled standard deviation. A positive *d* indicates that HW training led to a higher perceptual learning outcome, that is, to better visual graph recognition than control training. In line with Cohen (1988), ES was interpreted as small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$). When mean and standard deviation were not reported, if possible, Cohen's *d* was estimated from the reported *F* statistics or data were extracted with WebPlotDigitizer (v.3.9.; Rohatgi, 2015). Next, Cohen's *d* was used to calculate *Hedge's g* (Hedges, 1982) to correct for possible bias due to small samples. *Hedge's g* is very close to Cohen's *d* when the number of participants in each study is ≥ 10 (Borenstein et al., 2009). We also checked that both computations led to equivalent statistical results.

When studies reported more than one ES for the same sample (e.g., perceptual learning outcomes in more than one task), we adopted the *shifting unit of analysis* approach (Cooper, 2010), as it provides a good compromise between preserving the independence of effect sizes (necessary for the validity of meta-analyses) and retaining the maximum amount of information from each study. First, each ES associated with one sample was coded as an independent estimate. Next, the unit of analysis was shifted according to the moderator at test. In the overall mean ES analysis (summary ES) we used the sample as unit of analysis: the multiple effect sizes were aggregated into one weighted average ES so that each sample contributed only to one ES, and hence, sample-averaged effects were independent, while avoiding over-representation of multi-experiment studies (Rosenthal, 1991). In moderator

analyses (subgroups and meta-regressions) the same hold, except when testing a specific categorical moderator for which there were multiple effect sizes within-study (e.g., perceptual learning outcomes; granularity of visual discrimination). In this case, we shifted the unit of analysis from the sample to the ES, so that each sample contributed to one ES in each category of the moderator. To illustrate this procedure take the example of Kiefer et al. (2015): this study assessed post-training gains in visual graph recognition with two types of tasks. Therefore, Kiefer et al. contributed with one ES (combined ES) in the overall analysis and with two ES in the moderator analysis of perceptual learning tasks (one for visual discrimination, another for phonological mapping).

Analyses of effect sizes

The overall ES was estimated by calculating a weighted average of individual effect sizes (Rosenthal, 1991). A random-effects model was adopted in each analysis, given that heterogeneity among participant samples and experimental conditions could be presumed, and hence the true effect would vary from one study to another. A 95% *confidence interval* (CI) was calculated for each ES, to examine whether it was significantly larger than zero. To examine the variation in ES between studies, the Q test of homogeneity was used (Q_{within}). The descriptive statistic, I^2 , was also used to determine the degree of heterogeneity, which quantifies the proportion of the observed variance among ES due to real differences among studies rather than sampling error (Borenstein et al., 2009). By convention, I^2 values of 25%, 50%, and 75% indicate low, moderate, and high values of heterogeneity, respectively (Higgins et al., 2019). Therefore, when I^2 is moderate-to-large, the magnitude of effect sizes is variable across studies, and hence, there is enough heterogeneity to investigate whether these differences are meaningful and depend on moderator variables. Moderator analyses (subgroup analyses and meta-regressions) were conducted using mixed-effects models to investigate if the a-priori defined moderators influenced the magnitude of the benefit from HW training on visual graph recognition. For categorical moderators (e.g., perceptual learning tasks), we ran subgroup analyses with mixed-effects models and between-groups heterogeneity statistics, $Q_{between}$ (analogous to analysis-of-variance F test, which follows a chi-square distribution). Planned contrasts directly comparing two subgroups were conducted with one-tailed Z -test (a variant of the t -test), given that we had a-priori, theoretically driven predictions about the direction of these effects. For continuous moderators (e.g.,

set size), we ran meta-regressions based on the method of moments for random-effects models and the percentage of between-study variance explained, R^2 , is reported as a measure of ES. Note that moderator analyses typically need a large number of observations for high statistical power (Hedges & Pigott, 2004) which, in turn, is reduced by high levels of heterogeneity. Consequently, caution in warrant when interpreting null effects.

Outlier detection

Forest plots were used to examine the distribution of effect sizes and to detect potential outliers. Sensitivity analyses were conducted to determine the impact of outliers on the overall range of means, given that these analyses allow estimating an adjusted overall ES after removing studies one by one. They thus allow reporting the range of overall ES (i.e., highest and lowest ES) when outliers at each end of the distribution have been removed.

Publication bias

In a meta-analysis, the observed overall ES may be an overestimation of the true ES due to a biased sample of primary studies. We thus checked for a potential publication bias: first, with a *funnel plot*, which must be symmetric in the absence of a publication bias because the studies would be distributed symmetrically around the mean ES; next, with a *trim-and-fill* method for random-effects models to compare the distribution of observed ES to the predicted distribution (around the true mean; Duval & Tweedie, 2000). This method imputes values in the funnel plot to render it symmetrical and, on this basis, calculates an estimated overall ES. Finally, we also ran a *file-drawer* analysis to compute a *fail-safe N*, i.e., the number of missing studies necessary to nullify a statistically significant effect (Borenstein et al., 2009; Rosenberg, 2005; Rosenthal, 1991).

Results

Descriptive overview of the studies

The 50 independent studies (published in 33 articles) that entered this meta-analysis (corresponding to 1525 participants) reported accuracy data in perceptual learning tasks (only six

studies also reported RT; for a complete description, please see Table A1, Online Resource 1²). Most studies examined small samples (i.e., < 30), from 6 to 51 learners per training group. Twenty-seven studies adopted a between-participants design (54%; 16 with pre-/post-training assessment; 11 with post-training assessment only) and 23 adopted a within-participants design (46%; 20 with post-training assessment only). Post-training assessment occurred immediately after the last training session in most studies (36 studies: 72%) and 1-14 days post-training in 12 studies.

----- Table 1 about here -----

As shown in Table 1, the proportion of studies with visual (48%) and motor (52%) control training was similar, which in turn did not differ in training regime parameters (session duration, amount, and frequency of training) or set size (Table 2), *Mann-Whitney U-test*³, $U_s < 239.00$, $ps > .16$, $rs < 0.21$. Therefore, the moderator analysis on type of control training had no potential confounds.

Not surprisingly, 64% of the studies included phonological training (82% of the studies with children), given that the core mechanism of reading is phonological recoding (e.g., Share, 1995; Ziegler & Goswami, 2006). The distribution of Ph+ and Ph- studies was not fully balanced, hence, limiting a moderator analysis on this factor but these studies were matched in training regime (session duration, $U = 212.00$, $p = .21$, $r = -0.18$, amount of training, $U = 183.00$, $p = .06$, $r = 0.27$, or frequency, $U = 191.00$, $p = .09$, $r = 0.24$), and set size, $U = 264.00$, $p = .63$, $r = 0.07$ (see Table 2). Notably, in studies with adults, Ph+ and Ph- training was balanced (see Table 1). Thus, any significant moderator effect of phonological training would not be tainted by other factors in this sub-analysis.

Studies varied in HW tasks depending on age. Most studies with children adopted tracing (in 50%: 14 out of 28 studies), copying (25%), or both (11%), whereas studies with adults adopted either copying (68%: 15 out of 22) or writing from memory (32%). Therefore, the distribution of HW tasks was unbalanced and was also confounded with age, thus limiting the analysis of this moderator.

----- Table 2 about here -----

Participants were, on average, trained on 11 new graphs (SD = 7.4) for a total amount of 84

² All databases are available at <https://osf.io/8sqfa> (https://osf.io/wzdxn/?view_only=f31099e685454157b700955c4486a085).

