Research Article

Psychological Science
2021, Vol. 32(7) 1086–1103

DOI: 10.1177/0956797621993111
www.psychologicalscience.org/PS

The Effects of Handwriting Experience on Literacy Learning

Robert W. Wiley[1,2] and Brenda Rapp[2,3,4]
[1]Department of Psychology, University of North Carolina Greensboro; [2]Department of Cognitive Science, Johns Hopkins University; [3]Department of Neuroscience, Johns Hopkins University; and [4]Department of Psychological and Brain Sciences, Johns Hopkins University

Abstract
Previous research indicates that writing practice may be more beneficial than nonmotor practice for letter learning. Here, we report a training study comparing typing, visual, and writing learning conditions in adults (N = 42). We investigated the behavioral consequences of learning modality on literacy learning and evaluated the nature of the learned letter representations. Specifically, the study addressed three questions. First, are the benefits of handwriting practice due to motor learning per se or to other incidental factors? Second, do the benefits generalize to untrained tasks? And third, does handwriting practice lead to learning and strengthening only of motor representations or of other types of representations as well? Our results clearly show that handwriting compared with nonmotor practice produces faster learning and greater generalization to untrained tasks than previously reported. Furthermore, only handwriting practice leads to learning of both motor and amodal symbolic letter representations.

Keywords
literacy, handwriting, letters, learning

Received 1/17/20; Revision accepted 12/1/20

One consequence of the ever-increasing role of the Internet, cell phones, and laptops is the rising use of written language both in our personal and workplace communications. We are constantly consuming and producing written language as we read and tap out messages. As a result, many people spend more time reading and writing in digital formats than they do using pen and paper. Relatedly, in schools, although reading has traditionally been taught in tandem with handwriting, time spent teaching handwriting has been vastly reduced (Deardorff, 2011; Konnikova, 2014). These changes have naturally led parents and educators to ask how much time and how many resources should be spent on teaching children to write by hand. Presumably, to justify the teaching of handwriting, any benefits should extend beyond the acquisition of fine penmanship to strengthening core aspects of literacy. Clearly, a better understanding of the effects on literacy of writing experience, compared with typing or nonmotor experiences, has significant educational implications.

Although letters appear to be very simple objects—just a few lines on a piece of paper or screen—they are, in fact, surprisingly complex and rich. Their complexity stems from the wealth of information we have about them. For example, we know that a single letter, A, can look like “A” or like “a”; “A” is likely to be written beginning with an upward stroke slanted to the right; in English, its name is /eɪ/, but it can represent the sounds /æ/ or /ɑ/; it is the first letter of the alphabet; on the keyboard it is situated to the left of “S” in the center row; and, as an English word, it indicates the indefinite article. We use letters in many tasks, such as reading, writing, and spelling, in ways that involve numerous cognitive processes. Further, it has been well established that better letter knowledge among young

Corresponding Author:
Robert W. Wiley, University of North Carolina Greensboro, Department of Psychology
E-mail: rwwiley@uncg.edu

children (pre-K and kindergarten) is predictive of reading and writing skills even into the middle-school years (Bara & Bonneton-Botté, 2018; Berninger et al., 2006; Treiman & Kessler, 2004; Treiman et al., 1998; Zemlock et al., 2018).

Several studies have reported that handwriting experience is more beneficial for letter learning than are other, nonmotor learning experiences. Of course, it is not surprising that handwriting practice would improve handwriting itself—what is critical is that it has been argued that the benefits of handwriting extend to other skills, such as letter recognition, categorization, and retention (Bhide, 2018; Li & James, 2016; Longcamp et al., 2005, 2006, 2008; Naka, 1998; Zemlock et al., 2018). These researchers have hypothesized that the behavioral benefits of handwriting may arise not only from incidental factors (e.g., greater attention or more time on task for handwriting vs. nonmotor conditions) but also more fundamental ones—in particular, benefits arising from feedback from the visual output of the motor acts (see Cao et al., 2013; Li & James, 2016; Nakamura et al., 2012) and/or strengthened motor representations (see blue lines and boxes in Fig. 1). Accordingly, there are a number of outstanding issues concerning each of the components of the learning process depicted in Figure 1. In this investigation, we focused on the following questions: (a) Are the benefits of handwriting practice due to motor learning per se or to other incidental factors involved in handwriting practice? (b) Do the benefits generalize to tasks that were not practiced during the handwriting training? and (c) Does handwriting practice lead to learning and strengthening only of motor representations or of other types of representations as well?

The third question is especially important for understanding the cognitive bases of generalization to untrained tasks. However, it has been the least investigated. Specifically, we have little understanding of which of the multiple types of letter representations (e.g., visual, motor, phonological) are affected by the nature of the learning experience. Of particular interest is the possible role of amodal, symbolic letter representations, referred to as *abstract letter identities* (Caramazza & Hillis, 1990) or *symbolic letter identities* (Rothlein & Rapp, 2014). These are posited to be amodal representations of letter identities that unify the multiple domain and format-specific representations of letters. For example, the letter name ("ay"), visual formats ("A"/"a"), and the motor program for "A" all share a representation that is symbolic and lacks domain-specific content. The claim that symbolic letter identities play a key role in reading and spelling is supported by behavioral (Chen & Proctor, 2012; Lupyan et al., 2010; Rothlein & Rapp, 2017; Schubert, Gawthrop, & Kinoshita, 2018; Wiley et al., 2016), neuroimaging (Rothlein &

Statement of Relevance

With the ever-growing use of the Internet, cell phones, and laptops, many people spend more time reading and writing in digital formats than with pen and paper. This has led to questions about the time and resources that should be spent on teaching handwriting. To inform this debate, we conducted a training study with adults that compared the benefits of learning new letters by typing, visual, or handwriting practice. The clear winner was handwriting practice, which resulted in the best performance on a set of reading and spelling tasks that were not specifically trained. This implies that participants with handwriting practice used their knowledge of how to write the letters to strengthen their performance across multiple tasks. Further research is needed to determine the best implementation of handwriting practice with children in the classroom. For now, we know that the benefits of writing practice extend beyond penmanship to letter and word reading and spelling, thus indicating that handwriting can be a productive use of learning time.

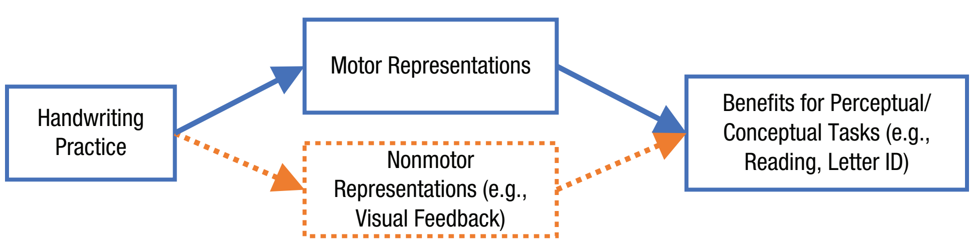

Fig. 1. Potential routes by which handwriting practice strengthens written language processing: through strengthening motor representations (blue arrows) learned through handwriting practice or through strengthening nonmotor representations (orange arrows), including visual feedback produced by writing.

Rapp, 2014), and neuropsychological (Caramazza & Hillis, 1990; Schubert & McCloskey, 2013) evidence. Symbolic letter identities have often been investigated in the Roman alphabet via analysis of tasks contrasting upper- and lowercase allographs on the basis of the logic that, under the symbolic-letter-identity hypothesis, upper- and lowercase letters share a common representation, despite visual dissimilarities (for examples in Hebrew, see Friedmann & Gvion, 2005; for examples in Arabic, see Carreiras et al., 2012; Wiley & Rapp, 2019; Wiley et al., 2016; and for examples in Japanese, see Kinoshita et al., 2019). Arabic, the script used in the present investigation, affords the opportunity to investigate the learning of symbolic letter identities, because most Arabic letters have multiple visual forms—allographs—that are used depending on the position of a letter within a word (see Fig. 2).

