



Enhancing literacy acquisition through spelling exercises: Comparing tasks for first graders

Cynthia Boggio, Marie-Line Bosse, Alexis Favre-Félix, Céline Pobel-Burtin,
Morgane Burnel, Maryse Bianco

► To cite this version:

Cynthia Boggio, Marie-Line Bosse, Alexis Favre-Félix, Céline Pobel-Burtin, Morgane Burnel, et al.. Enhancing literacy acquisition through spelling exercises: Comparing tasks for first graders. *Journal of Educational Research*, 2023, 116 (4), pp.187-197. 10.1080/00220671.2023.2233000 . hal-04237579

HAL Id: hal-04237579

<https://hal.science/hal-04237579>

Submitted on 12 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhancing literacy acquisition through spelling exercises: Comparing tasks for first graders

Cynthia Boggio^{a,b}, Marie-Line Bosse^a, Alexis Favre-Félix^c, Céline Pobel-Burtin^c, Morgane Burnel^{a,d} and Maryse Bianco^c

^a CNRS, LPNC, Univ. Grenoble Alpes, Grenoble, France; ^b Les Editions Hatier, Paris, France; ^c Univ. Grenoble Alpes, LaRAC, Grenoble, France; ^d Univ. Grenoble Alpes, MSH, Grenoble, France

ABSTRACT

Learning to spell is a difficult but essential task for children learning to read. Several spelling tasks involving different cognitive demands are used in classroom. To provide better guidance for teaching practices, a critical question is whether spelling performance of beginner readers is task dependent. The present study focuses on the performance of French first graders on two spelling tasks: picture to spelling and dictation to spelling. We hypothesize that spelling will be easier from a phonological entry (dictation) than from a semantic entry (picture). Sixty-three students spelled words in both tasks. Bayesian analyses showed moderate to strong evidence for the null hypothesis, suggesting that the dictation to spelling task is equivalent for first graders as the picture to spelling task. Exploratory analyses also suggest that first graders mobilize both the non-lexical and lexical routes for the two tasks. These results provide clues for spelling instruction in the first grades.

KEYWORDS

Spelling; education; first graders; dictation; written picture naming

ARTICLE HISTORY

Received 14 February 2023

Revised 5 June 2023

Accepted 30 June 2023

To cite this article:

Cynthia Boggio, Marie-Line Bosse, Alexis Favre-Félix, Céline Pobel-Burtin, Morgane Burnel & Maryse Bianco (2023): Enhancing literacy acquisition through spelling exercises: Comparing tasks for first graders, *The Journal of Educational Research*, DOI: 10.1080/00220671.2023.2233000

Introduction

Teaching reading and spelling is a major challenge for the schools. Academic success associated with social integration later in life, depends to a large extent on these initial acquisitions. Unfortunately, literacy acquisition has been shown to be particularly challenging in nontransparent written languages such as English or French (Fayol et al., 2008), where one sound unit may correspond to several graphical units and vice versa. Recently, almost 30% of French 2nd graders were identified as low reading achievers (Andreu et al., 2019) and the problem seems to persist among older students (Martin et al., 2017). As it is known that the majority of children who experience significant difficulties in the course of their schooling (e.g., school drop-out, low tests scores) have experienced past learning difficulties in the first grades (Ferrer et al., 2007; Landerl & Wimmer, 2008; Sprenger-Charolles et al., 2009), it is fundamental to determine the most effective methods for teaching literacy from the first grade.

Spelling words have been shown to be highly effective for improving literacy acquisition for young learners (e.g., Conrad et al., 2019; Ouellette & Sénéchal, 2008; Ouellette et al., 2017; Treiman, 1998; Uhry & Shepherd, 1993; Weiser, 2013) and struggling readers (e.g., Henbest & Apel, 2017). Including spelling instruction in early literacy instruction can be done using a variety of exercises, such as writing words after hearing them spoken aloud (dictation to spelling task) or under drawings or pictures (picture to spelling task). While on the surface these different tasks might appear similar in nature, recent studies on expert readers suggest that these two spelling tasks partially rely on different processing pathways (Bonin et al., 2015, 2019). Thus, they may also have different consequences on child performance and may be more or less suitable for early literacy acquisition. However, to our knowledge, this issue has never been investigated in the context of early literacy learning. In this article, for the first time, we will compare the performance of first graders on both dictation and picture to spelling tasks, to determine whether one task causes beginning spellers to write words more correctly than the other.

Literacy acquisition

In alphabetic languages, literacy is commonly described as involving two pathways (lexical and non-lexical) for both spelling (e.g., Houghton & Zorzi, 2003; Tainturier & Rapp, 2001), and reading (e.g., Coltheart et al., 2001). The non-lexical pathway consists of matching elements of language smaller than a word (e.g., phonemes) with their corresponding letters (e.g., graphemes). The lexical pathway involves the direct retrieval in memory of the orthographic form of the word to read or spell. Expert readers are known to rely more on lexical pathway to read or spell accurately and identify words quickly. The non-lexical pathway is considered to be essentially involved in reading and spelling new words or non-words (e.g., Tainturier & Rapp, 2001). In accordance with this dual route theoretical conception, reading and spelling acquisition of alphabetic languages have been described in step-by-step models (e.g., Ehri, 1989; Frith, 1985). The first steps allow the acquisition of phoneme-grapheme correspondences and the implementation of the non-lexical pathway. The last steps correspond to the implementation of the lexical pathway, through gradual memorization of lexical orthographic knowledge. However, even if it is often considered that this lexical memorization is not the object of the first steps of learning to read and spell, some research has shown that lexical orthographic knowledge begins to build from the very first steps of literacy learning (e.g., Bosse et al., 2003; Cassar & Treiman, 1997), even if it grows slowly during the first year of reading instruction.

On a practical level, currently, most researchers insist on the superiority of phonics instruction to teach reading and spelling to first grade pupils (e.g., Castles et al., 2018; Ehri et al., 2001; Hattie, 2008; National Reading Panel, 2000), even if the issue is still discussed (e.g., Bowers, 2020; Buckingham, 2020; Wyse & Bradbury, 2022). Phonics instruction means explicitly teaching the foundations of the non-lexical route, that is the relationship between the sounds of the spoken language and the letters or groups of letters of the written language. But this goal could be reached in very different ways (e.g.,

Henbest & Apel, 2017). In the classroom, a wide variety of exercises may help children acquiring phoneme-to-grapheme correspondences (e.g., Adams, 1990; Castles et al., 2018; Dahl et al., 2001) and observations of teachers' practices highlighted the multifaceted nature of phonics instruction (e.g., Mesmer & Griffith, 2005), including several reading and spelling exercises (e.g., Goigoux, 2016, for the description of French teacher practices).