³ Due to violation of normality assumptions, non-parametric, *Mann-Whitney U-test* was used; r ($r = Z / \sqrt{N}$) is reported as estimate of ES, following Cohen (1988): i.e., small, moderate, and large ES correspond to 0.10, 0.30, and 0.50, respectively.

minutes ($SD = 125.1$) distributed over 1.7 ($SD = 1.0$) sessions per week with an average duration of 16.5 minutes ($SD = 9.7$). Set size was significantly correlated with session duration, $r(45) = .31, p = .034$, and also tended to be associated with frequency of training, $r(43) = .27, p = .07$, but not with total amount of training, $r(45) = .12, p = .42$. Overall, the more graphs participants were taught, the longer each session and the more systematic the training program tended to be. Session duration was also significantly correlated with amount of training (for multiple-session studies), $r(28) = .47, p = .009$. Frequency of sessions per week and amount of training were significantly correlated in studies with adults and with children, both $r_s > .66, p < .001$ (this also held true when only multiple-session studies were considered). Therefore, longer programs (larger amount of training) were significantly associated with more systematic training and with longer sessions.

Eleven studies assessed the perceptual learning outcomes in visual graph recognition with more than one type of task. Nine studies examined categorization (all Ph+ and two were with children). Twenty-two studies assessed phonological mapping (only three studies with adults) and 27 studies assessed visual discrimination (12 with children of which seven were Ph+; 15 with adults of which two were Ph+).

Granularity of discrimination could be determined in 31 studies of which 19 studies assessed coarse-grained and 17 studies assessed fine-grained discrimination (within some studies, the two granularities were assessed with different tasks: e.g., Li & James, 2016; Zemlock et al., 2018). These studies were relatively balanced in age (10 studies with children in fine-grained and 11 in coarse-grained) but less in phonological training (Ph+ in 12 studies with fine-grained vs. in six with coarse-grained assessment) and in type of control training (visual, fine- vs. coarse-grained: 7 vs. 10; motor: 13 vs. 6, respectively).

A similar proportion of studies were conducted with children (56%; on average, 5.59 years old, $SD = 1.19$) and with adults (44%; on average, 22.44 years old, $SD = 2.60$), which differed on training regime and set size. Indeed, as shown in Table 2, children were trained on significantly less graphs, $U = 159.00, p = .003, r = -0.41$, but for significantly longer, $U = 148.50, p = .003, r = 0.42$, than adults. In single-session studies (18 studies: eight with children, 10 with adults), children were still trained on less graphs than adults, $U = 12.00, p = .014, r = -0.57$, but for a similar duration, $U =$

20.00, $p = .16$, $r = -.16$. In fact, session duration and frequency of training was similar for children and adults in multiple-session studies ($p = .19$, $r = 0.19$ and $p = .85$, $r = .03$). In sum, the moderator analysis of age would be tainted by amount of training and set size.

Overall ES: what is the magnitude of the benefit from HW training in visual graph recognition?

As shown in Table 3, all 50 independent effect sizes entered the computation of the summary ES of the post-training gain of HW relative to control training in visual graph recognition, given that the HW benefit was independent from experimental design: within- vs. between-participants, $Q(1) < 1$, $p = .95$; pre/post- vs. only-post- training assessment, $Q(1) = 1.08$, $p = .29$. The overall weighted mean ES was moderate-to-large and significant, $g = 0.58$, 95% CI [0.39, 0.76], $p < .001$. A sensitivity analysis showed that after removing potential outliers, the overall ES ranged from $g = 0.52$, 95% CI [0.35, 0.68], to $g = 0.60$, 95% CI [0.42, 0.78]. Evidence is thus clear that HW training leads to a significantly larger gain on visual graph recognition than control training.

----- Table 3 about here -----

As illustrated in the funnel plot in Figure 2, most studies were distributed symmetrically around the observed overall ES. This visual impression was confirmed by a trim-and-fill analysis, given that no study was imputed to the left of the mean. More important, the adjusted (predicted) ES was 0.58, 95% CI [0.39, 0.76], exactly the same as the observed ES. There was no evidence that studies with statistically non-significant results or with inconsistent results were underestimated. Rosenthal's *fail-safe N* indicated that the observed ES was robust: we would need to add 1880 studies to the analysis to nullify the observed advantage from HW training on visual graph recognition. In fact, if anything, the publication bias was one of underestimation of the true ES, as a trim-and-fill analysis revealed that 11 studies with larger ES (located to the right of the mean ES) were missing, after whose imputation the adjusted g was 0.78, 95% CI [0.59, 0.98].

----- Figure 2 about here -----

As shown in Table 3, there was substantial variance among the individual effect sizes, which cannot be accounted by the sampling error alone, $Q(49) = 245.74$, $p < .001$, $I^2 = 80.06$, $T^2 = 0.33$ (SE = 0.10). This heterogeneity suggests that there may be meaningful differences among studies, which were further explored through moderator analyses.

Moderator analyses: how does HW training benefit visual graph recognition?

Figure 3 and Table 4 summarize the results of the meta-analyses and meta-regressions.

----- Figure 3 about here -----

Training program

In line with symbolic accounts (Courrieu & de Falco, 1989; James, 2017; Li & James, 2016; Seyll et al., 2020), the magnitude of the benefit from HW training was moderated by the type of control training adopted, $Q(1) = 5.35$, $p = .02$, because the advantage from HW training was significantly smaller relative to visual control training ($g = 0.36$, $SE = 0.16$) than to motor control ($g = 0.87$, $SE = 0.15$), albeit both ES were significant, both $ps < .025$.

To ensure that the significant influence of control training was not contaminated by a third factor (e.g., phonological training), we re-analysed the data, considering studies with adults only, for which there was no bias for Ph+ or Ph-. Again, HW training led to a significantly smaller advantage relative to visual ($k = 8$; $g = 0.17$, $SE = 0.27$) than to motor control training ($k = 14$; $g = 0.90$, $SE = 0.22$), $Z = 2.02$, $p = .02$.

In the same vein, when phonological training was present (Ph+, $g = 0.47$, $SE = 0.11$), the benefit from HW training was significantly smaller than when it did not (Ph-, $g = 0.79$, $SE = 0.16$), $Z = -1.62$, $p = .05$. Given that this comparison was unbalanced (see Table 2), we ensured that restricting the analysis to studies with adults produced the same statistically significant difference (Ph+: $k = 9$; $g = 0.18$, $SE = 0.25$; Ph-: $k = 13$; $g = 0.95$, $SE = 0.22$), $Z = -2.31$, $p = .01$.

Regardless of the HW task, HW training led to an overall significantly larger improvement in visual graph recognition than control training, $Z = 5.76$, $p < .001$. As shown in Figure 3, whereas the benefit from HW training was robust for both copying, $g = 0.67$ ($SE = 0.15$) and tracing, $g = 0.75$ ($SE = 0.19$), both $ps < .001$, the effect was not reliable for writing from memory, $g = 0.27$ ($SE = .23$), $p = .24$, due to a large dispersion, possibly because a small number of studies adopted this task ($k = 9$). We found no significant difference between studies using different HW tasks, $Q(2) = 2.93$, $p = .23$, nor a significant advantage when copying was directly compared with tracing, $Z = -1.60$, $p = .11$ (bilateral test; if anything there was a tendency for studies with tracing to show a larger ES than those with copying). However, these analyses were tainted by age of participants and the distribution of studies

was unbalanced (see Table 2), and hence, may have been too underpowered to detect effects.

Specifically, with a current average sample size of $n = 21$ per study and $\alpha = .05$, a two-sample *t-test* has a power of merely .33 to detect a minimum significant difference (of 0.20 based on Cohen's convention: Griffin, 2020; Hedges & Pigott, 2004).

To avoid potential age confounds, meta-regressions on the set size and training regime were run separately for each age group. Set size was not a significant predictor: for children, $k = 28$, $\beta = -.015$ (SE = 0.02); for adults, $k = 22$, $\beta = -.019$ (SE = 0.02), both $Q(1) < 1$, $R^2 \cong .00$.