In this investigation, a total of 42 adult participants with no previous knowledge of Arabic learned 20 letters of that alphabet. Each participant was assigned to one of three learning conditions: *typing*, *visual*, or *writing*. The writing condition was a true motor condition, but the status of the typing condition was less clear. Although typing does require motor output, the relationship between the motor typing actions and the letter shapes is arbitrary (e.g., the action of typing a "T" as opposed to a "G" is almost the same, whereas writing

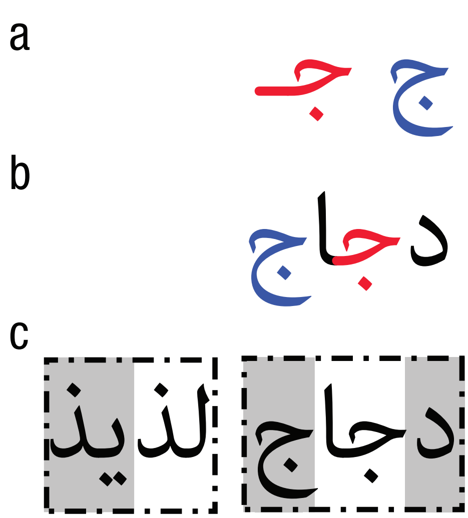

Fig. 2. Arabic allographs. In Arabic, different allographs of the same letter identity are used depending on the position of the letter within subwords. Subwords correspond to groupings of letters that are defined by nonligating letters (i.e., a letter that does not ligate on the left marks the end of a subword). The two allographs of the letter "jim" (/ʤi:m/) are shown in (a) in its isolated (blue) and ligating (red) forms. The two allographs occurring together in a word ("dajaaj," or "chicken") are shown in (b) along with nonligating (black) letters that create subwords. The relative spacing between words (boxes) and subwords (shading) is shown in (c).

Table 1. Tasks Administered at Each Time Point

Time point	Task
Before training	Same/different judgment
During training	Training task (typing, visual, or writing)
	Letter recognition
	Letter naming
After training	Letter recognition
	Letter naming
	Letter writing
	Word spelling
	Word reading
	Same/different judgment

them requires very different motor plans). The visual condition required only simple key presses and thus was similar in this way to the typing condition. Six behavioral assessments were administered at one or more time points: before, during, and after training (Table 1).

We used Arabic because it provides an existing organized alphabetic code (although an artificial script would also have been appropriate) and because of the ease of recruiting individuals with no prior knowledge of the language. The first question, regarding whether the behavioral benefits of handwriting are attributable to motor practice per se, was addressed through a study design that specifically addressed issues raised in previous work. These issues, which have thus far not been simultaneously addressed in a single study, were dealt with by presenting participants in all conditions with (a) dynamic visual information and variable exemplars (e.g., Cao et al., 2013; Li & James, 2016) and (b) phonological information about both letter names and sounds (Kiefer et al., 2015). The second question, concerning whether the benefits of handwriting training generalize to untrained tasks, was addressed by evaluating a more comprehensive set of processing tasks involving both letters and words than had been previously used in this type of research. Finally, the third question, concerning the nature of the learned representations, was evaluated via a same/different letter-judgment task, administered before and after training, which required participants to decide whether or not pairs of letters were physically identical. By measuring how the different learning conditions affect the perceived similarity of letters (Wiley & Rapp, 2019; Wiley et al., 2016), this task provides a window into various dimensions of representational similarity for the learned letters (visual, phonological, motor, and/or symbolic).

Briefly, the results of the investigation show that handwriting practice provides greater benefits than either typing or visual practice for a wide range of tasks. Furthermore, we found that only handwriting

Table 2. Participant Demographics

Learning condition	Participants enrolled		Mean age (years)	Mean education (years)	Participants who completed posttraining session
	Men	Women			
Typing	3	11	21.7 (2.6)	15.8 (1.4)	12
Visual	4	10	21.1 (4.6)	15.2 (1.8)	12
Writing	3	11	21.6 (3.1)	15.8 (2.2)	12

Note: Standard deviations are given in parentheses.

practice leads to developing motor and amodal symbolic representations, both of which influence letter perception. In the General Discussion, we consider the implications of these findings for our understanding of previous work and for literacy education, as well as their relevance to theories of embodied cognition that concern fundamental questions about the relationship between action and perception.

Method

Participants

Forty-two participants were recruited from the greater Johns Hopkins community, including both staff and students. The participant demographics are reported in Table 2. Participants had no history of learning disabilities, no previous experience with Arabic, and normal or corrected-to-normal vision. Participants received a payment of $10 per session. Consent was obtained using procedures approved by the Johns Hopkins University Institutional Review Board. Fourteen individuals were assigned to each of the three learning conditions, and 36 participants completed the study through the posttraining time point (12 per condition). These sample sizes reflect the maximum number of participants that could be enrolled and trained given the time and resources allotted to the study, and they improved on sample sizes in previous similar studies (e.g., Longcamp et al., 2006).

Stimuli

During training, both single letters and short words were presented. Each training session included four blocks of learning trials: three letter blocks and one word block. In all blocks, the stimuli were presented as though written on the screen in a dynamic display at a rate of 1 s per letter (using Adobe After Effects), respecting the standard order and direction of strokes, and participants performed tasks according to their learning condition. The only difference between letter blocks and word blocks was whether one letter appeared (with audio of both its name and phoneme—as in "K" = "kay," "/k/") or two to three letters appeared as a word (along with the pronunciation of the word).

The letter blocks were used to teach the letters' names and sounds. The Arabic alphabet consists of 28 unique letter identities, each of which has between one and four distinct shapes (allographs). Twenty letter identities—17 consonants and 3 vowels—were selected for training (see Section S1 in the Supplemental Material available online). Fourteen of the letters have two allographs and six have only one,[1] for a total of 34 letter shapes. Four different fonts (Adobe Arabic, Nadeem, Myriad, and Farisi; Fig. 3) were used for the training stimuli. The font Adobe Arabic was used in all other tasks. The rationale for including multiple fonts was (a) to expose participants in all three conditions to variable visual input, which has been found to be an important factor in learning to recognize letters (Li & James, 2016), and (b) to allow for matching the number of exemplars presented for each letter identity.

Word blocks were used to examine the effects of learning conditions beyond single-letter processing. Furthermore, training with words allowed participants to experience how allographs, despite their shape differences, corresponded to the same sounds within words (Fig. 2). Fifty-one words (one or two consonants with one vowel) were created; no definitions were given for the words, which included a mix of real words and pseudowords (see Section S1 in the Supplemental Material).

In each session, the first three blocks were letter blocks, consisting of 80 trials over 8 min; within these blocks, each letter identity was presented four times in random order: either two allographs in two fonts or one allograph in four fonts. In word blocks, each of the 51 words was presented once. Across each training session, each consonant was presented 16 times and each vowel 29 times in total.

Training protocols

Previous research (Bhide, 2018; James, 2010; Li & James, 2016; Longcamp et al., 2005, 2006, 2008; Naka, 1998; Zemlock et al., 2018) has raised the issue that learning conditions such as writing and typing involve complex processes and that it is therefore critical to determine which aspects of the complex learning experience are relevant to the questions at hand. This

Fig. 3. All stimuli used for the training sessions. The 20 letter identities are each presented with their allograph in the four fonts (from left to right: Adobe Arabic, Nadeem, Myriad, and Farisi).

includes incidental aspects, such as differences in time spent on task and the relative difficulty of the training tasks (see Bhide, 2018; James, 2010; Li & James, 2016; Longcamp et al., 2005), as well as more fundamental aspects, such as the visual feedback produced by motor output (see Kersey & James, 2013). We introduced five innovations to address these issues. First, we maximized the overall similarity across learning conditions by ensuring that all learning conditions involved exposure to the same stimuli for the same duration. Second, we specifically controlled for the possibility that dynamic displays could be the source of learning benefit for the handwriting condition by using the same set of dynamic displays of a letter or word on all the training trials (Fig. S1 in the Supplemental Material). Third, to specifically control for the possibility that simple exposure to variable visual letter forms could be the source of benefit for the handwriting condition, we included multiple allographs and fonts. The teaching of allographs is particularly novel and allowed us to evaluate the participants’ knowledge of allographs, which is critical to learning real written languages (i.e., readers must learn to map multiple exemplars onto the same letter identity, including extremely different visual forms such as

"G"/"g"). Fourth, to reduce the possibility that data analysis could be affected by different levels of achievement on the learning tasks across conditions, we administered posttraining assessments only after all conditions reached a common learning criterion on the letter-recognition task. Fifth, we trained all participants not only on the letters' names but also on their sounds (phonemes), which has not typically been done in previous studies (but see Kiefer et al., 2015). This is critical for testing generalization to reading and writing words, which requires knowledge of phoneme-grapheme mappings and not just letter names.

Participants were asked to learn the letters' shapes, names, and sounds, and they were told that they would be tested on this knowledge. Prior to beginning training, participants in all conditions were told the same basic facts about Arabic (e.g., that it is cursive and written from right to left), and the name, shape, and sound of each letter's allograph was previewed (see Sections S2 and S3 in the Supplemental Material). During each training trial, participants performed a task according to their assigned learning condition: typing, visual, or writing. In both letter and word blocks, across all learning conditions, a tone played at the end of each trial indicating a 1-s intertrial interval with a blank screen; both response time (RT) and accuracy were recorded on each trial. The task instructions for each learning condition are summarized as follows (see Section S2 in the Supplemental Material).