Spelling activity

Spelling activity is recognized as effective for improving literacy acquisition. (e.g., Ouellette & Sénéchal, 2008; Uhry & Shepherd, 1993). Proposing spelling exercises from the earliest learning stages in combination with reading exercises appears to be beneficial to reading and spelling acquisition (Adams, 1990; Ehri, 1989; Henbest & Apel, 2017; Treiman, 1998; Weiser, 2013; Weiser & Mathes, 2011), especially for struggling readers. For example, Weiser and Mathes (2011) synthesized 11 studies testing the effect of spelling instruction on the performance of primary school students who are poor spellers and readers. The results showed that spelling instruction in the reading program is very effective ($d = 0.81$) in improving students' performance in word reading and spelling (see also Ise & Schulte-Körne, 2010).

Despite the clearly positive effects of spelling tasks on early literary acquisition, it appears that first grade teachers allocate only a small amount of time to spelling activities (less than 4% in the U.S., according to Weiser & Mathes, 2011). A recent observational study suggests that first grade French teachers allocate only 54 min per week to spelling activities, representing 12% of the total time allocated to literacy instruction (Goigoux, 2016). Prior research (Håland et al., 2019) suggests the following reasons for limited spelling instruction in first grade: a conceptualization of literacy development in two stages (reading acquisition before learning to spell), a task too difficult for beginners, and a lack of pedagogical knowledge on how to implement effective spelling practices.

The most valuable support to teachers would be for researchers to study the precise relevance of spelling tasks for first grade learners. The question then becomes how best to structure spelling exercises for beginning readers. Indeed, a variety of tasks can be proposed that require the student to produce the written form of a given word, sentence or text. The item can be given to the child orally or through a visual image or picture. Additionally, the tool used by the child can vary (handwriting, keyboard, mobile letters, spelling aloud, etc.). A recent observational study suggests that the most frequent spelling tasks used by first-grade French teachers are dictation to spelling and picture to spelling (Goigoux, 2016). Research supporting the inclusion of spelling tasks in early literacy instruction does not specify the task to be used. However, in France, teachers have recently been instructed not to use pictures to teach spelling (MENJ, 2019). In order to help teachers choose the most appropriate spelling task(s), it seems important to better understand what makes them different.

In fact, the differences between the dictation to spelling task and the picture to spelling task have only rarely been studied, and never with beginning readers. Bonin et al. (2015) sought to determine the implications of lexical and non-lexical pathways in different spelling tasks, through the performances of 34 French adult students. Participants were prompted to write the same list of words from their picture form (picture to spelling), from their oral form (dictation to spelling) and from their visual form (copying). Overall, the picture to spelling task leads the writer to make more errors because of picture misidentification, and to be slower to initiate the handwriting movement (latency) than the dictation to spelling task. Moreover, the picture to spelling task seems more sensitive than the dictation to spelling task to the lexical frequency of the words to be written. In contrast, only the dictation to spelling task seems sensitive to the phoneme-grapheme consistency of the items. As lexical frequency is considered to be linked to the use of the lexical pathway and consistency to the use of the non-lexical pathway, these results suggest that the dictation to spelling task would mobilize more the non-lexical route, and less the lexical route, than the picture to spelling task. Accordingly, it has been suggested that during the picture to spelling task, as during the oral picture naming task (e.g., Alario et al., 2004), a lexical-

semantic processing is necessarily involved to generate the object's name from the picture (e.g., Bonin et al., 2012, 2015; Torrance et al., 2018). On its side, the dictation to spelling task is affected by the phoneme-grapheme consistency (e.g., Delattre et al., 2006), an indication of the use of the direct phonological to spelling non-lexical route.

Another difference between the dictation to spelling and the picture to spelling tasks could be the involvement of verbal working memory (Pérez et al., 2016). Indeed, during the dictation to spelling task, the item to encode is a spoken word that needs to be kept in memory until it is written down. For example, the dictation to spelling performance of children with learning disabilities is especially correlated with verbal working memory impairments (e.g., Brandenburg et al., 2015). In contrast, during the picture to spelling task, the item to encode is a visual input that, most often, remains visible throughout the exercise: there is no need to temporarily store it in memory. Accordingly, the picture to spelling may be easier than dictation to spelling, especially in the case of a weak verbal memory span (see Pérez et al., 2016, for a similar reasoning to compare dictation to spelling and copying tasks) and for young children for whom the functional capacity of working memory is still weak (e.g., Gathercole et al., 2004).

Relevance of the present study

As the studies on expert readers suggest that dictation and picture to spelling tasks may involve different cognitive processing and have different cognitive costs, it seems crucial to investigate this same issue with beginning readers. These two spelling tasks are both widely used from the first grade (e.g., Goigoux, 2016). Because beginning readers have limited orthographic lexical knowledge and mainly use the non-lexical pathway, one may assume that the dictation to spelling task is more appropriate for these children. However, this advantage of the dictation to spelling task may be counterbalanced by the need to maintain the item in working memory.

The purpose of this study was to determine whether one task is more suitable than the other for spelling training of beginning readers, beyond their potential differences related to a different involvement of working memory. At the end of first grade, French students were asked to spell words from their oral form (dictation to spelling) and from their drawing picture (picture to spelling). In both tasks, each item (oral form or picture) was presented to the children for a fixed time period and the children were asked to spell the word only after the cue disappeared. The two tasks were therefore equivalent in terms of working memory involvement, removing the possibility that any difference between the two tasks could be attributed to a difference at this level (Pérez et al., 2016) but only to a difference in spelling processing. We hypothesized that the dictation to spelling task would lead to better performance, knowing that at the beginning of learning, children more often use the non-lexical pathway, a pathway that could be more involved in the dictation to spelling task than in the picture to spelling task (Bonin et al., 2015, 2019). The hypotheses, variables and exclusion criteria were prerecorded on Open Science Framework (preregistration on OSF).

Method

Participants

Seventy French first graders (32 girls), randomly selected from six first-grade classrooms located in six different public elementary schools¹ in Grenoble and its surrounding communities (France), took part in this study. Seven students were excluded because they obtained a very low score (above the 5th percentile) in the Raven's nonverbal reasoning ability task (Raven et al., 1998). The 63 selected children (29 girls) ranged in age from 6.5 to 7.5 years ($M = 6.98$, $SD = 0.27$), with an average score of 24.8 on the Raven's task ($ET = 5.8$, range 13–35), which corresponds to the 50th percentile at the age of 7. All were of French mother tongue but eight also spoke one or two other languages. Eighteen participants

(29% of the sample) came from a school receiving mostly an underprivileged population. This percentage is slightly higher than the national percentage of such schools in France (in 2018, 19.5% of French elementary schools, DEPP, 2019). All 63 students received written approval from their parents to participate in this study and they also made their own agreement.

All the participants were also assessed on their reading level and verbal working memory skills. Verbal working memory was estimated with the backward digit span task of WISC V (Wechsler, 2016). Their total raw score on this task was 5.5 on average (SD = 1.9, range 0–10). The reading level was assessed by a One-Minute-Test including words and pseudo-words (Sprenger-Charolles et al., 2005). The mean performance was 32.5 items correctly read by minute (SD = 11.8, range 8–60), with 37.2 (SD = 14.4, range 10–73) for the words and 27.7 (SD = 9.9, range 6–46) for the pseudo-words. On Table 1, the characteristics of the participants were presented after the sample has been cut in half according to the reading level, a group of good readers (above the median) and a group of poor readers (below the median).

