



HAL
open science

Individual differences in the acquisition of non-linguistic audio-visual associations in 5 year-olds

Irene Altarelli, Ghislaine Dehaene-Lambertz, Daphne Bavelier

► **To cite this version:**

Irene Altarelli, Ghislaine Dehaene-Lambertz, Daphne Bavelier. Individual differences in the acquisition of non-linguistic audio-visual associations in 5 year-olds. *Developmental Science*, 2019, 10.1111/desc.12913 . hal-02324744

HAL Id: hal-02324744

<https://hal.science/hal-02324744>

Submitted on 22 Oct 2019

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.



Article type : Special Issue Article

TITLE: Individual differences in the acquisition of non-linguistic audio-visual associations in 5 year-olds

Running title: Non-linguistic audio-visual associations in 5 year-olds

AUTHORS:

Irene Altarelli¹⁻³, Ghislaine Dehaene-Lambertz¹ and Daphne Bavelier²

Authors' institutional affiliations:

1. Cognitive Neuroimaging Unit U992, INSERM, CEA DRF/Institut Joliot, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France.
2. Faculty of Psychology and Education Sciences, University of Geneva, Geneva, Switzerland.
3. CNRS UMR 8240, Laboratory for the Psychology of Child Development and Education (LaPsyDE), University Paris Descartes, Université de Paris, Paris, France.

Corresponding author: Irene Altarelli, LaPsyDE, 46 rue Saint-Jacques, 75005 Paris, France; irene.altarelli@parisdescartes.fr; phone: +330156771300.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/DESC.12913](https://doi.org/10.1111/DESC.12913)

This article is protected by copyright. All rights reserved

Conflict of Interest Statement: The last author, Daphne Bavelier, declares acting as a scientific advisor for Akili Interactive, Boston MA (USA) and for RedBull, Los Angeles CA (USA). The two other authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge all children, parents and teachers who participated in the study, Marie Palu for help with recruitment, Angela Pasqualotto for contributing to the audio-visual task programming, Adriano Sesser for designing the child-friendly visual stimuli of the audio-visual task, Virginie Van Wassenhove and Augustin Joessel for their help with the creation of the auditory stimuli used in the audio-visual association paradigm and Sylvain Charron for advice on statistical analyses. This research was supported by the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant no. 661667, LearningDeterminants to I.A.

RESEARCH HIGHLIGHTS

- This study presents a new child-friendly task measuring the acquisition of audio-visual associations between non-linguistic auditory signals and novel visual shapes.
- This new task is well calibrated for kindergartners, providing the necessary variability to explore relations between audio-visual association scores and other cognitive abilities.
- We found links between phonological awareness and scores in our audio-visual association task; language comprehension was also associated with scores in our novel task.
- These results underscore a key relation between foundational language competencies and audio-visual learning, even in the absence of linguistic input in the associative task.

ABSTRACT

Audio-visual associative learning – at least when linguistic stimuli are employed – is known to rely on core linguistic skills such as phonological awareness. Here we ask whether this would also be the case in a task that does not manipulate linguistic information. Another question of interest is whether executive skills, often found to support learning, may play a larger role in a non-linguistic audio-visual associative task compared to a linguistic one. We present a new task that measures learning when having to associate non-linguistic auditory signals with novel visual shapes. Importantly, our novel task shares with linguistic processes such as reading acquisition the need to associate sounds with arbitrary shapes. Yet, rather than phonemes or syllables, it uses novel environmental sounds – therefore limiting direct reliance on linguistic abilities.

Five-year-old French-speaking children (N=76, 39 girls) were assessed individually in our novel audio-visual associative task, as well as in a number of other cognitive tasks evaluating linguistic abilities and executive functions. We found phonological awareness and language comprehension to be related to scores in the audio-visual associative task, while no correlation with executive functions was observed. These results underscore a key relation between foundational language competencies and audio-visual associative learning, even in the absence of linguistic input in the associative task.

KEYWORDS

associative learning; audio-visual learning; language skills; pre-readers; cognitive correlates; individual differences

INTRODUCTION

Audio-visual learning is central to a variety of developmentally relevant processes. Among these, examples are the integration between vocal signals and visual cues (like facial movements) in social and language learning (Lewkowicz & Hansen-Tift, 2012; Weatherhead & White, 2017), the association of spoken words and visually-presented objects in the context of vocabulary learning (Pereira, Smith, & Yu, 2014), and the acquisition of speech sound-letter correspondences during reading development (Horbach, Scharke, Cröll, Heim, & Günther, 2015). Investigating how children come to master audio-visual associations is therefore key to many aspects of linguistic development.