Typing condition. In the typing condition, the task was to find the stimulus letter or letters on a keyboard and press the corresponding key or keys (in the correct sequence for words) as quickly and accurately as possible within the time limit. Opaque labels with Arabic letters were adhered to a regular U.S. English keyboard. The same keyboard layout was used for all participants and all sessions and was generated by randomizing the 34 letter shapes into three rows, with each allograph assigned to its own key; we were careful to avoid placing highly similar letters (Wiley et al., 2016) adjacent to each other. Feedback was provided in the form of tones indicating whether the response was correct or incorrect or whether no response was recorded at all.

Visual condition. In the visual condition, participants performed a visual detection task. The dynamic display of the target letter or word was identical to the displays in the typing and writing conditions but persisted for only 1 s before disappearing. It was followed by a 500-ms blank screen, a 66-ms fixation cross, and a 1,000-ms probe. The probe was either a nonalphabetic symbol, a symbol string (e.g., ?% #), or the target Arabic stimulus in a smaller font size. After the probe, the target returned for the remainder of the trial; thus, total trial length was equated across conditions. The task was to press a key to indicate whether the probe matched the identity of the target. Feedback was provided in the form of tones indicating whether the response was correct or incorrect or whether no response was recorded at all.

Writing condition. In the writing condition, the procedure was identical to that in the typing condition except that participants were instructed to copy the stimuli, writing the letters or words with a pen on ruled paper placed atop an Intuos electronic tablet (Wacom, Kazo, Saitama, Japan) connected to E-Prime software (Version 2.0; Schneider, Eschman, & Zuccolotto, 2012). The stroke patterns to be used were not explicitly prescribed but could be inferred from the presentation of the dynamic visual stimulus. No formal feedback was given, but participants were able to view their own handwritten exemplars and compare them with the stimuli on the screen.

Training sessions took place twice a week, with at least one day between sessions. To provide longitudinal measures of learning, we administered letter-recognition and letter-naming tasks at the end of each training session (see Behavioral Assessments).

Stopping criteria. Two criteria had to be fulfilled on the letter-recognition task administered at the end of each session: greater than 90% accuracy and a 25% reduction in RT relative to performance on the first session. These criteria had to be maintained across two consecutive sessions; otherwise, training continued for a maximum of six sessions. These criteria served to achieve general comparability in performance levels across the three learning conditions as well as to promote stability of the learned representations.

Behavioral assessments

We carried out extensive assessments that included tasks and/or stimuli that differed from those used during training to evaluate the extent to which learning experiences were generalized. The pretraining session consisted of the same/different judgment task. Training began on average 15 days later. At the end of each training session, the letter-recognition and letter-naming tasks were administered. The letter-recognition task was used to determine when stopping criteria were reached and to provide a longitudinal measure of the learning trajectory. These tasks are the closest match to actual classroom assessments that typically require students to associate letter shapes and names. Posttraining assessments (2–5 days after reaching training criteria) consisted of letter recognition, letter naming, letter writing, word spelling, and word reading. These five tasks were administered to all participants identically across all three conditions (see Table 1; for details, see Section S4 in the Supplemental Material). Feedback was provided only for letter recognition and only during training sessions.

Letter recognition. In the letter-recognition task, letter names were presented auditorily, and participants clicked on the corresponding letter shape from four choices displayed on the monitor. Both accuracy and RT were assessed.

Letter naming. In the letter-naming task, each of the 34 letter shapes were presented on the monitor, and participants were instructed to speak the letters' names into a microphone, which recorded the voice-onset time. Responses were scored as correct as long as they unambiguously referred to the correct letter. Both accuracy and RT for trials with correct responses were assessed. Specifically, responses were scored as (a) correct, (b) incorrect (i.e., the wrong letter name was produced), (c) nonletters (i.e., a name not matching any of the Arabic letters or otherwise unintelligible), or (d) no response. Interrater reliability in scoring of correct and incorrect responses was obtained for a set of nine participants (three per learning condition), each scored by two researchers. Scoring agreement was near perfect (Cohen's $\kappa = 0.98$).

Letter writing. In the letter-writing task, each of the 20 letter names were presented auditorily, and participants were asked to write the shapes from memory. Participants were reminded that most of the letters had two shapes, and they were prompted to produce both if they could remember them. There were, therefore, a maximum possible 34 points (one per letter shape). Participants wrote the letters with an ink pen on a sheet of paper placed atop an electronic tablet (Wacom Intuos pen tablet). Responses were scored as either correct or incorrect, and incorrect responses were further categorized into four types of errors: (a) mirror reversed but otherwise correct, (b) nonletters (i.e., deletion or addition of strokes that created a nonletter), (c) incorrect (i.e., the shape corresponded to a different letter), or (d) no response. Accuracy was evaluated by comparing correct with incorrect responses (i.e., collapsing across all four types of errors). Interrater reliability in scoring correct and incorrect responses was obtained for a set of nine participants (three per learning condition), each scored by two researchers. Scoring agreement was high (Cohen's $\kappa = 0.82$).

Word spelling. In the word-spelling task, 20 three- to six-letter words were presented auditorily, and participants were asked to write them on the electronic tablet. The stimuli included seven familiar words (i.e., used during training word blocks) and 13 novel words. Letters were scored as correct if the intended letter was unambiguous (e.g., there was no penalty for mirror-reversed letters or wrong allographs).

Word reading. In the reading-words task, 20 two- to six-letter words were presented on the monitor, and participants were asked to attempt to sound them out. The stimuli included seven words familiar from training; all of the words were different from those used in the spelling-to-dictation task. Accuracy was assessed by scoring the percentage of letters read correctly (e.g., reading "cat" as "at" was scored as 67%).

Same/different judgment. This task was used to provide behavioral evidence of how letter perception is affected by the type of learning experience and to reveal the multiple learned representations of letters, both sensorimotor and amodal (Fig. 4). On each trial, participants pressed a key to indicate whether a pair of letters was physically the same or different. The basic assumption underlying the task is that longer RTs to decide that two letters are different reflects greater representational similarity between the letters. For example, slow "different" responses for the two allographs of the letter "kaf" (ك and كـ) may be attributable to their shared identity. Importantly, the data were analyzed with simultaneous multiple regression to test for unique variance explained by different types of letter representations (see Analysis of Same/Different Judgment Task).

Data analysis

General analysis approach. To address the three questions of interest identified above, we analyzed the data with linear mixed-effects models (LMEMs; Baayen et al., 2008) using the R package *lme4* (Version 1.1-25; Bates et al., 2015). First, growth-curve analyses were used to determine whether the trajectory of letter learning (letter recognition and letter naming) differed across learning conditions. Second, analyses of posttraining tasks (e.g., word spelling) were used to determine whether training generalized to untrained tasks and whether any generalizations differed among conditions. Third, the analysis of the same/different judgment task specifically addressed the nature of the representational similarity underlying letter perception and whether the dimensions of similarity (e.g., visual, motor) differed among learning conditions.

Analysis of learning trajectories and posttraining tasks. RT data were log transformed (Gaussian family), and accuracy data were modeled with logistic regression (binomial family). In all models, learning condition was

Type of Representation	High Similarity		Low Similarity	
Pixel Overlap	ك	شـ	جـ	ش
Visual-Feature Overlap	ب	يـ	شـ	كـ
Motoric Similarity	ق	ش	كـ	جـ
Phonological Similarity	كـ	ق	ي	ق
Symbolic Identity	ك	كـ	بـ	يـ

Fig. 4. Examples of high- and low-similarity pairs of Arabic letters, along five different representational dimensions.

included as the primary regressor of interest (simple-coded with writing as the reference level), with control regressors appropriate to the task (e.g., trial order, word length; for a full description of regressors and covariates, see Sections S5–S7 in the Supplemental Material). Bootstrapped 95% confidence intervals (CIs) around the estimated β coefficients were provided by the bootMer function in the *lme4* package. All reported R^2 measures are pseudo-R^2s (specifically appropriate for generalized LMEMs), as provided by the R package *MuMIn* (Version 0.12; Bartoń, 2009). Random effects both by participant and by stimulus were included (for full model specifications, see Tables S1 and S2 in the Supplemental Material).