Table 1.

Characteristics of the participants according to their reading level.

One Minute Test	N	Age in months	Raven score	Reading score	Working memory
Poor readers	32	83.8 (2.9)	23.2 (5.8)	22.6 (6.9)	4.9 (1.9)
Good readers	31	83.7 (3.4)	26.4 (5.3)	42 (7.6)	6.1 (1.7)
Total	63	83.7 (3.2)	24.8 (5.8)	32.5 (11.8)	5.5 (1.8)

Note. N: number of participants. Raven score: mean raw score on the Raven test PM 47 (maximum score : 36). Reading score: mean global score of the One Minute Test by averaging the word score and the pseudoword score. Working memory: mean score of the Backward Digit Span (WISC V, Wechsler, 2016). Standard deviations are presented in parentheses.





















Stimuli, apparatus and setting

Each participant had to encode 20 items (see Table 2). These items consisted of 5-letter words frequently studied during first grade, paired with their drawing picture (chosen in a drawing picture bank, Cannard et al., 2006) and pronunciation (recorded by a unique female speech therapist). The material used for both tasks are available on Open Science Framework (material on OSF). Several criteria guided the choice of the items. The correct naming rate of the drawing pictures had to be higher than 80% (Cannard et al., 2006) in order to avoid incorrect recognition in the picture to spelling task. The spoken items had to be sufficiently frequent to be known by all the participants, so with an estimated lexical frequency in number of occurrences per million words higher than 50 (Manulex, Lété et al., 2004). In order to avoid a floor effect due to items too difficult to encode for first graders, we selected items with a phoneme-to-grapheme consistency higher than 50% (Manulex-infra, Peereman et al., 2007) and with no homophones. Two sets of 10 items, paired in both frequency and consistency, were established (Table 2). The mean frequency was 249 (SD = 173, range 64–545) for Set 1 and 247.2 (SD = 179.6, range 54–609) for Set 2. The mean phoneme-to-grapheme consistency was 82.8 (SD = 12.6, range 59–97) for Set 1 and 82.9 (SD = 6.6, range 75–98) for Set 2.

This study follows a within-subject design. For each participant, 10 items of one set were presented in the form of auditory stimuli (dictation to spelling condition) and 10 items of the other set in the form of visual stimuli (picture to spelling condition). Four blocks were created to counterbalance the order of conditions and word sets. Each participant has been randomly assigned to one block. Table 3 summarizes the number of participants per block and the block structure.

Table 2.

Items features of both sets: associated pictures, lexical frequency and phoneme-grapheme consistency.

Set 1	Items	singe <i>monkey</i>	poire <i>pear</i>	balai <i>broom</i>	bague <i>ring</i>	vache <i>cow</i>	avion <i>airplane</i>	robot <i>robot</i>	arbre <i>tree</i>	table <i>table</i>	porte <i>door</i>	Means
	Pictures											
	Frequency	64	71	108	131	176	235	282	398	480	545	249
	Consistency	59	82	92	67	84	88	73	92	94	97	82,8
Set 2	Items	crabe <i>crab</i>	piano <i>piano</i>	zèbre <i>zebra</i>	niche <i>niche</i>	poule <i>hen</i>	fleur <i>flower</i>	pluie <i>rain</i>	lapin <i>rabbit</i>	route <i>road</i>	livre <i>book</i>	Means
	Pictures											
	Frequency	54	63	120	132	184	205	294	387	424	609	247,2
	Consistency	75	81	81	84	77	90	81	81	81	98	82,9

Note. The lexical frequency is expressed in number of occurrences per million based on a corpus of words extracted from books used in French first grade classrooms. The Phoneme-to-Grapheme consistency of each item is expressed in percentages. It is equal to the mean consistency of all of its phonemes-to-grapheme mappings; the consistency of each phoneme-to-grapheme mapping is calculated as the frequency at which a particular phoneme-to-grapheme mapping occurs, divided by the total frequency of that phoneme multiplied by 100.

The experimental sessions were conducted on the Qualtrics questionnaire software running on a digital tablet (Asus ZenPad 10 Z301M P028, 10.1" 1280 × 800, LCD screen) with a touch screen and an ABC touch keyboard with large keys (from the Keyboard for Seniors application). All stimulus presentations were fully automated and touch screen responses were recorded by the software. Auditory stimuli were presented through headphones at a comfortable volume, as determined by the experimenter and participant during an initial volume test. All experimental sessions were conducted in a quiet room away from the participants' classrooms. The experimenter, seated close to the participant, monitored all sessions, checked for technical problems and administered verbal reinforcements only during training.

Table 3.

Counterbalancing procedure.

Condition order	First		Second		n
Block	Task	Set	Task	Set	
Block A	dictation to spelling	1	picture to spelling	2	18
Block B	dictation to spelling	2	picture to spelling	1	17
Block C	picture to spelling	1	dictation to spelling	2	15
Block D	picture to spelling	2	dictation to spelling	1	13

Procedure

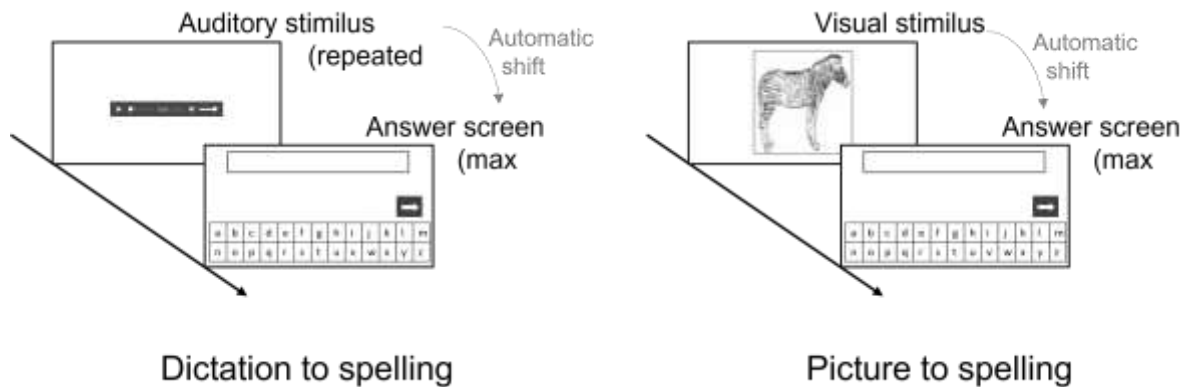
The procedure included a training phase followed by an experimental phase. The purpose of the training phase was to familiarize the participants with the device. The students were first asked to write the name of their favorite character in the right place using the keyboard and then validate their answer. The same task was repeated with their favorite color. The students were then asked to practice the task with the training words *filles* (girl) and *stylo* (pen) in the dictation to spelling condition and with the training words *pomme* (apple) and *train* (train) in the picture to spelling condition.