A number of studies have explored audio-visual associative learning, throughout development. On top of a rich literature documenting infants' audio-visual abilities in the context of simultaneously heard and seen speech (Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012), a few studies have also reported infant's abilities to learn arbitrary correspondences between artificial auditory and visual stimuli (e.g. in 6-8 month-old infants, Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017).

Similar paradigms have also successfully been developed in kindergarten children aged 4 to 6, prior to formal elementary schooling. In this 4 to 6-years age range, most studies assessed associative learning by requiring to pair auditory linguistic items (ranging from non-words to syllables to single phonemes) with visual objects, pictures or symbols. The majority of these studies necessitated an oral output, i.e. they required participants to produce an oral response. This potentially results in confounding associative learning with children's speech production abilities and phonological short-term memory. Studies requiring such oral output included a variety of experimental procedures, such as: (i) having to associate non-words with cuddle toys of unknown animals, and having to respond by pronouncing the corresponding non-word upon presentation of each cuddle toy (de Jong et al., 2000); (ii) having to articulate the correct non-words paired with pictures of unfamiliar children, fantasy animals or unknown letters (Lervåg, Bråten, & Hulme, 2009); (iii) having to pronounce the syllables associated with pictures of imaginary animals (Georgiou, Liu, & Xu, 2017); (iv) during and/or following a computer-based training (Ehm et al., 2019; Horbach et al., 2015) having to orally produce the syllables corresponding to either single elementary symbols (triangle/square/circle) (Ehm

et al., 2019) or to sequences of up to 4 symbols, combining dots and dashes (Horbach et al., 2015). Together, these studies have revealed that the capacity to learn cross-modal mappings varies among kindergarteners and is positively related to their phonological awareness (de Jong et al., 2000; Ehm et al., 2019; Georgiou et al., 2017; Lervåg et al., 2009), letter/character knowledge (Ehm et al., 2019), and rapid automatized naming skills (Georgiou et al., 2017; Lervåg et al., 2009). Among the rare studies that did not require oral production is that of Karipidis and colleagues (2017). These authors' paradigm required selecting a response among visual options; a computerized training was applied, which in each trial asked the child to select the correct false-font character among one to three distractors, after having heard a given phoneme through headphones. This study confirmed the role of phonological awareness in audio-visual association mastery, even when oral production is not required.

In slightly older elementary school children (6-11 years old), associations between non-words and symbols have been studied as well, confirming a positive relationship between performance and phonological awareness (Litt, de Jong, Bergen, & Nation, 2013; Windfuhr & Snowling, 2001).

Apart from highlighting how audio-visual associative learning relates to phonological awareness, letter knowledge and rapid automatized naming, some of the aforementioned studies also considered the link between audio-visual associations mastery and executive functions (EFs). EFs are emerging in recent literature as key factors in facilitating learning in childhood (Blair & Razza, 2007; Cardoso-Leite & Bavelier, 2014; Lipsey et al., 2017) and academic success (Clements, Sarama, & Germeroth, 2016; Sasser, Bierman, & Heinrichs, 2015). In line with this view, performance in audio-visual associative tasks in kindergartners was found to be related to verbal working memory, assessed by backward verbal span (Ehm et al., 2019) and to verbal short-term memory, assessed by forward verbal span (Lervåg et al., 2009), though the latter finding was not reported consistently in previous studies (Ehm et al., 2019; Windfuhr & Snowling, 2001).

Finally, these same associative skills have also been explored in the context of learning difficulties. For example, investigations of the acquisition of speech sounds-to-symbols correspondences revealed poorer performance in dyslexic children compared to age-matched controls, especially when evaluated under time constraints or when having to further apply the learnt rules to sequences of sounds/symbols (Aravena, Snellings, &

Tijms, 2013; Aravena, Tijms, Snellings, & Molen, 2017; Law et al., 2018). Similar observations were made in children with specific language impairment (SLI) for speech sounds-to-symbols correspondences (Bishop & Hsu, 2015). In contrast, the learning of non-linguistic sounds-to-visual shapes associations has also been assessed in children with SLI compared to age-matched and grammar abilities-matched controls. The task involved picking, among four options, the fractal-like visual pattern corresponding to its associated non-speech complex sound. The learning of such non-linguistic audio-visual associations was found to be comparable among children with SLI and the two controls groups (Bishop & Hsu, 2015). These findings suggest that, rather than fundamental anomalies in the learning process itself, impairments in speech sounds' representations and/or their retention may be key in explaining associative learning deficits in children with SLI.