Analysis of same/different judgment task. The dependent variable, RT to correct responses on *different* trials, was analyzed via generalized LMEM as a γ distribution (identity link; Lo & Andrews, 2015) with predictors indexing letter-pair similarity along different dimensions (e.g., visual similarity, motoric similarity).The pre- and posttraining Arabic-letter data sets were analyzed separately.[2] The pretraining data were analyzed (a) to determine which letter representations influenced performance for naive observers and (b) to verify that there were no significant differences across the three learning conditions prior to training. The posttraining results were analyzed in the same manner to determine whether, after reaching criteria on letter recognition, there were significant differences among learning conditions in terms of the types of representations that influenced performance on the task.

In addition to the predictor of condition, the primary variables of interest (fixed effects) were five predictors of representational similarity (Fig. 4): pixel overlap (low-level visual similarity), visual-feature overlap (proportion of shared visual features, such as lines and curves), motoric similarity (proportion of shared strokes, such as downward or clockwise), phonological representation (letter-name similarity), and symbolic letter identity (amodal representation; i.e., whether or not the letter pairs were allographs and shared symbolic letter identities)—as well as the interaction terms between these five predictors and the condition predictor. Random slopes for these effects by participant, as well as random intercepts by participant and by item, were also included (for full model specifications, see Tables S3–S7 in the Supplemental Material).

Results

The results are organized according to the three key questions of the investigation (see the Introduction) and summarized below (for detailed results, see Tables S1–S7 in the Supplemental Material).

Effects of learning conditions on learning trajectories

On average, participants in the writing condition required the least training (M = 3.67 sessions, SE = 0.36), followed by participants in the typing condition (M = 3.92, SE = 0.38) and the visual condition (M = 4.25, SE = 0.37).

Letter recognition. For accuracy in the letter-recognition task (Fig. 5, left), there were significant differences between the visual and writing conditions in both the linear ($p < .001$) and quadratic ($p < .05$) trends, indicating a faster rate of improvement in the writing condition. There were no significant differences between the typing and writing conditions. The total model R^2 was 59%. There were no significant differences among learning conditions in terms of changes in RT.

Letter naming. For accuracy in the letter-naming task (Fig. 5, right) there were significant differences between the typing and the writing conditions in the linear trend ($p < .05$) and between the visual and writing conditions in both the linear ($p < .001$) and cubic ($p < .05$) trends. These results indicate generally faster improvement for

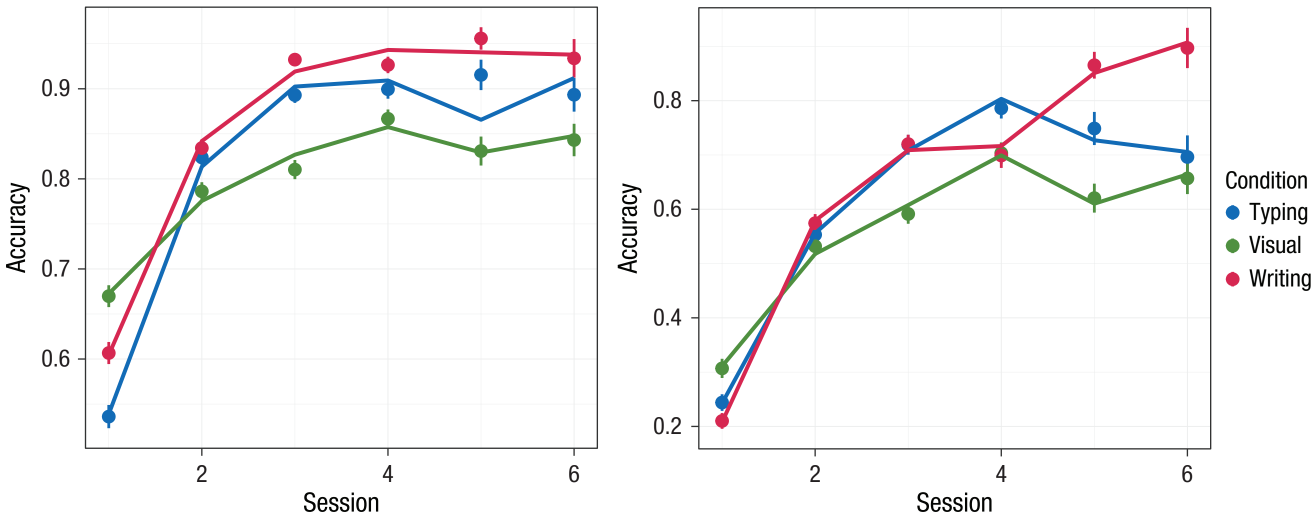

Fig. 5. Proportion of correct responses in the letter-recognition task (left) and the letter-naming task (right) across the training sessions, separately for the typing, visual, and writing learning conditions. Circles represent raw data (error bars reflect standard errors of the mean), and lines reflect model fits.

the writing condition compared with the other two conditions. The total model R^2 was 69%. There were no significant differences among conditions in terms of change in RT.

Summary of the effects of learning condition on learning trajectories. The growth-curve analyses confirmed that, overall, participants improved not only in their ability to recognize the letters but also in terms of their ability to name the letters (despite never receiving feedback on that task nor it being a part of the stopping criteria for training). There were no significant differences among conditions with regard to changes in RT. For accuracy, however, the writing condition was significantly different from the visual condition on letter recognition and from both the visual and typing conditions on letter naming. These differences indicate faster learning in the writing condition (Fig. 5).

Generalization of learning

Analysis of posttraining performance on letter recognition confirmed that, on average, the three learning conditions were equivalent after training was completed, despite differences in the amount of training required to reach the stopping criteria: For the typing, visual, and writing conditions, mean RT on trials with correct responses was 2,188 ms (SD = 559), 1,945 ms (SD = 430), and 1,847 ms (SD = 250), respectively, and mean accuracy was 96.7% (SD = 3.5%), 94.9% (SD = 5.6%), and 97.3% (SD = 1.2%), respectively.[3] None of these differences were significant (ps > .1).

The results for the four posttraining generalization tasks are presented in Figure 6 (for detailed results, see Tables S1–S2 in the Supplemental Material). For all of these tasks, number of sessions and days between the training and posttraining sessions were included as covariates to control for the possibility that differences in performance were the result of unequal amounts of exposure to the stimuli or longer delays before returning to the lab for the posttraining evaluation session. Analyses of the letter-naming and letter-writing tasks included the additional covariates of recognition RT and recognition accuracy (computed from the posttraining letter-recognition task) to control for item-specific differences in the participants' ability to recognize the letters that they were asked to name or write. For each model, the total model R^2 reported is for the marginal plus the conditional effects (i.e., the combination of both the fixed and random effects).

Figure 6 presents the performance for each condition with violin plots, which show the distribution of individual scores within the groups (shaded areas) overlaid with box-and-whisker plots showing the medians and first and third quartiles. From this figure it can be seen that the median performance was best for all generalization measures in the writing condition.

Letter naming. Mean RT on letter-naming trials with correct responses was 1,545 ms (SD = 337), 1,393 ms (SD = 368), and 1,308 ms (SD = 235) for the typing, visual, and writing conditions, respectively. There was a significant difference between the typing and writing conditions (p < .05); RTs in the writing condition were faster. RTs in