The experimental phase then started with 10 items in one condition and 10 other items in the other condition, according to the block randomly assigned. An example of a trial for each condition is illustrated in Figure 1. Each trial started with the presentation of the item that had to be encoded. In the dictation to spelling condition, the target word was successively heard twice for a total listening time of

6.18 s on average (between 5 and 9 s according to the item length). In the picture to spelling condition, the drawing of the target word was present in the center of the screen for 8 s, a sufficient time for the child to explore and name the image. In both conditions, at the end of the item presentation, the response screen appeared. The response screen comprised a rectangular cell for typing the word, a blue arrow to validate the answer and a tactile keyboard which appeared at the bottom of the screen after the child had clicked in the rectangular cell. Participants spelled words using the keyboard, controlled their response in the rectangular cell, and validated their answer. If they didn't validate after 60 s, the next screen automatically appeared. Sessions lasted 10 min on average.

Figure 1.

Study procedure: example of trial in dictation to spelling to spelling and in the picture to spelling tasks.



Measures

Three different scores were given to assess participants' spelling performance. Note that for each score, we ignored accent errors (on the item *zèbre*) because they are not easily or quickly accessible on the keyboard used.

The orthographic measure (scoreO) scored 1 if the written word contained no error and 0 if the written word contained at least one error (e.g., target word: balai (a broom), written word: balai = 1; balei = 0; balia = 0).

The phonological measure (scoreP) scored 1 if the written word was pronounced the same as the target word and 0 if the pronunciation was different (e.g., target word: balai, written word: balai = 1; balei = 1; balia = 0).

The letter measure (scoreL) corresponded to a score between 0 and 5. 1 point was given for each correct letter and 1 point was removed for each incorrect letter and for each letter inversion (e.g., target word: balai, written word: balai = 5; balei (4 correct letters – 1 wrong letter) = 3; balia (5 correct letters – 1 inversion) = 4).

The length of the response time (RT) was registered using the Qualtrics software, from the end of the stimulus screen until the moment the participant validated their answer. However, due to technological constraints, this time measure included not only the time to produce the spelling (time to find the letters on the keyboard and click on them) but also the reaction time (time between the appearance of the response screen and the spelling of the first letter) and the response validation time (time between the spelling of the last letter and the validation by the arrow button). This measure was therefore not informative and has not been analyzed.

Analyses

Data preparation

We expected to record 1260 observations for the four measures, 630 for both tasks. To clean our data, all responses associated with a misidentification were excluded (incorrect picture or sound recognition), as for example if *cabane* (hut) was written instead of *niche* (doghouse) in the picture to spelling condition. Indeed, the goal was to measure the spelling skills of the participants and not their recognition performance. The exclusion of misidentification errors resulted in a 2% loss of total data. All answers affected by a technical problem related to the use of digital tablets (e.g., no answer or only partial answer recorded) have also been deleted. This represented a loss of 17% of the dataset. Finally, the analyzed data included 503 observations for the dictation to spelling task (79.8%) and 518 observations for the picture to spelling task (82.2%).

Data analysis plan

To test our hypothesis, we conducted confirmatory analyses using a frequentist analysis method and then a Bayesian method. In addition, exploratory analyses were carried out to explore the relations between child performance and items' characteristics. These different analyses were conducted on R (Version 4.0.2) and RStudio (R Core Team, 2019). The data (data on OSF) and the script (analysis on OSF) are available on Open Science Framework.

Confirmatory analyses

Frequentist method

Data were analyzed using linear mixed effects (LME) models (Baayen et al., 2008). ScoreO, scoreP and scoreL were the key dependent variables. In a first analytical approach (model 1), the only fixed factor for the three measures was the type of task (dictation to spelling vs. picture to spelling). Participants, nested in blocks, and items were specified as random factors. In a second analytical approach (model 2), we added the reading score in continuous value (One Minute Test total) and its interaction with the type of task as additional fixed factors, in order to control the potential effect of reading skills on the dependent variables. For the orthographic measure and the phonological measure, in which the dependent variable was binomial, a logistic model was fitted by using a Binomial family. For the letters score measure, in which the dependent variable was ordinal and ranged between 0 and 5, a logistic model was fitted by using a Poisson family. In order to determine the most optimal model for our data, we followed the recommendations of Bates et al. (2018). Initially, a maximal model was specified, including all participants and items random intercepts and random slopes. Next, we fit a zero-correlation model. Then, the parameters not significantly explaining a part of the variance were removed one by one, until the likelihood ratio test showed no further improvement. We reported p-values for these tests with a significance threshold set at $p < .05$, obtained by Satterthwaite approximation with the lmerTest package (Version 3.1.2; Kuznetsova et al., 2017). The confidences intervals (CI) were indicated.

Bayesian method

As a complement to the frequentist analysis, we used a Bayesian analysis approach. The Bayes factors (BF) allow us to determine the degree of certainty of the observed effect. Compared to frequentist analyses, Bayesian analyses allow us to interpret insignificant results by indicating whether observations support a null hypothesis or the data are simply insensitive (Dienes, 2014). We will interpret the statistical results using the BF, which corresponds to the degree of likelihood of a given model. BF10 is used to determine the degree of likelihood of the alternative model (H1) over the null (H0), and BF01 determines the inverse, the likelihood degree of H0 over H1. As proposed by Jeffreys (1998) and repeated many times by other authors (e.g., Wagenmakers et al., 2011), the BF has been interpreted with the following classification: A BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides

very strong evidence and $100 < BF$ provides extreme/decisive evidence. In other words, a BF_{10} higher than 30 indicates very strong evidence for the alternative hypothesis relative to null and a BF_{01} higher than 30 indicates very strong evidence for the reverse pattern.

Exploratory analyses

In exploratory correlational analyses, we tried to investigate the involvement of the lexical and non-lexical pathways in both tasks. A relation between the children's performance and the items' frequency will be used as an index of the lexical pathway mobilization and a relation between the children's performance and the items' consistency will be used as an index of the non-lexical pathway mobilization (Bonin et al., 2015). Correlations were calculated between the non-binomial measure (scoreL) and the frequency and consistency characteristics of the items. The Spearman correlation coefficient was used because the distribution was not normal (Artusi et al., 2002). To determine the importance of one relationship relative to another, tests of significance were applied. The website Psychometrica was used to conduct comparison of correlations from dependent samples (Lenhard & Lenhard, 2014, calculation according to Eid et al., 2017).

We also explored the involvement of verbal working memory in both tasks, computing correlation between the students' performance on the working memory test (Backward Digit Span) and their performance (scoreL) on the two tasks. We explored whether students' performance in working memory was differently related to their performance in spelling across tasks. Tests of significance (Lenhard & Lenhard, 2014) were used to determine the importance of one relationship relative to another. Since in both tasks the items disappeared when the child could respond, we supposed the involvement of working memory to be equivalent in the two tasks.