----- Table 4 about here -----

The three parameters of training regime were correlated with each other, and hence, were considered together in the meta-regressions conducted separately for children and for adults. Note that multicollinearity does not impede conducting such regression analyses nor questions the robustness of these models given that it has no impact on the overall regression model or associated statistics (R^2 and p -value). However, multicollinearity does distort the regression weights of each individual predictor, and hence, we refrained from interpretation of each predictor separately. As shown in Table 4, we conducted two meta-regressions: one including all studies, another with only multiple-session studies. Note that, for single-session studies, duration of session and amount of training were the same, and hence, in this case the two predictors considered were total amount of training and frequency (sessions per week). Overall, it was only in studies with children that training regime significantly moderated the size of the benefit from HW training in visual graph recognition: it explained ~13% of the variance in the ES observed, $Q(2) = 10.38$, $p < .001$, $R^2 = .13$, $I^2 = 62.75\%$, $T^2 = 0.16$. In studies with adults, training regime was not significant, $Q(2) < 1$, $R^2 = .00$, $I^2 = 89.33\%$, $T^2 = 0.82$.

In meta-regressions on multiple-sessions studies only, the three parameters were jointly considered and, interestingly, training regime was a reliable moderator of the HW benefit for children, $Q(3) = 12.22$, $p = .007$, $R^2 = .38$, $I^2 = 68.14\%$, $T^2 = 0.22$, and also for adults, $Q(3) = 10.05$, $p = .018$, $R^2 = .10$, $I^2 = 83.54\%$, $T^2 = 0.52$.

Perceptual learning tasks

The benefit from HW training was significant in all perceptual learning tasks considered:

visual discrimination ($k = 27$), $g = 0.70$ ($SE = 0.15$), $p < .001$; phonological mapping ($k = 24$), $g = 0.38$ ($SE = 0.16$), $p = .017$; categorization ($k = 9$), $g = 0.62$ ($SE = 0.26$), $p = .017$. The moderator effect of perceptual learning task was not significant, $Q(2) = 2.10$, $p = .50$. Note, however, that this moderator analysis was tainted by confounds. Across tasks the benefit from HW training in visual recognition of graphs was reliable, $Z = 5.54$, $p < .001$.

More important, the benefit from HW training was moderated by granularity of visual discrimination, $Q(1) = 8.91$, $p = .003$. The overall ES was reliable across grain-size, $Z = 4.27$, $p < .001$, but it was much larger when visual discrimination tapped into fine-grained ($k = 17$), $g = 0.93$ ($SE = 0.18$), $p < .001$, than into coarse-grained features ($k = 19$, $g = 0.19$, $SE = 0.17$), $p = .26$ (see Figure 3). Thus, the benefit from HW training was especially strong when visual graph recognition implied discrimination of distinctive, diagnostic features such as orientation of highly confusable graphs.

Age of participants

There was no significant moderator effect of age, $Q(1) < 1$: both studies with children ($g = 0.57$, $SE = 0.13$) and adults ($g = 0.60$, $SE = 0.14$) showed larger gains in visual graph recognition from HW training, $ps < .001$. However, this analysis had low power (.21), and children were overall trained for longer and on less symbols than adults (see Table 2), which might have disguised a potential moderator effect of age.

Discussion

In the last decade, cognitive science has shown a renewed interest in the role of HW training when learning to read (for an overview, see, James, 2017). Indeed, the type of learning experience influences the processing strategies and neural substrates recruited in subsequent visual graph recognition (James, 2010; James & Atwood, 2009; James & Engelhardt, 2012; Kersey & James, 2013; Longcamp et al., 2008; Vinci-Booher et al., 2016, 2021). However, the nature of the underlying cognitive mechanism has been elusive. In the present work, we reported the first meta-analysis of the benefit from HW (relative to other types of training) on subsequent visual recognition of graphs. We have thus statistically synthesized the available research and provided two original contributions.

First, the present results indicated a moderate-to-large and statistically robust overall effect of

HW training in visual graph recognition (Hedge's $g = 0.58$). Although the benefit from graphomotor training when learning visual graphs was discovered fifty years ago (e.g., Jeffrey, 1958; Williams, 1969), and most research has since been consistent, discrepant results have also been reported (e.g., Kiefer et al., 2015; Koenigsberg, 1973; Naka & Naoi, 1995). Thus, it was timely to determine the size of this effect.

Second, the overall analysis showed a substantial amount of heterogeneity between-studies, suggesting that the adoption of the diverse methodological parameters might likely explain some of the observed discrepant results. More important, such heterogeneity allowed testing the predictions about the operation of the underlying cognitive mechanism formulated by two proposals: the sensorimotor hypothesis framed by embodied cognition (Bara et al., 2004; Labat et al., 2020; Longcamp et al., 2008; Longcamp et al., 2006; Longcamp et al., 2005; Yin & Zhang, 2021) and the symbolic accounts of the visual analysis and the perceptual variability hypotheses (Courrieu & de Falco, 1989; James & Engelhardt, 2012; Li & James, 2016; Seyll et al., 2020; Vinci-Booher & James, 2020; Williams, 1975). To this aim, we examined four types of moderators: training program (type of control training; presence/absence of phonological training; HW tasks adopted); set size and training regime (amount of training, duration and frequency of sessions); perceptual learning tasks adopted and granularity of visual discrimination; and age of learners.

The present meta-analysis showed that the benefit from HW training (over control training) was reliable across different outcomes of perceptual learning, but, more important, this effect was larger when fine-grained than coarse-grained discrimination of graphs was involved. Note that, albeit relevant, this result is compatible with the operation of both sensorimotor and perceptual learning mechanisms. Crucially, the two accounts predict opposing effects in what regards both the magnitude of the advantage from HW relative to different types of control training and the moderator role played by phonological training. These results were thus more enlightening.

We found that the benefit from HW training was larger relative to motor control training, such as typewriting, than when control training was merely visual. In the same vein, the benefit from HW training was larger when phonological training was absent than present. This pattern of results agrees with predictions from symbolic accounts which hold that the cognitive mechanism driven by HW is

perceptual learning rather than the motor experience (e.g., Li & James, 2016; Seyll et al., 2020; Williams, 1975). HW is the most natural activity promoting a visual analysis of the component parts of graphs and awareness of the dimensions of shape, orientation, and spatial relations. This seems to be the reason why HW training benefits reading (e.g., Courrieu & de Falco, 1989; Seyll et al., 2020).

It is hard to see how the sensorimotor hypothesis would explain a larger advantage from HW relative to motor control than to visual-only training, given that in the latter no sensorimotor integration occurred and the learning experience was the poorest in terms of brain-body-environment interaction. In the same vein, if the graphomotor component was backbone of graph representations, then it is also difficult to explain how sensorimotor mechanisms could accommodate a smaller advantage from HW when phonological training was also part of the learning experience.

The present results agree with previous evidence discrepant with the sensorimotor account (e.g., Mayer et al., 2020; Z. Xu et al., 2020). Some authors have suggested that post-training evidence of enhancement in functional activity at vOT (e.g., James, 2010) and of connectivity with frontal and parietal (motor) regions when viewing graphs trained via HW, relative to control conditions, directly support the sensorimotor hypothesis (James & Atwood, 2009; James & Engelhardt, 2012; Longcamp et al., 2008; Longcamp et al., 2016; Vinci-Booher et al., 2016). However, correlational evidence does not allow for a causal interpretation of the role of motor representations in visual graph recognition. Even if we (incorrectly) assumed it, then the prediction that follows is that damage in brain regions responsible for HW would necessarily lead to reading impairments, which is not the case (e.g., Kurosaki, Hashimoto, Tatsumi, & Hadano, 2016; Vandenborre, van Dun, Engelborghs, & Marien, 2015). Furthermore, Vinci-Booher et al. (2021) recently showed that visual-motor functional connections observed immediately after HW training are transient and no longer found one week after the end of training, which demonstrates that they are not directly related to the gains promoted by HW training in subsequent visual graph recognition.