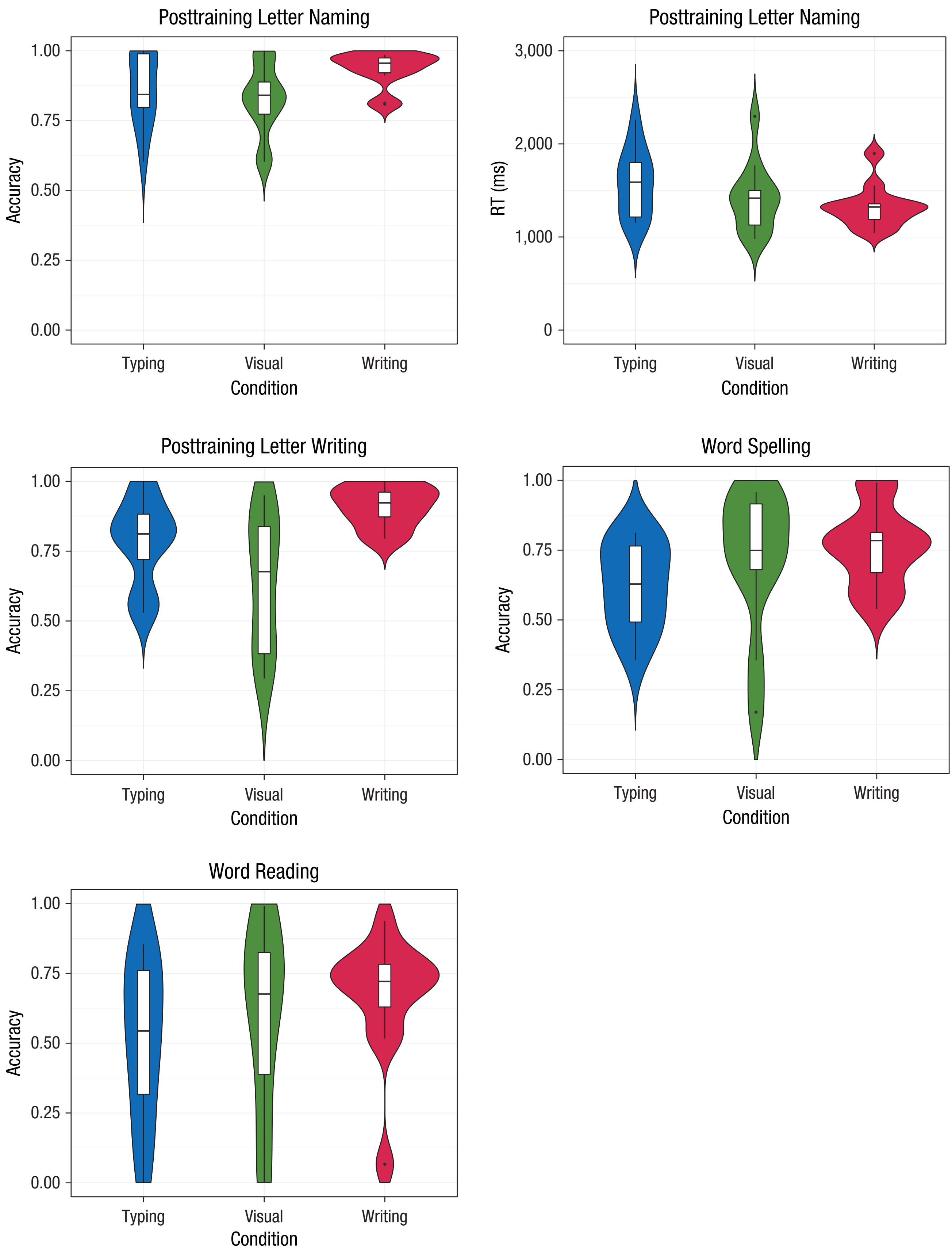

Fig. 6. Violin plots depicting performance on the four posttraining generalization tasks, separately for the typing, visual, and writing learning conditions. The shaded areas indicate the distribution of individual scores. In the box-and-whisker plots, the horizontal lines indicate medians; the areas above and below the medians indicate the first and third quartiles, respectively; and the whiskers extend 1.5 times the interquartile range.

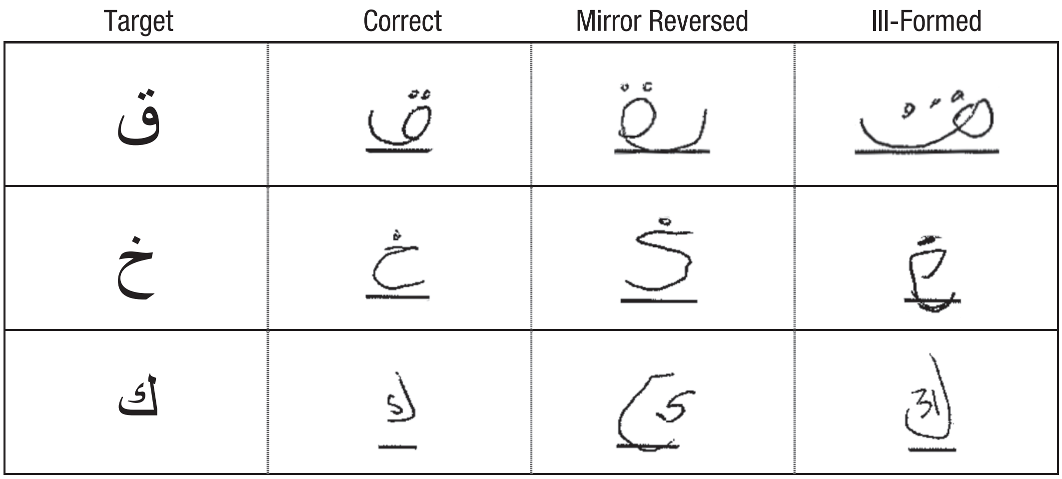

Fig. 7. Example of correct responses, mirror-reversed errors, and ill-formed errors from the posttraining letter-writing task. Examples are shown for three different letters: ق ("qaf"), خ ("kha"), and ك ("kaf"). The examples are selected from the responses of 7 participants. The lines under the letters are the actual lines that appeared on the paper.

the visual and writing conditions were not significantly different from one another (p > .05). The total model R^2 was 43%.

Mean accuracy was 86.2% (SD = 13.3%), 82.8% (SD = 12.8%), and 93.4% (SD = 6.2%) for the typing, visual, and writing conditions, respectively. There was a significant difference between the visual and writing conditions (p < .05); participants were more accurate in the writing condition. Accuracy in the typing and writing conditions was not significantly different (p > .05). The total model R^2 was 47%.

Letter writing. Mean accuracy in the letter-writing task was 78.5% (SD = 15.3%), 64.5% (SD = 23.8%), and 91.0% (SD = 7.2%) for the typing, visual, and writing conditions, respectively. There was a nonsignificant trend of higher accuracy in the writing condition compared with the typing condition (p < .09) and a significant effect of higher accuracy in the writing condition compared with the visual condition (p < .001). The total model R^2 was 46%.

The differences among learning conditions were not driven by a single type of error (mirror reversals, ill-formed shapes, or failure to produce any recognizable response), as all three types of errors were most common in the visual condition (10.5% of all responses were mirror reversals, 9.2% of all responses were ill-formed, and there was no response on 15.7% of trials), followed by the typing condition (5.5% mirror reversals, 6.5% ill-formed, and 9.5% no response). The fewest errors of each type were found in the writing condition (1.7% mirror reversals, 3.5% ill-formed, and 3.7% no response). Example errors are reported in Figure 7.

Word spelling. For the word-spelling task, mean accuracy was 62.3% (SD = 15.4%), 72.0% (SD = 25.2%), and 76.3% (SD = 14.5%) for the typing, visual, and writing conditions, respectively. There was no effect of familiarity (previously trained words vs. novel words), p > .1. Accuracy in word spelling was significantly higher in the writing than in the typing condition (p < .05), and there was no significant difference in accuracy between the visual and writing conditions. The total model R^2 was 38%.

Word reading. In the word-reading task, mean accuracy was 50.8% (SD = 28.5%), 59.2% (SD = 33.4%), and 66.6% (SD = 22.0%) for the typing, visual, and writing conditions, respectively. There was no effect of familiarity (previously trained words vs. novel words), p > .1. Participants in the writing condition were not significantly more accurate than those in either the typing or the visual conditions ($p \approx$ .18 and .47, respectively). The total model R^2 was 57%.

Interim discussion

The fact that there were no significant differences among learning conditions in letter-recognition ability at the posttraining session confirms that each group reached the criteria, although individuals in the writing condition learned more quickly and required fewer training trials to do so (see Fig. 5). Across all five generalization measures, participants in the writing condition were numerically the highest performing (Table 3). Many of these differences were statistically significant even when we controlled for item-specific individual differences in letter recognition and differences in the number of training

Table 3. Performance on Each of the Generalization Measures, According to Learning Condition

Measure	Typing	Visual	Writing
Letter-naming response time	3*	2	1
Letter-naming accuracy	2	3*	1
Letter-writing accuracy	2	3*	1
Word spelling	3*	2	1
Word reading	3	2	1

Note: The values indicate the ranking of the average performance for each learning condition. An asterisk indicates a statistically significant difference in the comparison with the writing condition ($p < .05$).

sessions and calendar days between the last training session and the posttraining session.

In recognition of the limitations of null-hypothesis significance testing (see Cumming, 2013) and arbitrary thresholds for significance, we conducted a Monte Carlo analysis to determine the probability that participants in one of the three learning conditions would perform—by chance—the best on all five measures. The Monte Carlo analysis was conducted by randomly permuting the condition labels across participants, computing the average performance on each of the five measures for each relabeled condition, and determining the rankings for the relabeled writing condition for each permutation. Running 10,000 iterations of this analysis indicated that the probability that a random group of 12 participants would be ranked as the best-performing group on all measures was only 0.0329 (see Fig. S2 in the Supplemental Material).

The results represent compelling evidence that letter learning in the writing condition led to the best performance on a range of tasks beyond letter recognition or writing and that these learning benefits could not be explained by differences in the amount of stimulus exposure or greater familiarity with the letter shapes.