Results

Descriptive results are shown in Table 4. The children's spelling performance (scoreO) was on average 50% successful. In other words, on average, children wrote half of the words with the correct spelling. ScoreP was higher. Indeed, roughly 76% of the answers were phonologically correct. We observe neither a floor effect nor a ceiling effect for these two scores. However, for both measures, very large standard deviations are a sign of important inter-individual variability. The score per letter (scoreL) was around 83% correct, with smaller standard deviations.

The correlation matrix showed that the three scores were highly and positively correlated with each other ($r = .55, p < .001$ for scoreO and scoreP; $r = .75, p < .001$ for scoreO and scoreL; $r = .67, p < .001$ for scoreP and scoreL). They were all correlated with reading performance, working memory, and Raven. Reading performance measure was positively correlated with the working memory measure ($r = .33, p < .001$) and Raven measure ($r = .37, p < .001$), and working memory measure with Raven measure too ($r = .52, p < .001$).

Confirmatory analyses

Did participants score better on one task than another?

The mixed model 1 revealed that the difference between the scores in picture to spelling and those in dictation to spelling are not significant for the orthographic score (scoreO: $\beta = 0.13, z(1021) = 0.84, p = .404, 95\% \text{ CI } [-0.18, 0.45]$). The Bayes factors provided strong evidence of the absence of difference ($BF_{01} = 13.7$). For the phonological score, we observed no significant difference either (scoreP: $\beta = -0.16, z(1021) = -0.94, p = .347, 95\% \text{ CI } [-0.53, 0.7]$), with a moderate evidence for the null hypothesis ($BF_{01} = 7.1$). The letter score measure also seems equivalent in the two conditions (scoreL: $\beta = 0.01, z(1021) = 0.39, p = .699, 95\% \text{ CI } [-0.05, 0.07]$), with a strong evidence of the absence of difference ($BF_{01} = 13.5$).

Table 4.

Descriptive statistics of the scores measures as a function of spelling tasks and children's performance in reading.

Tasks	One Minute Test	ScoreO (%)	ScoreP (%)	ScoreL (%)
Dictation to spelling	Poor readers	33.3 (47.3)	68.1 (46.7)	77.4 (23.6)
	Good readers	61.8 (48.7)	82.7 (37.9)	87.4 (20.8)
	Total	47.9 (50)	77.4 (41.9)	82.8 (22.2)
Picture to spelling	Poor readers	32.1 (46.8)	67.4 (47)	77.2 (23)
	Good readers	64.5 (47.9)	84.1 (36.7)	88.2 (19.6)
	Total	50.2 (50)	75.1 (43.3)	83.6 (22)

Note. ScoreO, scoreP and scoreL represent the mean percentage of correct responses. Poor readers refer to children who read fewer than 33 items on average on the One Minute Test, and good readers refer to children reading 33 or more items on average. Standard deviations are presented in parentheses.

When children's reading performance was included in the models as an additional fixed factor (model 2), we still did not observe any effect of the task on performance, neither for scoreO ($\beta = -0.23$, $z(985) = -0.46$, $p = .646$, 95% CI [-1.26, 0.78], $BF_{01} = 13.6$), nor for scoreP ($\beta = -0.06$, $z(985) = -0.12$, $p = .901$, 95% CI [-1.15, 0.9], $BF_{01} = 8.2$) and nor for scoreL ($\beta = -0.03$, $z(985) = -0.32$, $p = .750$, 95% CI [-0.22, 0.15], $BF_{01} = 13.5$). However, all three measures varied significantly with children reading performance (scoreO: $\beta = 0.04$, $z(985) = 6.18$, $p < .001$, 95% CI [0.03, 0.06]; scoreP: $\beta = 0.04$, $z(985) = 6.29$, $p < .001$, 95% CI [0.2, 0.05]); scoreL: $\beta = 0.003$, $z(985) = 5$, $p < .001$, 95% CI [0.002, 0.005]). No interaction effect between the task and reading performance was found for any of the three measures (scoreO: $\beta = 0.005$, $z(985) = 0.644$, $p = .519$, 95% CI [-0.01, 0.02]; scoreP: $\beta = -0.003$, $z(985) = -0.235$, $p = .741$, 95% CI [-0.02, 0.01]); scoreL: $\beta = 0.03$, $z(985) = 0.44$, $p = .658$, 95% CI [-0.002, 0.003]). The estimates of model 2 fixed parameters are summarized in Table 5.

In sum, these results indicate that participants did not score better on one task than the other. However, unsurprisingly, children who were effective at reading were also effective at spelling and vice versa, regardless of the task (e.g., Georgiou et al., 2020).

Exploratory analyses

Did a pathway get mobilized more than the other depending on the task?

Although the material used in this experiment was simple and suitable for first graders and was not designed to measure extremes such as high versus low item frequency or consistency, we did explore the relationship between item characteristics and performance levels for each task. As observed by Bonin et al. (2015), an effect of item frequency may point to involvement of the lexical pathway, whereas an effect of item consistency may suggest involvement of the non-lexical pathway.

Correlation analyses between letter scores obtained on the two tasks and the frequency and consistency characteristics of the items are presented in Table 6. Positive, weak but statistically significant correlations were observed between the scoreL measure and both frequency and consistency. This means that the more frequent and consistent a word is, the greater the number of correct letters in a spelling task, whatever the nature of the input. The correlations with both frequency and consistency were not statistically significantly stronger for the scoreL in the picture to spelling task ($p = .396$) than for the scoreL in the dictation to spelling task ($p = .45$).

Table 5.

Standardized parameter estimates for model 2 computed independently on the three dependent variables.

Model 2 : Score ~ Task + ReadingLevel + Task*ReadingLevel			
Dependent variables:	ScoreO	ScoreP	ScoreL
Fixed variables:			
Intercept	-2.83***	-0.77	1.41***
Task [ref: dictation to spelling]	-0.23	-0.06	0.01
Reading level	0.04 ***	0.04 ***	0.16 ***
Task*reading level	0.004	-0.003	0.03
Observations	985	985	985

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Did the students' performance in working memory correlate differently with their performance on the letter score according to the task?

Correlation analyses between the letter score and working memory (Digit Span Backward) showed a positive link for both the picture to spelling task and the dictation to spelling task (see Table 6). They suggest that the better the student's working memory, the greater the number of correct letters. Significance tests did not reveal one correlation stronger than another ($p = .368$). As expected, working memory seems to play an equivalent role in the two tasks.

Table 6.

Correlations between scoreL on spelling tasks and the frequency and consistency of items, and the children's working memor.