It is for sure possible to acquire reading in the absence of writing, which was the norm until the XX century (Lyons, 2020). More important, the neural tuning for graphs is already visible after short-term training in letter-speech sound correspondences without HW (Brem et al., 2010; Brem et al., 2018). In sum, the present results and the current body of literature do not provide compelling

evidence that embodied representations underpin the benefit from HW training in visual graph recognition. Instead, the present results seem to agree with prior evidence showing that as long as learners are trained on diagnostic features of graphs (e.g., differences in orientation between highly confusable graphs), no additional benefit from HW training is found (Courrieu & de Falco, 1989; Koenigsberg, 1973; Seyll et al., 2020; Williams, 1975). Perceptual training on distinctive features of graphs seems to be the critical factor for facilitating subsequent letter-speech sound learning (Nelson & Wein, 1974; Samuels, 1973; Torres et al., 2020), given that mere visual exposure or explicit training in non-diagnostic graph features (e.g., size, shape) do not enhance subsequent letter discrimination (Caldwell & Hall, 1969; Nelson & Wein, 1974; Pick, 1965; Tawney, 1972; Williams, 1969). Therefore, it is because HW is a natural way of attending to visual diagnostic features that its practice can foster visual graph recognition and learning to read. Indeed, we found that the advantage from HW training was significant both in visual discrimination and in phonological mapping tasks, and hence, the present results do speak about learning to read.

The visual analysis and the perceptual variability hypotheses focus on different operations of perceptual learning, but they are not mutually exclusive. Nonetheless, the perceptual variability hypothesis puts the tonic into category learning, given that when two tasks of HW were directly compared, namely copying and tracing, only the former led to greater functional activation at the vOT for trained letters than geometric shapes (James & Engelhardt, 2012). In the present work, type of HW task was not a significant moderator of the HW benefit. This null result must be interpreted with caution given that statistical power was low. Note, however, that brain and behavioral differences between the two graphomotor tasks might rather be transient (Vinci-Booher & James, 2020) and uninformative about the underlying mechanism (Vinci-Booher et al., 2021).

Regarding the potential role of age (e.g., Longcamp et al., 2005), we found no significant moderator role of this factor. However, children were overall trained for longer and on less symbols than adults, and training regime (amount of training, duration and frequency of sessions) was a reliable moderator of the benefit from HW training for both groups, at least in multiple-session programs. Therefore, an effect of age might have been disguised by confounds.

Note, however, that for training studies on real letters by preliterate children, the influence of

age (e.g., Labat et al., 2010; Longcamp et al., 2005) could rather reflect uncontrolled exposure to graphs outside training. This would be more likely for the oldest children, who were closer to school entrance, and hence, possibly more motivated to benefit from such training. Moreover, several studies have shown that age is not a relevant factor when learning to read. On the one hand, learning to read enhances orientation processing and mirror-image discrimination both in children and in adults (e.g., Fernandes et al., 2016; Kolinsky et al., 2011) and the same neural circuitry comprising the vOT supports visual graph recognition and reading regardless of the age at which literacy was acquired (Brem et al., 2010; Dehaene-Lambertz et al., 2018; Dehaene et al., 2010; Hervais-Adelman et al., 2019). On the other hand, HW training (in an invented-letter task) follows the same curve of learning, and the same improvement in consolidation and retention in preliterate children, in beginning readers, and in young adults (Julius & Adi-Japha, 2015).

Limitations and future directions for research

To our knowledge, no primary study has hitherto confronted the three hypotheses considered here nor some of the moderators that we found to be relevant in the present meta-analysis. Given that a synthesis of the available primary studies can only consider the parameters that have thus far been examined, we must acknowledge that the present work can only shed some light on the underlying mechanisms, which will hopefully be addressed in future research.

In particular, the available training studies varied substantially in set size of the script (from 3 to 30 graphs: e.g., Suggate et al., 2016; Guan et al., 2011) and in training regime (from a single session of 1-30 min: Naka & Naoi, 1995; Guan et al., 2011; to 28 sessions of ~25 min each, with four sessions per week: Mayer et al., 2020). On average, participants were trained on 11 new graphs for 84 min distributed over 1-2 sessions per week of 17 min each (with a significant dispersion on these parameters). Note that training regime has been overlooked (although with references by Jefferson, 1953; Longcamp et al., 2006), and probably for this reason, we found significant correlations with set size. Overall, the more graphs participants were taught, the longer and the more frequent each session. Longer programs (larger amount of training) were also significantly associated with more systematic training. Given that these parameters were correlated, it was not possible to analyze their independent role on the benefit from HW training. Yet, these parameters significantly predicted the size of the

benefit from HW training in multiple-session programs, especially for children.

Both perceptual and sensorimotor learning involve a complex interplay between an online phase of (transient) improvement, corresponding to learning within-session, and a second phase, post-training, reflecting memory processes and stabilization of cognitive representations underpinning long-term permanent effects (e.g., Ghanamah et al., 2020; Prettyman, 2019; Stickgold, James, & Hobson, 2000). Indeed, when viewing (real) letters, only fluent adult readers showed functional activation of a distributed network including parietal and frontal regions, whereas activation was restricted to the vOT in 8-year-olds, and it was only in 6-year-old beginning readers (who still did not know all letters of the alphabet) that functional activity at the bilateral temporal cortex was modulated by letter forms (handwritten vs. printed: Vinci-Booher & James, 2020). Training along multiple sessions also enables the involvement of sleep, which is important in perceptual learning and visual discrimination (Stickgold et al., 2000) and also specifically in mirror-image discrimination of graphs when learning to read (Torres et al., 2020). Additionally, follow-up assessment is of critical importance, and albeit referred by Longcamp et al. (2005, 2008) and Vinci-Booher and James (2020), it was only included in few studies and with a disperse interval. In sum, the nature of graph representations may change as a function of amount of training and parameters of training regime seem relevant, and hence, should be the focus of future research.

Implications for practice

The present results provide key insights into the effectiveness of education practices of literacy instruction and on the role of HW training when learning to read, answering a recurrent question from parents, teachers, and policy makers. This translational outcome is particularly relevant given that the golden age of HW has already been gone (Lyons, 2020), antagonist forces conspire against this practice (Heuer, 2016), but it is still common in education (Santangelo & Graham, 2016), as it should be. Indeed, the present work highlights the importance of increasing HW practice in preschool and primary school years, especially when combined with perceptual training on the diagnostic features of graphs, including orientation-differences of highly confusable pairs, and accompanied by exposure to variable instances of letters. The learner will thus have faced the first challenge when learning to read, becoming able to quickly recognize visual graphs, and will then be

ready for the next challenge of mapping these orthographic symbols into phonological representations, reaching the essence of reading.

Conclusion

The overall benefit of HW training when learning to read is now established in different written systems and types of scripts. The present meta-analysis synthesized the available primary studies and provided two important contributions. This work provides support for the significant role of HW training when learning visual graphs. More important, by examining potential moderators of this HW advantage, we have shed some light on the underlying mechanism, which seems to be about perceptual learning rather than the motor act per se. Future research should clarify the nature of this mechanism by disentangling the intricate role of the moderator factors on the benefit from HW training when learning to read.