The nature of the learned representations: results of the same/different judgment task

As discussed earlier, the same/different judgment task allowed us to evaluate the nature of underlying and learned letter representations by examining the influence on RTs of five types of similarity between the letter-pair stimuli on which same/different judgments were made (Fig. 4). The rationale is that RTs for same/different judgments will be slowed on the basis of the strength of the underlying representational types.

Pretraining time point. At the pretraining session, the only measure of representational similarity[4] that significantly predicted performance on deciding whether two letters were the same or different was pixel overlap. Participants were significantly slower to respond to more visually similar letter pairs ($p < .001$). This effect is depicted in Figure 8, which shows the estimated RT difference between pairs with low similarity (0.2-pixel overlap) and pairs with high similarity (0.8-pixel overlap); p values for the difference scores were obtained using the R package *emmeans* (Version 1.1.4; Lenth, 2019). The difference between high and low pixel overlap was estimated at 99.8 ms, 73.4 ms, and 99.7 ms for the typing, visual, and writing conditions, respectively (all ps $< .05$), and there were no significant interactions with condition. This shows that naive viewers of Arabic letters are sensitive to the pixel-based similarity between pairs of letters with which they are not familiar.

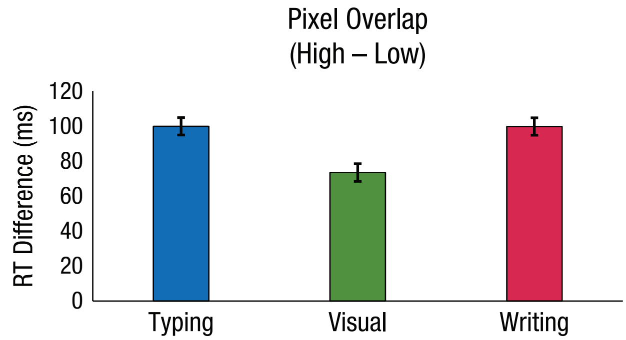

Fig. 8. Mean response time (RT) difference on trials with correct responses to stimuli with high and low pixel overlap at the pretraining time point. Results are shown separately for the typing, visual, and writing learning conditions. RT differences were calculated by subtracting responses to stimuli with low pixel overlap from responses to stimuli with high pixel overlap. Error bars reflect standard errors as generated by the R package *emmeans*.

Posttraining time point. At the posttraining session, as with the pretraining session, there was still a significant effect of pixel overlap (slower RT on more similar letters, $p < .05$) and no significant interactions with condition. The effects plots showing the estimated RT difference between pairs with low and high visual-feature similarity (0.2 vs. 0.8 feature overlap), low and high motoric similarity (0.2 vs. 0.8 shared motor bistrokes), and symbolic letter identity (different vs. same identities) are depicted in Figure 9.

The effect of visual-feature similarity was significantly stronger for the typing than the writing condition ($p < .05$; see Fig. 9, top)—the estimated difference between high and low similarity was 54.3 ms ($p < .05$), 26.1 ms ($p > .1$), and 10.7 ms ($p > .1$) for the typing, visual, and writing conditions, respectively. The effect of motoric similarity was significantly stronger for the writing condition compared with both the typing and visual conditions ($p < .05$ and $p < .01$, respectively; see Fig. 9, middle)—an estimated difference for high-low similarity

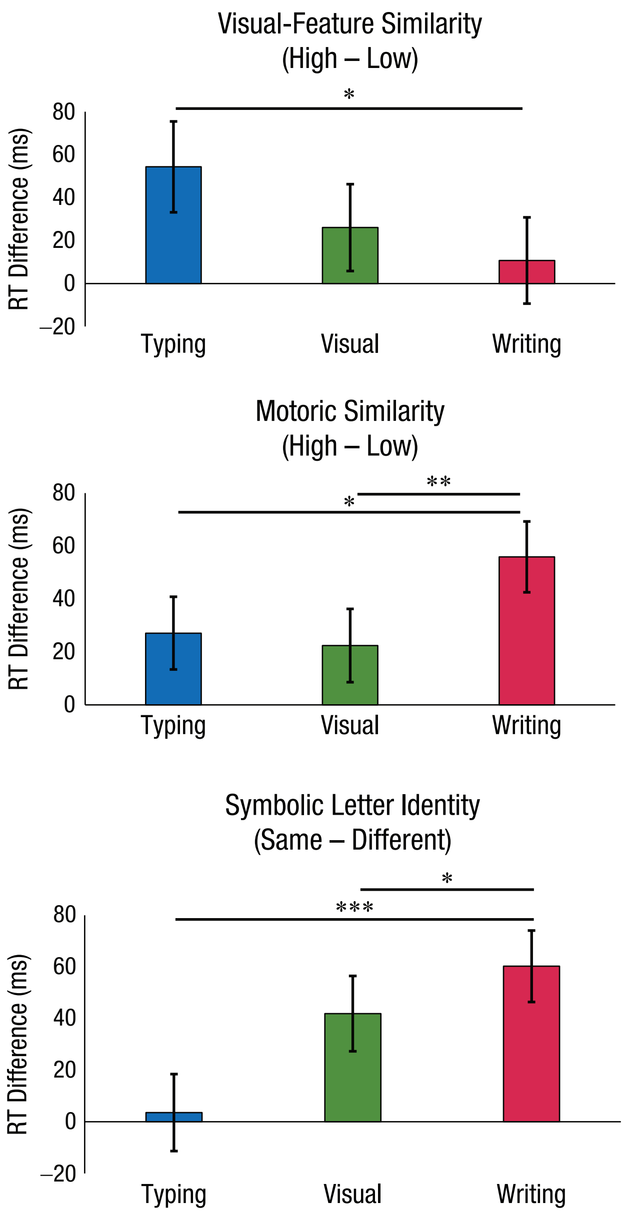

Fig. 9. Mean response time (RT) difference on trials with correct responses to stimuli with high and low visual-feature similarity (top), high and low motoric similarity (middle), and the same and different symbolic letter identity (bottom) at the posttraining time point. Results are shown separately for the typing, visual, and writing learning conditions. RT differences were calculated by subtracting responses to stimuli with low similarity or different symbolic letter identity from responses to stimuli with high similarity or the same symbolic letter identity. Error bars reflect standard errors as generated by the R package *emmeans*. Asterisks indicate significant differences between learning conditions (*p < .05, **p < .01, ***p < .001).

of 27.1 ms ($p < .05$), 22.4 ms ($p > .1$), and 55.9 ms ($p < .001$) for the typing, visual, and writing conditions, respectively. The effect of symbolic letter identity was significantly stronger for the writing condition compared with both the typing ($p < .001$) and visual conditions ($p < .05$; Fig. 9, bottom)—an estimated difference between same- and different-identity pairs of 3.5 ms ($p > .1$), 41.9 ms ($p < .01$), and 60.3 ms ($p < .001$) for the typing, visual, and writing conditions, respectively.

Comparing the pre- and posttraining sessions directly, we found that the only significant changes were for the writing condition and only for motoric similarity ($p < .01$) and symbolic letter identity ($p < .05$). For all other measures and conditions, no changes were significant (ps > .1; see Tables S5–S7 in the Supplemental Material).

Summary regarding learned representations (same/different judgment task). Not surprisingly, before completing any training, participants across all learning conditions were naive and showed only significant effects of pixel overlap. After participants completed training, however, all four types of representational similarity were found to predict RTs for one or more groups. Visual-feature similarity was significant only for the typing group, which may indicate a shift from reliance on low-level visual features to more abstract ones. Indeed, pixel overlap was no longer significant for the typing condition at the posttraining session, although the difference from the pre- to posttraining sessions was not significant.

The effect of motoric similarity was driven by the writing condition and was weakest (in fact, nonsignificant) in the visual condition (mirroring the results of the letter-writing task). More striking is that the effect of symbolic letter identity was likewise driven by the writing condition; discrimination of same-identity pairs was an estimated 60 ms slower than discrimination of different-identity pairs. The symbolic-letter-identity effect was also present in the visual condition but was significantly smaller (42 ms), and it was virtually absent in the typing condition (4 ms; n.s.). In the General Discussion, we return to these results and in particular to their implications for understanding the effects of handwriting training.

General Discussion

What are the benefits of handwriting training for letter learning and literacy?