	Task	Items frequency	Items consistency	Children's working memory
ScoreL	dictation to spelling	.10*	.13**	.23***
	picture to spelling	.12**	.09*	.28***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The purpose of this study was to compare the performance of students at the end of first grade on two different spelling exercises; dictation to spelling and picture to spelling. We hypothesized that the dictation to spelling task would lead to better performance, knowing that at the beginning of learning, children more often use the non-lexical pathway, a pathway that may be more involved in the dictation to spelling task than in the picture to spelling task (Bonin et al., 2015, 2019). To test this hypothesis, children were asked to write two equivalent sets of words, one from the auditory form of the words (dictation to spelling) and the other from the visual form of the words (picture to spelling). Contrary to the hypothesis, our results suggest that students spelling performance at the end of first grade does not vary significantly according to the task. Bayesian statistical analyses allow us to say that this lack of effect is not due to a problem of too low sensitivity of our data. More precisely, the Bayes factors provided moderate (for phonological score) to strong evidence (for orthographic and letter scores) of the absence of difference. We can be confident, therefore, that first-grade children's performance is equivalent between the two tasks.

Data on adult expert readers suggest that these two tasks are not equivalent in terms of the pathways involved. More precisely, according to the results of Bonin et al. (2015), the lexical pathway is involved in all spelling tasks but makes a stronger contribution in the picture to spelling task than in the dictation

to spelling task. In contrast, the non-lexical pathway seems to be more involved during a dictation to spelling task than during a picture to spelling task. Our results suggest that these conclusions obtained on adult expert performance are not transposable to beginning readers and spellers. Indeed, both the orthographic score of first graders, which may be considered as an estimation of the lexical pathway performance, and their phonological score, which may reflect the correct use of the non-lexical pathway, obtained the same performance in the picture to spelling and the dictation to spelling tasks.

To explain this result, our primary hypothesis is that French first graders have difficulty using the lexical pathway to spell words. Consequently, they preferentially use the non-lexical pathway to spell almost all items (Goigoux, 2016), whatever the task used. Because French written language is opaque and difficult to learn (Peereman & Content, 1999; Ziegler et al., 1996), they may still have too little orthographic lexical knowledge, or even if they have some orthographic lexical knowledge (e.g., Sprenger-Charolles et al., 1998), they have more difficulty mobilizing this knowledge during spelling tasks than during reading tasks. The fact that the orthographic score is not very low in the present experiment could be explained by the fact that several items are fully consistent and could be correctly spelled even using the non-lexical pathway. However, it has been shown that children, at the end of first grade, may acquire and use some orthographic lexical knowledge (e.g., Bosse et al., 2003; Cassar & Treiman, 1997; Martinet et al., 2004; Sprenger-Charolles et al., 1998), suggesting that the combined use of the two pathways is plausible at the end of first grade. The present results simply suggest that both pathways are used in the same way in both a picture to spelling task and a dictation to spelling task.

We found that reading performance explains performance in the two spelling tasks. Those children proficient at reading were also proficient at spelling. These findings are in accordance with the literature, which indicates that reading and spelling are closely related skills (e.g., Abbott et al., 2010). Indeed, it is important to teach these skills together, as they are mutually supportive (e.g., Henbest & Apel, 2017). However, an interaction effect may have been expected between reading and spelling performance depending on the task. As the lexical pathway seems more involved in the picture to spelling task (Bonin et al., 2015, 2019) and knowing that poor readers have particular difficulties to build this lexical pathway, we might have expected a larger difference between the tasks (in favor of the dictation to spelling task) in poor readers than in good readers. This is not what the results of this study show. One task is not better than the other, even for poor readers. This is another argument in favor of the idea that all first-grade children use the same process to encode the target words, whatever the task. Because learning to spell, at least in opaque alphabetic languages, begins by matching a sound with its letter (Ehri, 1987; Fayol & Jaffré, 2016), even good readers at the end of first grade are likely to use preferentially the phoneme-grapheme conversion strategy for both tasks.

Exploratory analyses showed the presence of a correlation between item consistency and performance. This correlation is equivalent in both tasks. As we mentioned earlier, an influence of item consistency is a manifestation of the mobilization of the non-lexical pathway. This correlation confirms that in both the dictation to spelling task and the picture to spelling task, children use the non-lexical pathway to spell the words. However, a frequency effect, equivalent in both tasks, is also observed. Correlations show that the more frequent a word is, the more correct the number of letters written. This frequency effect confirms that, actually, both pathways are involved during spelling tasks, even if, as already suggested, the non-lexical pathway seems to prevail when spellers are first graders.

Finally, a significant correlation was also observed between spelling performance and working memory span. Whatever the task, the correlations were of equal strength, suggesting that working memory is as involved in the dictation to spelling task than in the picture to spelling task. It is important to remember that we chose to make the picture disappear in the picture to spelling task, so that this task requires, as much as the dictation to spelling task, the retention in memory of the word to be written. In both tasks, the item has to be recalled until the end of spelling, which can increase cognitive load (Sweller, 1988, 2011). This is different from the classic picture to spelling task where the picture remains visible during spelling production. In such a case, the picture can play the role of an external memory (Pérez et al.,

2016). One could expect that, for young children having a short working memory span, this last situation will be easier than a dictation to spelling task.

Regarding teaching practice, the main results of this study suggest that first graders may process different inputs, such as oral words or drawing pictures, and spell them in the same way. According to the Bayesian analyses, there is strong evidence that young children spell words under dictation as successfully as they spell words in a picture to spelling task. Then, it seems that various spelling exercises could be proposed to first graders with the same benefit, there is no evidence to say that one task is to be preferred for practicing spelling, even for poor readers.

The results of this experiment are not in accordance with the recommendation to avoid the use of pictures during both spelling and reading exercises (MENJ, 2019). In contrast, knowing that spelling is beneficial to reading acquisition (e.g., Adams, 1990; Henbest & Apel, 2017; Weiser, 2013; Weiser & Mathes, 2011), teachers should be encouraged to provide frequent and varied spelling activities, such as dictation to spelling, picture to spelling, spelling with mobile letters, copying, etc. Moreover, publishers of teaching tools for first grade have to be encouraged to publish various spelling exercises, including exercises using pictures. It is important to specify that the tasks of the present experiment have been built for experimental purposes. They are not optimal teaching situations. For example, a picture to spelling task where the picture is maintained visible during spelling production is preferable for teaching practice. It would allow children with poor working memory to do the task, without forgetting the item before the end of spelling production. For the same reasons, during a dictation to spelling task, the item could be repeated more than twice.

We do want to call attention to some of our methodological decisions for this study that may merit consideration before generalization. Given that our participants were beginning readers and spellers, the study was conducted on a limited number of frequent and fairly easy to spell words. Generalizing our findings to a whole curriculum will require additional studies conducted on older children, with more items for encoding, and with items featuring greater variability in consistency and frequency. It seems likely that a difference between dictation to spelling and picture to spelling tasks, in terms of the lexical versus non-lexical pathways involved, would appear during the years of learning written language, as this difference is observed in expert readers (Bonin et al., 2015).