References

References marked with * correspond to studies included in this meta-analysis

- Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology*, 85(3), 478-508. doi:10.1037/0022-0663.85.3.478
- Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. P. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 32-60). Edinburgh, UK: Churchill Livingstone.
- Araujo, S., Faisca, L., Bramaio, I., Reis, A., & Petersson, K. M. (2015). Lexical and sublexical orthographic processing: An ERP study with skilled and dyslexic adult readers. *Brain & Language*, 141, 16-27. doi:10.1016/j.bandl.2014.11.007
- * Bara, F., Fredembach, B., & Gentaz, É. (2010). Rôle des procédures exploratoires manuelles dans la perception haptique et visuelle de formes chez des enfants scolarisés en cycle 2. [Role of manual exploratory procedures on the haptic and visual perception of shapes in children.]. *L'Année psychologique*, 110(2), 197-225. doi:10.4074/S0003503310002022
- Bara, F., & Gentaz, E. (2011). Haptics in teaching handwriting: The role of perceptual and visuo-motor skills. *Human Movement Science*, 30(4), 745-759. doi:10.1016/j.humov.2010.05.015
- * Bara, F., Gentaz, E., & Colé, P. (2007). Haptics in learning to read with children from low socioeconomic status families. *British Journal of Developmental Psychology*, 25(4), 643-663. doi:10.1348/026151007X186643
- * Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, 19(3), 433-449. doi:10.1016/j.cogdev.2004.05.003
- Bara, F., Morin, M.-F., Alamargot, D., & Bosse, M.-L. (2016). Learning different allographs through handwriting: The impact on letter knowledge and reading acquisition. *Learning and Individual Differences*, 45, 88-94. doi:10.1016/j.lindif.2015.11.020
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On Staying Grounded and Avoiding Quixotic Dead Ends. *Psychonomic Bulletin & Review*, 23(4), 1122-1142. doi:10.3758/s13423-016-1028-3

* Bhide, A. (2018). Copying helps novice learners build orthographic knowledge: methods for teaching Devanagari akshara. *Reading and Writing*, 31(1), 1-33. doi:10.1007/s11145-017-9767-8

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). Comprehensive Meta-Analysis (Version 3) [Computer software]. Englewood, NJ: Biostat.

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

Bornstein, M. H., Gross, C. G., & Wolf, J. Z. (1978). Perceptual similarity of mirror images in infancy. *Cognition*, 6, 89-116. doi:10.1016/0010-0277(78)90017-3

Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E., Lyytinen, H., . . . Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939-7944. doi:10.1073/pnas.0904402107

Brem, S., Hunkeler, E., Machler, M., Kronschnabel, J., Karipidis, II, Pleisch, G., & Brandeis, D. (2018). Increasing expertise to a novel script modulates the visual N1 ERP in healthy adults. *International Journal of Behavioral Development*, 42(3), 333-341. doi:10.1177/0165025417727871

Caldwell, E. C., & Hall, V. C. (1969). The Influence of Concept Training on Letter Discrimination. *Child Development*, 40(1), 63-71. doi:10.2307/1127155

Cao, F., Rickles, B., Vu, M., Zhu, Z. H., Chan, D. H. L., Harris, L. N., . . . Perfetti, C. A. (2013). Early stage visual-orthographic processes predict long-term retention of word form and meaning: A visual encoding training study. *Journal of Neurolinguistics*, 26(4), 440-461. doi:10.1016/j.jneuroling.2013.01.003

* Cao, F., Vu, M., Chan, D. H. L., Lawrence, J. M., Harris, L. N., Guan, Q., . . . Perfetti, C. A. (2013). Writing affects the brain network of reading in Chinese: A functional magnetic resonance imaging study. *Human Brain Mapping*, 34(7), 1670-1684. doi:10.1002/hbm.22017

- Chang, L. Y., Chen, Y. C., & Perfetti, C. A. (2018). GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages. *Behavior Research Methods*, 50(1), 427-449. doi:10.3758/s13428-017-0881-y
- * Chang, L. Y., Xu, Y., Perfetti, C. A., Zhang, J., & Chen, H.-C. (2014). Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language. *International Journal of Disability, Development and Education*, 61(3), 288-305. doi:10.1080/1034912X.2014.934016
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed. ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed. ed.). Thousand Oaks, CA: Sage.
- Courrieu, P., & de Falco, S. (1989). Segmental vs. dynamic analysis of letter shape by preschool children. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 9(2), 189-198.
- Daniels, P. T., & Share, D. L. (2018). Writing System Variation and Its Consequences for Reading and Dyslexia. *Scientific Studies of Reading*, 22(1), 101-116. doi:10.1080/10888438.2017.1379082
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol*, 16(3), e2004103. doi:10.1371/journal.pbio.2004103
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16(4), 234-244. doi:10.1038/nrn3924
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9(7), 335-341. doi:10.1016/j.tics.2005.05.004
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes, G., Jobert, A., . . . Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330, 1359-1364. doi:10.1126/science.1194140
- Duval, S., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463.

- Fernandes, T., Leite, I., & Kolinsky, R. (2016). Into the looking glass: Literacy acquisition and mirror invariance in preschool and first-grade children. *Child Development*, 87(6), 2008 - 2025.
doi:10.1111/cdev.12550
- * Gentaz, E., Colé, P., & Bara, F. (2003). Evaluation of multisensoriel reading training in kindergarten children: The role of haptic exploration. *L'Année psychologique*, 103, 561-584.
doi:10.3406/psy.2003.29652
- Ghanamah, R., Eghbaria-Ghanamah, H., Karni, A., & Adi-Japha, E. (2020). Too little, too much: A limited range of practice 'doses' is best for retaining grapho-motor skill in children. *Learning and Instruction*, 69, 14. doi:10.1016/j.learninstruc.2020.101351
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT, US: Appleton-Century-Crofts.
- Gibson, E. J. (1970). The ontogeny of reading. *American Psychologist*, 25(2), 136-143.
doi:10.1037/h0029419
- Gibson, E. J., Pick, A. D., Osser, H., & Gibson, J. J. (1962). A developmental study of discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, 55, 897-906.
doi:10.1037/h0043190
- Gilbert, C. D., & Li, W. (2012). Adult Visual Cortical Plasticity. *Neuron*, 75(2), 250-264.
doi:10.1016/j.neuron.2012.06.030
- Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 23(4), 959-978. doi:10.3758/s13423-015-0860-1
- Goldstone, R. L., & Byrge, L. A. (2015). Perceptual learning. In *The Oxford handbook of philosophy of perception*. (pp. 814-832). New York, NY, US: Oxford University Press.
- Grainger, J. (2018). Orthographic processing: A 'mid-level' vision of reading: The 44th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 71(2), 335-359.
doi:10.1080/17470218.2017.1314515
- Griffin, J. W. (2020). MetapowerR: an R package for computing meta-analytic statistical power. . R package version 0.2.1, <https://CRAN.R-project.org/package=metapower>.

- * Guan, C. Q., Liu, Y., Chan, D. H. L., Ye, F. F., & Perfetti, C. A. (2011). Writing Strengthens Orthography and Alphabetic-Coding Strengthens Phonology in Learning to Read Chinese. *Journal of Educational Psychology*, 103(3), 509-522. doi:10.1037/a0023730
- * Guan, C. Q., Perfetti, C. A., & Meng, W. (2015). Writing quality predicts Chinese learning. *Reading and Writing*, 28(6), 763–795. doi:10.1007/s11145-015-9549-0
- * Guan, C. Q., & Wang, Y. (2017). The Effect of Handwriting Training on Language Learning Among Deaf Children and Their Matched Hearing Peers in China. *American Annals of the Deaf*, 162(3), 265-276. Retrieved from <https://www.jstor.org/stable/26235346>
- Guralnick, M. J. (1972). Alphabet Discrimination and Distinctive Features: Research Review and Educational Implications. *Journal of Learning Disabilities*, 5(7), 427-434. doi:10.1177/002221947200500707
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490-499. doi:10.1037/0033-2909.92.2.490
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9(4), 426-445. doi:10.1037/1082-989x.9.4.426
- Hervais-Adelman, A., Kumar, U., Mishra, R. K., Tripathi, V. N., Guleria, A., Singh, J. P., . . . Huettig, F. (2019). Learning to read recycles visual cortical networks without destruction. *Science Advances*, 5(9), eaax0262. doi:10.1126/sciadv.aax0262
- Heuer, H. (2016). Technologies shape sensorimotor skills and abilities *Trends in Neuroscience and Education*, 5(3), 121–129.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., . . . (Eds). (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (2nd ed.). Hichester, UK: John Wiley & Sons.
- * Hulme, C., Monk, A., & Ives, S. (1987). Some experimental studies of multi-sensory teaching: the effects of manual tracing on children's paired-associate learning. *British Journal of Developmental Psychology*, 5(4), 299-307. doi:10.1111/j.2044-835X.1987.tb01066.x