Overall, the existing literature has revealed that audio-visual associative learning capacities exist early in life and persist in childhood, while being potentially impaired in some learning-disabled populations. While laying the ground for further investigations, these studies have suffered from a few limitations. First of all, much of the previous work on audio-visual associative learning capacities in childhood used linguistic auditory stimuli – such as phonemes or combinations of phonemes (syllables, non-words) –, and pictures of animals or elementary shapes (circle, square, triangle) as visual stimuli. This means the used stimuli presented some degree of familiarity for participants. While it is essential to have a clear representation of the unimodal stimuli over which the associations have to be learnt, the quality or access of these representations is likely to add confounding variability in the learning of audio-visual pairings. Previous work has indeed suggested that varying levels of familiarity may influence the processes (and neural areas) involved in audio-visual binding (Hein et al., 2007; Li et al., 2016). Because we are interested in the very process of building cross-modal associations, we decided to limit as much as possible the share of variance due to pre-existing smaller or greater knowledge of the auditory and of the visual stimuli employed. Thus, our task required to pair novel environmental sounds with unknown, unfamiliar visual shapes. This choice also guarantees that variance potentially shared between our task and language-related

variables – as reported by previous studies (e.g., Karipidis et al., 2017; Windfuhr & Snowling, 2001) – cannot be attributed to the linguistic nature of the stimuli employed.

Developing audio-visual associative tasks involving non-linguistic materials is also important in the context of future studies that will examine learning-impaired populations, such as dyslexic children, whom have been shown to suffer from weak phonological skills, i.e. difficulties in representing and manipulating the sounds of speech, as measured by tasks of rhyme detection or spoonerisms (Peterson & Pennington, 2015; Ramus, 2003). Are these children impaired in linguistic audio-visual associative learning paradigms (Aravena, Snellings, & Tijms, 2013; Aravena, Tijms, Snellings, & Molen, 2017; Law et al., 2018), because of their reliance on phonological representations that are already weak – thus impacting the representation or manipulation of unimodal linguistic stimuli? Or are they suffering from core deficits in the very process of building audio-visual associations, even when non-linguistic information is used? These questions remain largely unexplored and will be interesting to address in future work, in order to better characterize the cognitive deficits that may be associated with learning difficulties, with the long-term goal of offering tailored intervention strategies.

In addition to studying cross-modal learning free of linguistic confounds, another aim of the present work was to track learning progression and inter-individual variability therein. Most audio-visual learning studies have focused on the outcome of learning, with some commendably considering how performance may change over time by comparing performance in the initial versus final stages of training (Bishop & Hsu, 2015), or separating acquisition (first phase, feedback provided) and retrieval (second phase, no feedback provided, e.g., Ehm et al., 2019). Yet most of the studies gather total scores within each phase. In the current study, we propose to examine individual variability in audio-visual associative learning capacities not only as concerns learning *outcome*, but also in terms of learning *progression* over the whole training period.

Tracking learning progression is particularly important for the general aim of finely examining transfer of learning, the major aim of this Special Issue. A dominant view in the field is that transfer results in facilitating performance in a new task. The standard procedure to assess transfer is that of first training individuals on a given task, and then asking whether the training benefits performance in a new, untrained task. A less

common view of transfer entails facilitating the *learning* of a new task, in contrast to just its performance (Bavelier, Green, Pouget, & Schrater, 2012; Bengio, 2012). In this view, when initially presented with a new task, individuals may not necessarily show any benefits from the previous training. Unless that new task be designed as a learning task, the opportunity to observe transfer may be lost. One of the goals of the present work was therefore to design a new non-linguistic audio-visual learning task that allows to track learning *as it unfolds*. To do so requires to continuously sample participants' response as participants learn novel audio-visual associations.

Therefore, in the current study, we present an audio-visual associative task that does not employ linguistic material, and examine the cognitive correlates (including linguistic abilities such as phonological awareness and executive functions skills such as visuo-spatial attention) of both learning *outcome* and learning *progression* in our task, in a population of pre-reading kindergarten children (N=76; mean age: 68 months, sd: 4 months; 39 girls; 10 left-handers).