To summarize, our study showed, with moderate to strong evidence, that first-grade children spell words in the same way during a dictation to spelling task or a picture to spelling task. The processes they used to perform these two tasks seem similar, including both the non-lexical and lexical pathways, whether the input is auditory or visual. As guidance for informing pedagogical choices on a scientific basis, studies evaluating and comparing pedagogical practices are still too few and should be developed further. In line with that objective, findings from the present study suggest that both picture to spelling task and the dictation to spelling task may be proposed to first graders with the same benefit.

Note

1. In France, education in public schools is free of charge. Public schools represent 98% of French elementary schools (DEPP, 2019).

Acknowledgements

This study was carried out with the support of both the National Technology Research Association (ANRT, France) and of the industrial partner specialised in school book publishing “Les Éditions Hatier”, within the framework of a CIFRE grant (grant for collaborations between research laboratories

and industrial partners) for the PhD of the first author. We thank the teachers and pupils of the schools of the Grenoble area for their participation.

Ethical approval

The research meets ethical guidelines and adheres to the legal requirements of the study country. In French, when research involving human falls under the law n° 2012-300 of March 5th, 2012 of the French Public Health Code (known as law Jardé), that is to say when they are organized and practiced on the human being in view the development of biological or medical knowledge, they must be examined by a Committee for the Protection of Persons (CPP). The present research does not aim at increasing medical or biological knowledge. Therefore, the Ethics Committee for Research of the University of Grenoble Alpes (CERGA), certify that this research not require researchers to seek ethics approval from a CPP.

References

- Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, 102(2), 281–298. <https://doi.org/10.1037/a0019318>
- Adams, M. J. (1990). *Beginning to read: Learning and thinking about print*. MIT Press.
- Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 36(1), 140–155. <https://doi.org/10.3758/BF03195559>
- Andreu, S., Cioldi, I., Conceicao, P., Eteve, Y., Fabre, M., Le Breton, S., Persem, E., Portelli, T., Rue, G., Rocher, T., Vourc'h, R., & Wuillamier, P. (2019). *Evaluations 2020 Point d'étape CP: Premiers résultats*. <https://www.education.gouv.fr>
- Artusi, R., Verderio, P., & Marubini, E. (2002). Bravais-Pearson and Spearman correlation coefficients: Meaning, test of hypothesis and confidence interval. *The International Journal of Biological Markers*, 17(2), 148–151. <https://doi.org/10.5301/JBM.2008.2127>
- 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. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *arXiv:1506.04967 [stat]*. <http://arxiv.org/abs/1506.04967>
- Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental Psychology (2006)*, 68(7), 1268–1294. <https://doi.org/10.1080/17470218.2014.978877>
- Bonin, P., Méot, A., Laroche, B., Bugaiska, A., & Perret, C. (2019). The impact of image characteristics on written naming in adults. *Reading and Writing*, 32(1), 13–31. <https://doi.org/10.1007/s11145-017-9727-3>
- Bonin, P., Roux, S., Barry, C., & Canell, L. (2012). Evidence for a limited-cascading account of written word naming. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 38(6), 1741–1758. <https://doi.org/10.1037/a0028471>

- Bosse, M.-L., Valdois, S., & Tainturier, M.-J. (2003). Analogy without priming in early spelling development. *Reading and Writing*, 16(7), 693–716. <https://doi.org/10.1023/A:1025883815395>
- Bowers, J. S. (2020). Reconsidering the Evidence That Systematic Phonics Is More Effective Than Alternative Methods of Reading Instruction. *Educational Psychology Review*, 32(3), 681–705. <https://doi.org/10.1007/s10648-019-09515-y>
- Brandenburg, J., Kleszczewski, J., Fischbach, A., Schuchardt, K., Büttner, G., & Hasselhorn, M. (2015). Working memory in children with learning disabilities in reading versus spelling: Searching for overlapping and specific cognitive factors. *Journal of Learning Disabilities*, 48(6), 622–634. <https://doi.org/10.1177/0022219414521665>
- Buckingham, J. (2020). Systematic phonics instruction belongs in evidence-based reading programs: A response to Bowers. *The Educational and Developmental Psychologist*, 37(2), 105–113. <https://doi.org/10.1017/edp.2020.12>
- Cannard, C., Bonthoux, F., Blaye, A., Scheuner, N., Schreiber, A.-C., & Trinquart, J. (2006). BD2I : Normes sur l'identification de 274 images d'objets et leur mise en relation chez l'enfant français de 3 à 8 ans. *L'Année Psychologique*, 106(03), 375. <https://doi.org/10.4074/S0003503306003034>
- Cassar, M., & Treiman, R. (1997). The Beginnings of Orthographic Knowledge: Children's Knowledge of Double Letters in Words. *Journal of Educational Psychology*, 89(4), 631–644. <https://doi.org/10.1037/0022-0663.89.4.631>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Conrad, N. J., Kennedy, K., Saoud, W., Scallion, L., & Hanusiak, L. (2019). Establishing word representations through reading and spelling: Comparing degree of orthographic learning. *Journal of Research in Reading*, 42(1), 162–177. <https://doi.org/10.1111/1467-9817.12256>
- Dahl, K. L., Scharer, P. L., Lawson, L. L., & Grogan, P. R. (2001). *Rethinking Phonics: Making the Best Teaching Decisions*. Heinemann.
- DEPP. (2019). *Repères et références statistiques sur les enseignements, la formation et la recherche - 2019*. Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. <http://www.enseignementsup-recherche.gouv.fr/cid144369/reperes-et-references-statistiques-edition-aout-2019.html>
- Delattre, M., Bonin, P., & Barry, C. (2006). Written spelling to dictation: Sound-to-spelling regularity affects both writing latencies and durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1330–1340. <https://doi.org/10.1037/0278-7393.32.6.1330>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Ehri, L. C. (1987). Learning to Read and Spell Words. *Journal of Reading Behavior*, 19(1), 5–31. <https://doi.org/10.1080/10862968709547585>
- Ehri, L. C. (1989). The Development of Spelling Knowledge and Its Role in Reading Acquisition and Reading Disability. *Journal of Learning Disabilities*, 22(6), 356–365. <https://doi.org/10.1177/002221948902200606>

- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71(3), 393–447. <https://doi.org/10.3102/00346543071003393>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. <https://sfbs.tu-dortmund.de/handle/sfbs/3110>
- Fayol, M., Bonin, P., & Collay, S. (2008). La consistance orthographique en production verbale écrite : une brève synthèse. *L'Année Psychologique*, 108(03), 517–546. <https://doi.org/10.4074/S0003503308003059>
- Fayol, M., & Jaffré, J.-P. (2016). L'orthographe : Des systèmes aux usages. *Pratiques*, [En ligne], 169–170. <https://doi.org/10.4000/pratiques.2984>
- Ferrer, E., McArdle, J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, 43(6), 1460–1473. <https://doi.org/10.1037/0012-1649.43.6.1460>
- Frith, U. (1985). Beneath the surface of developmental dyslexia. *Surface Dyslexia : Neuropsychological and Cognitive Studies of Phonological Reading*, 32, 301–330. <https://ci.nii.ac.jp/naid/10022405906/>
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, 40(2), 177–190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Georgiou, G. K., Torppa, M., Landerl, K., Desrochers, A., Manolitsis, G., de Jong, P. F., & Parrila, R. (2020). Reading and spelling development across languages varying in orthographic consistency: Do their paths cross? *Child Development*, 91(2), e266–e279. <https://doi.org/10.1111/cdev.13218>
- Goigoux, R. (2016). Lire et Écrire, étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages. *Institut français de l'éducation (IFE)*. http://ife.ens-lyon.fr/ife/recherche/lire-ecrire_e.
- Håland, A., Hoem, T. F., & McTigue, E. M. (2019). Writing in First Grade: The Quantity and Quality of Practices in Norwegian Classrooms. *Early Childhood Education Journal*, 47(1), 63–74. <https://doi.org/10.1007/s10643-018-0908-8>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Henbest, V., & Apel, K. (2017). Effective Word Reading Instruction: What Does the Evidence Tell Us? *Communication Disorders Quarterly*, 39(1), 303–311. <https://doi.org/10.1177/1525740116685183>
- Houghton, G., & Zorzi, M. (2003). Normal and Impaired Spelling in a Connectionist Dual-Route Architecture. *Cognitive Neuropsychology*, 20(2), 115–162. <https://doi.org/10.1080/02643290242000871>
- Ise, E., & Schulte-Körne, G. (2010). Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia*, 60(1), 18–39. <https://doi.org/10.1007/s11881-010-0035-8>
- Jeffreys, H. (1998). *The Theory of Probability*. OUP Oxford.
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). *lmerTest Package: Tests in Linear Mixed Effects Models*. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

- Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, 100(1), 150–161. <https://doi.org/10.1037/0022-0663.100.1.150>
- Lenhard, W., & Lenhard, A. (2014). *Hypothesis Tests for Comparing Correlations*. Psychometrica. <https://www.psychometrica.de/correlation.html>.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156–166. <https://doi.org/10.3758/BF03195560>
- Martin, M. O., Mullis, I. V., & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. International Association for the Evaluation of Educational Achievement.
- Martinet, C., Valdois, S., & Fayol, M. (2004). Lexical orthographic knowledge develops from the beginning of literacy acquisition. *Cognition*, 91(2), B11–B22. <https://doi.org/10.1016/j.cognition.2003.09.002>
- MENJ. (2019). *Un guide fondé sur l'état de la recherche pour enseigner la lecture et l'écriture au CP*. Eduscol.
- Mesmer, H., & A., E., Griffith, P., L. (2005). Everybody's Selling It - But Just What Is Explicit, Systematic Phonics Instruction? *The Reading Teacher*, 59(4), 366–376. <https://doi.org/10.1598/RT.59.4.6>
- National Reading Panel. (2000). *Teaching Children to Read : an Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction : Reports of the Subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.
- 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. <https://doi.org/10.1111/j.1467-8624.2008.01166.x>
- Ouellette, G., Martin-Chang, S., & Rossi, M. (2017). Learning from our mistakes: Improvements in spelling lead to gains in reading speed. *Scientific Studies of Reading*, 21(4), 350–357. <https://doi.org/10.1080/10888438.2017.1306064>
- Peereman, R., & Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 31(2), 376–379. <https://doi.org/10.3758/bf03207735>
- Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, 39(3), 579–589. <https://doi.org/10.3758/BF03193029>
- Pérez, M., Giraudo, Hélène., & Tricot, A. (2016). Les processus cognitifs impliqués dans l'acquisition de l'orthographe : dictée vs. copie. *A.N.A.E*, 118, 280–286. <https://hal.archives-ouvertes.fr/hal-01403772>
- R Core Team. (2019). R A language and environment for statistical computing R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raven, J. C., Court, J. H., & Raven, J. E. (1998). *Raven's coloured progressive matrices*. Harcourt Assessment.

- Sprenger-Charolles, L., Bogliotti, C., Piquard-Kipffer, A., & Leloup, G. (2009). Stabilité dans le temps des déficits en et hors lecture chez des adolescents dyslexiques (données longitudinales). *ANAE - Approche Neuropsychologique des Apprentissages Chez L'enfant*, 21(103), 243–253.
- Sprenger-Charolles, L., Colé, P., Béchenec, D., & Kipffer-Piquard, A. (2005). French normative data on reading and related skills from EVALEC, a new computerized battery of tests (end Grade 1, Grade 2, Grade 3, and Grade 4). *European Review of Applied Psychology*, 55(3), 157–186. <https://doi.org/10.1016/j.erap.2004.11.002>
- Sprenger-Charolles, L., Siegel, L. S., & Bonnet, P. (1998). Reading and Spelling Acquisition in French: The Role of Phonological Mediation and Orthographic Factors. *Journal of Experimental Child Psychology*, 68(2), 134–165. <https://doi.org/10.1006/jecp.1997.2422>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J. (2011). Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Éds.), *Psychology of Learning and Motivation*. (Vol. 55, p. 37–76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Tainturier, M.-J., & Rapp, B. (2001). In B. Rapp (Éd.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*. (p. 263–289).
- Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., Dimakos, I., Fidalgo, R., Hyönä, J., Jóhannesson, Ó. I., Madjarov, G., Pauly, D. N., Upstad, P. H., van Waes, L., Vernon, M., & Wengelin, Å. (2018). Timed written picture naming in 14 European languages. *Behavior Research Methods*, 50(2), 744–758. <https://doi.org/10.3758/s13428-017-0902-x>
- Treiman, R. (1998). Beginning to spell in English. In C. Hulme, & R. M. Joshi (Ed.), *Reading and spelling: Development and disorders*. (p. 371–393).
- Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/Spelling Instruction as Part of a First-Grade Reading Program: Effects on Several Measures of Reading. *Reading Research Quarterly*, 28(3), 218. -. <https://doi.org/10.2307/747995>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011.) *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Wechsler, D. (2016). *WISC-V - Échelle d'intelligence de Wechsler pour enfants et adolescents - 5ème édition*. Harcourt Assessment. <https://www.pearsonclinical.fr/wisc-v>
- Weiser, B. L. (2013). Ameliorating Reading Disabilities Early: Examining an Effective Encoding and Decoding Prevention Instruction Model. *Learning Disability Quarterly*, 36(3), 161–177. <https://doi.org/10.1177/0731948712450017>
- Weiser, B., & Mathes, P. (2011). Using Encoding Instruction to Improve the Reading and Spelling Performances of Elementary Students At Risk for Literacy Difficulties: A Best-Evidence Synthesis. *Review of Educational Research*, 81(2), 170–200. <https://doi.org/10.3102/0034654310396719>
- Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading. *Review of Education*, 10(1), e3314. <https://doi.org/10.1002/rev3.3314>
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments & Computers*, 28(4), 504–515. <https://doi.org/10.3758/BF03200539>