- * James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental Science*, 13(2), 279-288. doi:10.1111/j.1467-7687.2009.00883.x
- James, K. H. (2017). The Importance of Handwriting Experience on the Development of the Literate Brain. *Current Directions in Psychological Science*, 26(6), 502-508. doi:10.1177/0963721417709821
- * James, K. H., & Atwood, T. P. (2009). The role of sensorimotor learning in the perception of letter-like forms: tracking the causes of neural specialization for letters. *Cognitive Neuropsychology*, 26(1), 91-110. doi:10.1080/02643290802425914
- James, K. H., & Engelhardt, L. (2012). The effects of handwriting experience on functional brain development in pre-literate children. *Trends in Neuroscience and Education*, 1(1), 32-42. doi:10.1016/j.tine.2012.08.001
- James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory-motor brain network. *Neuropsychologia*, 44(14), 2937-2949. doi:10.1016/j.neuropsychologia.2006.06.026
- Jeffrey, W. E. (1958). Variables in early discrimination-learning. I. Motor-responses in the training of a left-right discrimination. *Child Development*, 29(2), 269-275.
- * Jensen, N. J., & King, E. M. (1970). Effects of different kinds of visual-motor discrimination training on learning to read words. *Journal of Educational Psychology*, 61(2), 90-96. doi:10.1037/h0028913
- Julius, M. S., & Adi-Japha, E. (2015). Learning of a simple grapho-motor task by young children and adults: similar acquisition but age-dependent retention. *Frontiers in Psychology*, 6, 11. doi:10.3389/fpsyg.2015.00225
- * Kersey, A. J., & James, K. H. (2013). Brain activation patterns resulting from learning letter forms through active self-production and passive observation in young children. *Frontiers in Psychology*, 4, 567. doi:10.3389/fpsyg.2013.00567
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805-825. doi:10.1016/j.cortex.2011.04.006

- * Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or Typewriting? The Influence of Pen- or Keyboard-Based Writing Training on Reading and Writing Performance in Preschool Children. *Advances in cognitive psychology*, 11(4), 136-146. doi:10.5709/acp-0178-7
- Kirk, U. (1980). Learning to Copy Letters: A Cognitive Rule-Governed Task. *The Elementary School Journal*, 81(1), 29-33. doi:10.1086/461203
- Koenigsberg, R. S. (1973). An evaluation of visual versus sensorimotor methods for improving orientation discrimination of letter reversals by preschool children. *Child Development*, 44(4), 764-769.
- Kolinsky, R., Verhaeghe, A., Fernandes, T., Mengarda, E. J., Grimm-Cabral, L., & Morais, J. (2011). Enantiomorphy through the looking glass: literacy effects on mirror-image discrimination. *Journal of Experimental Psychology: General*, 140, 210-238. doi:10.1037/A0022168
- Kurosaki, Y., Hashimoto, R., Tatsumi, H., & Hadano, K. (2016). Pure agraphia after infarction in the superior and middle portions of the left precentral gyrus: Dissociation between Kanji and Kana. *Journal of Clinical Neuroscience*, 26, 150-152. doi:10.1016/j.jocn.2015.08.030
- * Labat, H., Boisson, A., Brunel, L., Ecalle, J., Versace, R., & Magnan, A. (2020). Multisensory letter integration and implicit learning of reading with 5-year-old children. *European Review of Applied Psychology*, 70(1), 100477. doi:10.1016/j.erap.2019.100477
- Labat, H., Ecalle, J., & Magnan, A. (2010). Effet d'entraînements bimodaux à la connaissance des lettres. Étude transversale chez des enfants de trois et cinq ans. *Psychologie Française*, 55(2), 113-127. doi:10.1016/j.psfr.2009.12.002
- * Labat, H., Vallet, G., Magnan, A., & Ecalle, J. (2015). Facilitating Effect of Multisensory Letter Encoding on Reading and Spelling in 5-Year-Old Children. *Applied Cognitive Psychology*, 29(3), 381-391. doi:10.1002/acp.3116
- * Li, J. X., & James, K. H. (2016). Handwriting generates variable visual output to facilitate symbol learning. *Journal of Experimental Psychology: General*, 145(3), 298-313. doi:10.1037/xge0000134

Logotheitis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552-563.

* Longcamp, M., Boucard, C., Gilhodes, J. C., Anton, J. L., Roth, M., Nazarian, B., & Velay, J. L. (2008). Learning through hand- or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience*, 20(5), 802-815. doi:10.1162/jocn.2008.20504

* Longcamp, M., Boucard, C., Gilhodes, J. C., & Velay, J. L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human Movement Science*, 25(4-5), 646-656. doi:10.1016/j.humov.2006.07.007

Longcamp, M., Richards, T. L., Velay, J. L., & Berninger, V. W. (2016). Neuroanatomy of Handwriting and Related Reading and Writing Skills in Adults and Children with and without Learning Disabilities: French-American Connections. *Pratiques*, 171-172. doi:10.4000/pratiques.3175

* Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: a comparison between handwriting and typing. *Acta Psychologica*, 119(1), 67-79. doi:10.1016/j.actpsy.2004.10.019

Lyons, M. (2020). The Common Writer since 1500. In S. Eliot & J. Rose (Eds.), *A Companion to the History of the Book* (pp. 759-772): John Iley & Sons.

Lyu, B., Lai, C., Lin, C.-H., & Gong, Y. (2021). Comparison studies of typing and handwriting in Chinese language learning: A synthetic review. *International Journal of Educational Research*, 106, 101740. doi:10.1016/j.ijer.2021.101740

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*, 23(4), 1090-1095. doi:10.3758/s13423-015-0878-4

Mahon, B. Z. (2015). What is embodied about cognition? *Language Cognition and Neuroscience*, 30(4), 420-429. doi:10.1080/23273798.2014.987791

Mahon, B. Z., & Hickok, G. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bulletin & Review*, 23(4), 941-958. doi:10.3758/s13423-016-1045-2

Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping*, 36(5), 1963-1981. doi:10.1002/hbm.22749

* Mayer, C., Wallner, S., Budde-Spengler, N., Braunert, S., Arndt, P. A., & Kiefer, M. (2020). Literacy Training of Kindergarten Children With Pencil, Keyboard or Tablet Stylus: The Influence of the Writing Tool on Reading and Writing Performance at the Letter and Word Level. *Frontiers in Psychology*, 10, 17. doi:10.3389/fpsyg.2019.03054

Merritt, E., Swain, S. N., Vinci-Booher, S., & James, K. H. (2020). Constraining Stroke Order During Manual Symbol Learning Hinders Subsequent Recognition in Children Under 4 1/2 Years. *Frontiers in Psychology*, 11, 8. doi:10.3389/fpsyg.2020.00500

Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. *Journal of Neurolinguistics*, 49, 71-83. doi:10.1016/j.jneuroling.2018.08.005

* Naka, M., & Naoi, H. (1995). The effect of repeated writing on memory. *Memory & Cognition*, 23(2), 201-212. doi:10.3758/BF03197222

Nakamura, K., Honda, M., Hirano, S., Oga, T., Sawamoto, N., Hanakawa, T., . . . Shibasaki, H. (2002). Modulation of the visual word retrieval system in writing: A functional MRI study on the Japanese orthographies. *Journal of Cognitive Neuroscience*, 14(1), 104-115. doi:10.1162/089892902317205366

Nakamura, K., Kuo, W.-J., Pegado, F., Cohen, L., Tzeng, O. J. L., & Dehaene, S. (2012). Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proceedings of the National Academy of Sciences*, 109(50), 20762-20767. doi:10.1073/pnas.1217749109