We examined this question in an investigation that examined the benefits of handwriting training for learning novel letters, comparing handwriting with typing and visual training. The findings show that the benefits of handwriting training not only included a faster learning trajectory but also extended beyond the tasks on which participants were trained (letter recognition and writing) to untrained tasks, such as letter naming and word reading. Moreover, a Monte Carlo analysis revealed that the finding that the writing condition was numerically superior to other learning conditions on all five generalization measures (Table 3) was highly unlikely to have occurred by chance ($p \approx .0329$). Thus, overall, the results revealed that handwriting practice fundamentally and positively affects written language learning, both for letter and word comprehension and production.[5]

The representational consequences of handwriting practice

That handwriting has profound effects on letter processing was further supported by the findings from a perceptual same/different judgment task. These findings showed that the visual perception of the learned letters of individuals who learned by writing was influenced not only by their knowledge of how to write the letters but also by their knowledge of the symbolic identities of the letters. Critically, these perceptual-learning effects were strongest in the writing condition. The finding that the effect of motor similarity on this task was strongest in the writing condition is important when one considers that it need not have been the case that knowledge of the motor plans for writing would play any role in a visual same/different judgment task. In fact, although some researchers have found evidence of motor similarity affecting visual-perception tasks with letters, either behaviorally or in neural-activity patterns (e.g., Nashaat et al., 2016; Schubert, Reilhac, & McCloskey, 2018; Wiley & Rapp, 2019; Wiley et al., 2016), others have not (e.g., Bi et al., 2009; Rothlein & Rapp, 2017; Zhai & Fischer-Baum, 2019). To the best of our knowledge, this is the first study to directly associate handwriting experience (or the lack thereof) with the effect of motor similarity on letter perception, supporting the proposition that motor similarity is in fact an aspect of motor knowledge and not merely correlated with some other type of representation. With regard to the significantly stronger effect of symbolic letter identity for the writing condition, a similar conclusion applies—Although many researchers have argued for this type of abstract representation (e.g., Caramazza & Hillis, 1990; Lupyan et al., 2010; Rothlein & Rapp, 2014; Wiley et al., 2016), the existence of such abstract representations, for letters or other objects, has been questioned (e.g., Barsalou, 2008; Chen & Proctor, 2012). Most significantly, this study is the first to demonstrate an association between symbolic-letter-identity learning and writing experience.

The implication of these findings is that handwriting practice most strongly supported the learning of the multidimensional representations of letters that have been documented in expert readers—visual, motor, and symbolic representations (e.g., Kinoshita et al., 2019; Rothlein & Rapp, 2014, 2017; Wiley & Rapp, 2019; Wiley et al., 2016). This interpretation is consistent with that of Cao and colleagues (2013), who stated that the motor component learned through handwriting experience is "highly interactive with the other components and may become especially helpful for perception when other components are impaired or weak" (p. 1680). Our findings are also consistent with evidence from neuroimaging studies reporting that patterns of neural activity during letter viewing in children are similar to those observed in adults only when the children are viewing letters that they had experience writing (e.g., James, 2010).

Thus, it appears that participants in the current study used their representational knowledge gained from writing letters to strengthen their performance across multiple tasks.

What is the basis of the observed differences in outcomes among the learning conditions?

Our results indicate that neither variable visual input nor dynamic information, two indirect effects of handwriting practice, is sufficient alone to explain the handwriting benefits. However, the possibility remains that the handwriting benefits arise because of visual feedback that is generated by self-production or, at the very least, that the benefits of visual feedback are not equivalent when it is externally provided rather than internally generated. This would be consistent with the finding of Kersey and James (2013) that brain activity in the sensorimotor network was observed only after children were trained to write letters themselves, not after they passively viewed an instructor writing letters. It is important to recognize that writing and typing are both complex processes and that further work is still needed to identify which aspects of these tasks are responsible for the various effects that are reported in this and previous studies.

With regard to symbolic letter identities, one possibility is that they serve as hubs for the cross-modal letter processing (Fig. 10) used in reading and spelling, either supported by multimodal association areas, as proposed by several researchers (e.g., Binder, 2016; Buckner & Krienen, 2013), and/or by brain regions supporting amodal orthographic processing (Dehaene et al., 2004; Rothlein & Rapp, 2014). The writing-training paradigm used here, which required learning allographs and associating visual, motoric, and phonological representations, may have especially facilitated learning symbolic letter identities. This may be because the writing training required participants to learn all pairwise associations between different modalities. The near-zero effect of

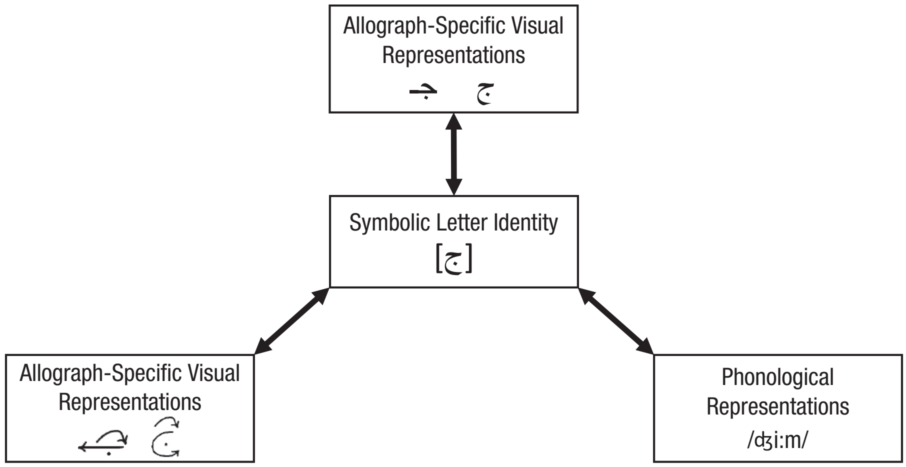

Fig. 10. Cognitive architecture of letter representations. Modality-specific representations (visual, motor, and phonological) are associated with one another through their common amodal representation: symbolic letter identity.

symbolic letter identity on the same/different judgment task for the typing condition may have been influenced by the fact that the keyboard, unlike standard keyboards, was arranged with a unique key for each allograph. What is certainly true is that the typing and visual training only required visuo-phonological associations, not visuomotor or phonological-motor associations. In this way, handwriting (compared with the other learning conditions) naturally promotes additional experience with transcoding inputs from one modality into another. Pattamadilok and colleagues (2016) suggested that activation in the dorsal premotor cortex, which is associated with sensorimotor processing of letters, "goes beyond a simple coactivation between the motor and visual regions [and] suggests that the contribution of the [dorsal premotor cortex] to reading reflects shared cognitive processes in writing and reading rather than an evocation of writing-related motor representations during reading" (p. 1541). To the extent that this shared cognitive process between writing and reading includes symbolic letter identities, their conclusions would be supported by evidence from the current study.

The evidence in support of symbolic-letter-identity learning is also relevant to the broader debate between embodied-cognition and abstractionist views. A strong embodied-cognition position holds that learning (including letter learning; Tenpenny, 1995) exclusively involves laying down sensory and motor memories that are reactivated when needed for task performance (see Barsalou, 2008). In contrast, abstractionist views propose that at least some aspects of human knowledge, such as letter knowledge, involve abstract, amodal representations that cannot be reduced to sensorimotor memories (e.g., Leshinskaya & Caramazza, 2016; Mahon, 2015; Mahon & Hickok, 2016). Our findings favor the abstractionist view and add to various existing lines of evidence of symbolic letter identities from cognitive psychology (e.g., Schubert, Gawthrop, & Kinoshita, 2018; Wiley et al., 2016), neuropsychology, and neuroimaging (e.g., Petit et al., 2006; Rothlein & Rapp, 2014). Additionally, the finding that symbolic-letter-identity learning was strongest in the context of writing experience would seem to be an especially timely reminder that evidence of motor learning should not be automatically interpreted as favoring an embodiment position (but see Barsalou, 2008, for a review of a range of different positions on this question).

Implications for educational practices and directions for future research

The clear evidence presented here that handwriting training during letter learning strengthens reading and spelling skills has implications for best practices in education—although it must be emphasized that these results were observed in adults learning a second orthography and not in children. Future research should also investigate whether the findings reported here generalize to older adults or to individuals with lower levels of education or specific learning disabilities. Nonetheless, at a minimum, it is a warning against prematurely abandoning or significantly reducing handwriting education. The finding that the benefits of writing practice extend beyond penmanship to both letter and word comprehension and production indicates that handwriting can be a productive use of learning time. However, there are certainly a great number of implementations of handwriting practice that could be adopted, and they are likely to vary in their effectiveness.