- Nelson, R. O., & Wein, K. S. (1974). Training letter discrimination by presentation of high-confusion versus low-confusion alternatives. *Journal of Educational Psychology*, 66(6), 926-931.
doi:10.1037/h0021537
- * Ouellette, G., & Sénéchal, M. (2008). Pathways to literacy: a study of invented spelling and its role in learning to read. *Child Development*, 79(4), 899-913. doi:10.1111/j.1467-8624.2008.01166.x
- Pegado, F., Nakamura, K., & Hannagan, T. (2014). How does literacy break mirror invariance in the visual system? *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00703
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46(28), 4646-4674. doi:10.1016/j.visres.2006.04.023
- Pick, A. D. (1965). Improvement of visual and tactual form discrimination. *Journal of Experimental Psychology*, 69(4), 331-339. doi:10.1037/h0021772
- Planton, S., Jucla, M., Roux, F. E., & Demonet, J. F. (2013). The "handwriting brain": a meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex*, 49(10), 2772-2787. doi:10.1016/j.cortex.2013.05.011
- Prettyman, A. (2019). Perceptual learning. *Wiley Interdisciplinary Reviews-Cognitive Science*, 10(3), 15. doi:10.1002/wcs.1489
- Rohatgi, A. (2015). WebPlotDigitizer Version 3.9 [Computer software]. Retrieved from <http://arohatgi.info/WebPlotDigitizer>
- Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59(2), 464-468.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. London, England: Sage.
- Rothlein, D., & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the multiple representations of letters. *Neuroimage*, 89, 331-344. doi:10.1016/j.neuroimage.2013.11.054
- Samuels, S. J. (1973). Effect of distinctive feature training on paired-associate learning. *Journal of Educational Psychology*, 64(2), 164-170. doi:10.1037/h0034607

Santangelo, T., & Graham, S. (2016). A Comprehensive Meta-analysis of Handwriting Instruction.

Educational Psychology Review, 28(2), 225-265. doi:10.1007/s10648-015-9335-1

* Seyll, L., Wyckmans, F., & Content, A. (2020). The impact of graphic motor programs and detailed visual analysis on letter-like shape recognition. *Cognition*, 205, 104443.

doi:10.1016/j.cognition.2020.104443

Share, D. L. (1995). Phonological recoding and Self-teaching - sine-qua-non of reading acquisition.

Cognition, 55(2), 151-218. doi:10.1016/0010-0277(94)00645-2

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3(12), 1237-1238. doi:10.1038/81756

* Suggate, S., Pufke, E., & Stoeger, H. (2016). The effect of fine and grapho-motor skill demands on preschoolers' decoding skill. *Journal of Experimental Child Psychology*, 141, 34-48.

doi:10.1016/j.jecp.2015.07.012

Suggate, S., Pufke, E., & Stoeger, H. (2018). Do fine motor skills contribute to early reading

development? *Journal of Research in Reading*, 41(1), 1-19. doi:10.1111/1467-9817.12081

Tawney, J. W. (1972). Training letter discrimination in four-year-old children. *Journal of Applied Behavior Analysis*, 5(4), 455-465. doi:10.1901/jaba.1972.5-455

Torres, A. R., Mota, N. B., Adamy, N., Naschold, A., Lima, T. Z., Copelli, M., . . . Ribeiro, S. (2020). Selective Inhibition of Mirror Invariance for Letters Consolidated by Sleep Doubles Reading Fluency. *Current Biology*. doi:10.1016/j.cub.2020.11.031

Vandenborre, D., van Dun, K., Engelborghs, S., & Marien, P. (2015). Apraxic agraphia following thalamic damage: Three new cases. *Brain & Language*, 150, 153-165.

doi:10.1016/j.bandl.2015.05.011

Vinci-Booher, S., & James, K. H. (2020). Visual experiences during letter production contribute to the development of the neural systems supporting letter perception. *Developmental Science*, e12965. doi:10.1111/desc.12965

Vinci-Booher, S., James, T. W., & James, K. H. (2016). Visual-motor functional connectivity in preschool children emerges after handwriting experience. *Trends in Neuroscience and Education*, 5(3), 107-120. doi:10.1016/j.tine.2016.07.006

- * Vinci-Booher, S., James, T. W., & James, K. H. (2021). Visual-motor contingency during symbol production contributes to short-term changes in the functional connectivity during symbol perception and long-term gains in symbol recognition. *Neuroimage*, 227, 117554. doi:10.1016/j.neuroimage.2020.117554
- * Williams, J. (1969). Training Kindergarten Children To Discriminate Letter-Like Forms. *American Educational Research Journal*, 6(4), 501-514. doi:10.3102/00028312006004501
- Williams, J. (1975). Training children to copy and to discriminate letterlike forms. *Journal of Educational Psychology*, 67(6), 790-795. doi:10.1037/0022-0663.67.6.790
- * Xu, Y., Chang, L. Y., Zhang, J., & Perfetti, C. A. (2013). Reading, Writing, and Animation in Character Learning in Chinese as a Foreign Language. *Foreign Language Annals*, 46(3), 423-444. doi:10.1111/flan.12040
- Xu, Z., Liu, D., & Joshi, R. M. (2020). The Influence of Sensory-Motor Components of Handwriting on Chinese Character Learning in Second- and Fourth-Grade Chinese Children. *Journal of Educational Psychology*, 112(7), 1353-1366. doi:10.1037/edu0000443
- Yin, Y., & Zhang, Q. (2021). Chinese characters are read using not only visual but also writing motor information. *Psychophysiology*, 58(1), e13696. doi:10.1111/psyp.13696
- * Zemlock, D., Vinci-Booher, S., & James, K. H. (2018). Visual-motor symbol production facilitates letter recognition in young children. *Reading and Writing*, 31(6), 1255-1271. doi:10.1007/s11145-018-9831-z
- * Zhao, P., Li, S., Zhao, J., Gaspar, C. M., & Weng, X. (2015). Training by visual identification and writing leads to different visual word expertise N170 effects in preliterate Chinese children. *Dev Cogn Neurosci*, 15, 106-116. doi:10.1016/j.dcn.2015.09.002
- * Zhao, P., Zhao, J., Weng, X., & Li, S. (2018). Event-related potential evidence in Chinese children: Type of literacy training modulates neural orthographic sensitivity. *International Journal of Behavioral Development*, 42(3), 311-320. doi:10.1177/0165025417708341
- Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science*, 9(5), 429-436. doi:10.1111/j.1467-7687.2006.00509.x

Table 1

Distribution of studies by age of participants, type of control training (visual or motor control) and phonological training (Ph+ vs. Ph-).

Age	Type of training			
	Visual control (<i>k</i> = 24)		Motor control (<i>k</i> = 25)	
	Ph+	Ph-	Ph+	Ph-
Children (<i>k</i> = 28)	14	2	9	3
Adults (<i>k</i> = 22)	5	2	4	9
Across-age	19	4	13	12

Note. *k* corresponds to number of studies included in each subgroup.

Table 2

Descriptive statistics ($M \pm SD$, Median, Mdn, and range) for set size and training regime by categorical moderators.

		Training regime		
Categorical moderator	Set size	Session dur ^a	Frequency ^b	Amount ^c
Age				
children	8.28 ± 5.39	19.33 ± 8.80	1.63 ± 1.00	126.44 ± 153.59
	Mdn = 6.00 (3 - 26)	Mdn = 25.00 (3 - 30)	Mdn = 1.00 (1 - 4)	Mdn = 90.00 (3 - 700)
adults	14.23 ± 8.33	12.97 ± 9.82	1.71 ± 1.10	32.15 ± 37.61
	Mdn = 11.00 (5 - 30)	Mdn = 11.50 (1 - 30)	Mdn = 2.00 (1 - 5)	Mdn = 20.50 (1 - 150)
Control training				
visual	9.20 ± 5.67	16.78 ± 9.13	1.54 ± 0.91	81.87 ± 87.92
	Mdn = 8.00 (3 - 26)	Mdn = 15.00 (1 - 30)	Mdn = 1.00 (1 - 2)	Mdn = 40.00 (1 - 350)
motor	12.46 ± 8.50	16.21 ± 10.38	1.77 ± 1.14	86.09 ± 152.38
	Mdn = 11.00 (3 - 30)	Mdn = 15.50 (1 - 30)	Mdn = 1.00 (1 - 5)	Mdn = 30.00 (1.4 - 700)
Phonological training				
Ph+	11.16 ± 8.43	18.07 ± 9.26	1.87 ± 1.14	106.89 ± 146.59
	Mdn = 8.00 (3 - 30)	Mdn = 22.50 (1.4 - 30)	Mdn = 1.50 (1 - 5)	Mdn = 33.00 (1.4 - 700)
Ph-	10.44 ± 5.25	13.47 ± 19.12	1.33 ± 0.77	41.23 ± 47.95
	Mdn = 10.00 (3 - 26)	Mdn = 13.00 (1 - 30)	Mdn = 1.00 (1 - 4)	Mdn = 29.00 (1 - 180)

Note. $M \pm SD$ correspond to mean and standard deviation; minimum and maximum are in parenthesis.

^aDuration of each session in minutes. ^bNumber of sessions per week. ^cTotal duration of the training program in minutes.

Table 3

Primary studies: observed ES under random-effect weights, 95% CI (lower and upper limits, LL and HL), and statistics (z -value for comparison with null effect).

Primary study	Hedge's g	SE	95 % CI		Z
			LL	HL	
Williams, 1969 •	0.62	0.49	-0.33	1.57	1.272
Jensen & King, 1970	0.14	0.22	-0.29	0.57	0.630
Hulme et al., 1987 (Exp 1)	0.89	0.36	0.18	1.60	2.468 *
Hulme et al., 1987 (Exp 2)	1.26	0.43	0.41	2.11	2.906 *
Hulme et al., 1987 (Exp 3)	1.57	0.40	0.80	2.35	3.964 *
Hulme et al., 1987 (Exp 4)	2.83	0.49	1.86	3.80	5.731 *
Naka & Naoi, 1995 (Exp 1; pairs of symbols)	0.33	0.31	-0.28	0.94	1.057
Naka & Naoi, 1995 (Exp 1; single symbols)	0.19	0.31	-0.42	0.80	0.614
Naka & Naoi, 1995 (Exp 2: American participants)	0.17	0.25	-0.32	0.66	0.698
Naka & Naoi, 1995 (Exp 2: Japanese participants)	0.16	0.25	-0.33	0.64	0.638
Naka & Naoi, 1995 (Exp 3)	0.08	0.31	-0.53	0.68	0.245
Naka & Naoi, 1995 (Exp 4)	0.10	0.22	-0.33	0.54	0.459
Gentaz et al., 2003 •	0.44	0.39	-0.33	1.21	1.117
Bara et al., 2004 •	0.27	0.31	-0.35	0.88	0.851
Bara et al., 2004 (Control study)	1.07	0.32	0.43	1.70	3.290 *
Longcamp et al., 2005 (middle-age children)	-0.18	0.39	-0.94	0.58	-0.456
Longcamp et al., 2005 (older children)	1.10	0.42	0.28	1.92	2.641 *
Longcamp et al., 2005 (younger children)	0.37	0.38	-0.38	1.13	0.978
Longcamp et al., 2006 •	0.13	0.40	-0.64	0.91	0.335
Bara et al., 2007 •	0.28	0.34	-0.38	0.95	0.837
Longcamp et al., 2008	0.21	0.40	-0.57	0.98	0.523
Ouellette & Sénéchal, 2008 •	0.43	0.30	-0.14	1.01	1.470 +
James & Atwood, 2009	-0.32	0.54	-1.37	0.73	-0.598
Bara et al., 2010 (Study 1) •	0.15	0.31	-0.46	0.76	0.484
James, 2010	0.79	0.56	-0.30	1.88	1.417
Guan et al., 2011 (Exp 1)	0.55	0.26	0.04	1.06	2.106 *
Guan et al., 2011 (Exp 2)	0.28	0.32	-0.35	0.91	0.868
Kersey & James, 2013	0.20	0.35	-0.47	0.88	0.590
Xu et al., 2013 •	0.36	0.17	0.03	0.69	2.16 *
Cao et al., 2013 •	0.41	0.42	-0.41	1.23	0.98
Chang et al., 2014 (distributed learning approach)	-0.28	0.29	-0.84	0.28	-0.975
Chang et al., 2014 (grouped learning approach)	-0.28	0.29	-0.84	0.28	-0.975
Guan et al., 2015 •	0.37	0.17	0.04	0.71	2.18 *
Kiefer et al., 2015 •	-0.04	0.29	-0.60	0.52	-0.137
Labat et al., 2015	1.53	0.49	0.57	2.50	3.115 *
Zhao et al., 2015, 2018 •	0.93	0.28	0.38	1.47	3.354 *
Li & James, 2016 (copy, single typed symbols) •	0.40	0.28	-0.15	0.96	1.431
Li & James, 2016 (tracing, handwritten symbols)	0.76	0.29	0.20	1.33	2.636 *
Li & James, 2016 (tracing, single typed symbols) •	-0.43	0.28	-0.99	0.12	-1.539 +
Suggate et al., 2016	0.15	0.20	-0.23	0.54	0.766
Guan et al., 2017	0.58	0.36	-0.13	1.29	1.594 +
Bhide, 2018 •	0.13	0.17	-0.20	0.46	0.785
Zemlock et al., 2018 •	0.33	0.19	-0.03	0.70	1.793 +
Mayer et al., 2020 •	0.15	0.14	-0.13	0.43	1.046
Labat et al., 2020 •	0.66	0.34	0.00	1.31	1.961 *
Seyll et al., 2020 (Exp 1) •	3.58	0.51	2.58	4.59	6.980 *
Seyll et al., 2020 (Exp 2) •	3.33	0.50	2.35	4.31	6.681 *
Seyll et al., 2020 (Exp 3A) •	1.75	0.31	1.15	2.35	5.693 *
Seyll et al., 2020 (Exp 3B) •	3.25	0.39	2.48	4.02	8.297 *
Vinci-Booher et al., 2021	0.35	0.31	-0.26	0.97	1.133
Overall ES	0.58	0.09	0.39	0.76	6.154 *

• Studies on which more than one outcome measure contributed to the aggregated (average combined) ES.

* $p \leq .05$, + $.05 < p < .15$ (bilateral tests).

Table 4

Meta-regressions run with training regime parameters (i.e., session dur, session duration; frequency, number of sessions per week; amount, total duration of training) separately for adults and children.

Meta-regressions run with training regime separately for adults and children.

	parameter	β	Z	p-value ^a
<i>all studies included</i>				
children	intercept	0.217 (SE = 0.201)		
(k = 28)	frequency	.407 (SE = .142)	2.87	.004
	amount	-.003 (SE = .001)	-3.07	.002
adults	intercept	0.929 (SE = 0.409)		
(k = 22)	frequency	-.215 (SE = .294)	-0.73	.464
	amount	.003 (SE = .008)	0.41	.683
<i>multiple-session studies</i>				
children	intercept	0.891 (SE = 0.694)		
(k = 10)	frequency	.439 (SE = .257)	1.71	.087
	amount	-.003 (SE = .001)	-1.83	.067
	session dur	-.039 (SE = .036)	1.28	.199
adults	intercept	-3.844 (SE = 1.630)		
(k = 9)	frequency	.641 (SE = .411)	1.56	.119
	amount	-.073 (SE = .026)	-2.81	.005
	session dur	.409 (SE = 0.131)	3.11	.002

Note. p-value is two-sided (bilateral test). SE = standard error.

Figure captions

Fig. 1 Flowchart of the selection process of the primary studies included in this meta-analysis

Fig. 2 Funnel plot of the studies included in this meta-analysis (represented with blue dots in online version) according to their estimated ES (the overall ES is represented by the full diamond) by standard error (the Y-axis is reversed with low values at the top).

Fig. 3 Mean effect size (g) and 95% CI for each categorical moderator examined and the overall effect, k , and N of participants on whom these effect sizes were based upon.

Note. *** $p \leq .001$ ** $p < .01$; * $p < .05$.