This research does not address the pressing issue of understanding the optimal characteristics of the writing practice needed to maximize literacy learning. Furthermore, if the hypothesis is correct that the benefits of handwriting arise at least in part from strengthening the amodal symbolic-letter-identity representations used for mapping between different letter representations, then there should be other letter-learning conditions that also serve to develop and strengthen symbolic letter identities. In other words, although writing may be a natural way of learning to link the multiple modalities of letter representation, there may be other beneficial approaches as well. For example, it is possible that more extensive typing training with a standard keyboard arrangement could improve performance for that condition. Alternatively, training conditions that require participants to learn the dynamic visual information (even without self-production through handwriting) or that provide more equivalent visual feedback across conditions could be used to further identify the critical elements of the writing experience that produce the reported behavioral benefits. Despite the many remaining questions, the findings of this investigation provide a valuable foundation on which to continue to build our understanding of the role of handwriting in the development of literacy.

Transparency

Action Editor: Barbara Knowlton
Editor: Patricia J. Bauer
Author Contributions
Both of the authors contributed to the study concept and design. Testing, data collection, and data analysis were performed by R. W. Wiley. Both of the authors interpreted the results, drafted the manuscript, and approved the final version of the manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by a Distinguished Science of Learning Fellowship to R. W. Wiley from The Johns Hopkins University Science of Learning Institute.
Open Practices
Data and materials for this study have not been made publicly available, and the design and analysis plan were not preregistered.

ORCID iD

Robert W. Wiley https://orcid.org/0000-0003-1853-8063

Acknowledgments

We acknowledge Delaney Ubellacker and Dane Morgan for their contributions to this project.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797621993111

Notes

1. Nonligating letters, those that do not connect with the following letters, have only one allograph. There are eight such letters in Arabic, six of which were included in this study (see Fig. 3).
2. LMEM analysis testing for significant differences between performance in the pretraining and posttraining sessions (i.e., three-way interactions among time point, learning condition, and types of representational similarity) had intolerably high multicollinearity. As an alternative, pretraining versus posttraining models with two-way interactions (Time Point × Representational Similarity) were evaluated with separate LMEMs for each learning condition. For full details, see the Supplemental Material.
3. These values refer to the posttraining session, which is distinct from the training sessions depicted in the growth-curve analyses in Figure 6.
4. We determined that simultaneous regression of both symbolic-letter-identity and phonological-similarity (letter-name) regressors was not possible because of high multicollinearity (variance inflation factors 10.1 and 9.4, respectively—unacceptably high by any reasonable criterion). As a solution, separate LMEM analyses were run with either symbolic-letter-identity or phonological-similarity regressors. The phonological-similarity regressor was not significant either as a main effect or in interaction with condition, and model comparisons using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) favored the model with symbolic letter identity only compared with the model with phonological similarity only (AIC and BIC ≈ 2 points different in favor of the symbolic-letter-identity model). Therefore, there is no evidence supporting the hypothesis that it was phonological letter-name representations and not amodal symbolic-letter-identity representations that contributed to letter perception in the same/different judgment task.
5. Whether typing or visual training is superior was not evaluated in the study, but Table 3 indicates that their rankings were generally comparable.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bara, F., & Bonneton-Botté, N. (2018). Learning letters with the whole body: Visuomotor versus visual teaching in kindergarten. *Perceptual and Motor Skills*, *125*(1), 190–207.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Bartoń, K. (2009). *MuMIn: Multi-model inference* (R Package Version 0.12) [Computer software]. http://R-Forge.R-project.org/projects/mumin

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., Shimada, S., & Apel, K. (2006). Early development of language by hand: Composing, reading, listening, and speaking connections; three letter-writing modes; and fast mapping in spelling. *Developmental Neuropsychology*, *29*(1), 61–92.

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

Bi, Y., Han, Z., & Zhang, Y. (2009). Reading does not depend on writing, even in Chinese. *Neuropsychologia*, *47*(4), 1193–1199.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*(4), 1096–1108.

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*, *17*(12), 648–665.

Cao, F., Vu, M., Lung Chan, D. H., Lawrence, J. M., Harris, L. N., Guan, Q., Xu, Y., & 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.

Caramazza, A., & Hillis, A. E. (1990). Levels of representation, co-ordinate frames, and unilateral neglect. *Cognitive Neuropsychology*, *7*(5–6), 391–445.

Carreiras, M., Perea, M., & Mallouh, R. A. (2012). Priming of abstract letter representations may be universal: The case of Arabic. *Psychonomic Bulletin & Review*, *19*, 685–690. https://doi.org/10.3758/s13423-012-0260-8

Chen, J., & Proctor, R. W. (2012). Influence of category identity on letter matching: Conceptual penetration of visual processing or response competition? *Attention, Perception, & Psychophysics*, *74*(4), 716–729.

Cumming, G. (2013). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Deardorff, J. (2011, June 15). *The many health perks of good handwriting*. Chicago Tribune. https://www.chicagotribune.com/lifestyles/health/sc-health-0615-child-health-handwriti20110615-story.html

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J.-B., Le Bihan, D., & Cohen, L. (2004). Letter binding and invariant recognition of masked words. Behavioral and neuroimaging evidence. *Psychological Science*, *15*(5), 307–313.

Friedmann, N., & Gvion, A. (2005). Letter form as a constraint for errors in neglect dyslexia and letter position dyslexia. *Behavioural Neurology*, *16*(2–3), 145–158.

James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental Science*, *13*(2), 279–288.

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*, Article 567. https://doi.org/10.3389/fpsyg.2013.00567

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*, 136–146.

Kinoshita, S., Schubert, T., & Verdonschot, R. G. (2019). Allograph priming is based on abstract letter identities: Evidence from Japanese kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 183–190.

Konnikova, M. (2014, June 2). *What's lost as handwriting fades*. The New York Times. https://www.nytimes.com/2014/06/03/science/whats-lost-as-handwriting-fades.html

Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means* (R Package Version 1.1.4) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review*, *23*(4), 991–1001.

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.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, Article 1171. https://doi.org/10.3389/fpsyg.2015.01171

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. https://doi.org/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.

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.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, *21*(5), 682–691. https://doi.org/10.1177/0956797610366099

Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology*, *69*(2), 172–178.

Mahon, B. Z., & Hickok, G. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bulletin & Review*, *23*(4), 941–958.

Naka, M. (1998). Repeated writing facilitates children's memory for pseudocharacters and foreign letters. *Memory & Cognition*, *26*(4), 804–809.

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, USA*, *109*(50), 20762–20767.

Nashaat, N. H., Kilany, A., Hasan, H. M., Helal, S. I., Gebril, O. H., & Abdelraouf, E. R. (2016). Dysgraphia in Egyptian dyslexic children: Related abilities to writing performance in Arabic. *Journal of Innovations in Pharmaceutical and Biological Sciences*, *3*, 110–115.

Pattamadilok, C., Ponz, A., Planton, S., & Bonnard, M. (2016). Contribution of writing to reading: Dissociation between cognitive and motor process in the left dorsal premotor cortex: Contribution of writing to reading. *Human Brain Mapping*, *37*(4), 1531–1543.

Petit, J. E. A. P., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2006). On the time course of letter perception: A masked priming ERP investigation. *Psychonomic Bulletin & Review*, *13*(4), 674–681.

Rothlein, D., & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the multiple representations of letters. *NeuroImage*, *89*, 331–344.

Rothlein, D., & Rapp, B. (2017). The role of allograph representations in font-invariant letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1411–1429.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime 2 reference guide*. Psychology Software Tools.

Schubert, T., Gawthrop, R., & Kinoshita, S. (2018). Evidence for cross-script abstract identities in learners of Japanese kana. *Memory & Cognition*, *46*, 1010–1021. https://doi.org/10.3758/s13421-018-0818-4

Schubert, T., & McCloskey, M. (2013). Prelexical representations and processes in reading: Evidence from acquired dyslexia. *Cognitive Neuropsychology*, *30*(6), 360–395.

Schubert, T., Reilhac, C., & McCloskey, M. (2018). Knowledge about writing influences reading: Dynamic visual information about letter production facilitates letter identification. *Cortex*, *103*, 302–315. https://doi.org/10.1016/j.cortex.2018.03.020

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*(3), 339–363.

Treiman, R., & Kessler, B. (2004). The role of letter names in the acquisition of literacy. *Advances in Child Development and Behavior*, *31*, 105–135.

Treiman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development*, *69*(6), 1524–1540.

Wiley, R. W., & Rapp, B. (2019). From complexity to distinctiveness: The effect of expertise on letter perception. *Psychonomic Bulletin & Review*, *26*(3), 974–984.

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1186–1203.

Zemlock, D., Vinci-Booher, S. A., & James, K. H. (2018). Visual–motor symbol production facilitates letter recognition in young children. *Reading and Writing*, *31*(6), 1255–1271.

Zhai, M., & Fischer-Baum, S. (2019). Exploring the effects of knowledge of writing on reading Chinese characters in skilled readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 724–731. https://doi.org/10.1037/xlm0000604