Will the cognitive correlates of audio-visual associative learning reported in the literature, and particularly phonological awareness, be found in non-linguistic audio-visual associative learning in pre-readers? We see two alternative possibilities. If in past studies phonological awareness skills were related to audio-visual associative learning by allowing better representations/access to the linguistic auditory stimuli used in the cross-modal tasks, then they should *not* contribute to learning outcome in the task presented here. Indeed, given the specific audio-visual associative task we have designed, which does not make use of any linguistic stimuli, we hypothesize that language-related skills (i.e. phonology, vocabulary, language comprehension) will not be associated to learning in our task. We do recognize that an alternative hypothesis is possible: if however phonological awareness abilities and audio-visual learning are both linked to a common, underlying factor, they should still be related to each other, even in the absence of linguistic information within the associative task.

Finally, in the present study, by assessing visuo-spatial attention and short-term memory in the same kindergarten children, we examine the putative relation between non-linguistic associative learning and a subset of executive functions, in line with recent

literature on EFs and learning (Blair & Razza, 2007; Cardoso-Leite & Bavelier, 2014; Lipsey et al., 2017).

MATERIALS AND METHODS

Participants

A total of 107 French-speaking children (mean age: 68 months, $sd=4$) were recruited from kindergarten schools in the Paris area. To facilitate recruitment, we tested the whole class, then a-posteriori excluded beginning readers (i.e. being able to decode simple, frequent words in capital letters, $N=13$) and children non-native in French ($N=1$). Four more children refused or were not present at preschool when scheduled to perform the audio-visual associative task ($N=4$). Moreover, the data from a further subset of children ($N=7$) was not included in the final sample based on low performance in the Matrices Reasoning test, assessing non-verbal reasoning abilities (WPPSI-III published norms – score below 7, i.e. 1 standard deviation from the age-adjusted mean).

All children were in their final year of kindergarten and none of them had received formal training in reading within the school system. All procedures and information/consent forms were approved by the ethics committee of Paris-Sud University, which abides by the declaration of Helsinki.

Experimental procedure:

Children were tested individually by a neuropsychologist, in a quiet room, at each kindergarten school. Individual assessments took place over two sessions, each lasting up to 50 min. Children underwent a battery of classic, normed tests from the Bilan de Santé Evaluation du Développement pour la Scolarité 5 à 6 ans (BSEDS, Azzano et al., 2011), detailed below.

Phonological awareness

To assess phonological awareness abilities, we collected performance on two tasks, namely a syllable suppression task and a rhyming task (Zorman & Jacquier-Roux, 1999) within BSEDS. In the syllable suppression task, the child is asked to repeat a known

word, omitting one syllable (e.g. the French for rabbit, “lapin”, to be repeated omitting “pin”). All words are bisyllabic and one point is given per correct suppression, with a total of 10 points. In the rhyming task, three response options are proposed to the child and he/she has to select the word that rhymes with the target word (e.g. target word “souris”, possible options “chapeau, mari, tortue”). There are 8 trials, with correct responses counting 1 point each, for a total of 8. The score obtained in each task (correct responses in each task) were summed for further use in our analyses.

Vocabulary size

Vocabulary was estimated through the corresponding task in the BSEDS battery (Azzano et al., 2011), where it is based on a subset of items from the Test de vocabulaire actif passif pour enfants de 5 à 8 ans (TVAP. 5-8; Deltour & Hupkens, 1980). The child is presented with words in isolation of increasing difficulty as the trials unfold (first word: winter; last word: beam). In each trial, the child is requested to pick among 6 images the one that corresponds to the presented word. There are 15 trials, with 2 points for each correct response, for a maximum score of 30. Total score was considered for further analyses.

Oral comprehension skills

Similarly, the test of oral comprehension within BSEDS battery presents a subset of items from the classic comprehension test in French E.CO.S.SE (Lecocq, 1996). In each trial, the child is presented with a sentence (read by the experimenter) and he/she has to select the correct drawing among four possibilities. Sentences include relative clauses, active and passive forms. 1 point per correct trial is attributed, for a total of 10. The sum of correct responses was used in further analyses.

Rapid automatized naming (RAN)

Rapid automatized naming tasks (RAN) assess lexical access, i.e. the access and retrieval of phonological labels. Two RAN tasks were used, one based on colours and another on objects (Denckla & Rudel, 1976). The child is requested to name as quickly as possible and making as few errors as possible a series of either colours or objects, depicted on a sheet of paper, starting from the upper left corner of the sheet and ending

in the lower right. In each task, he/she needs to name a total of 25 items, displayed in 5 columns and 5 rows. Time in seconds is measured for each task, and the average time of the two tasks was considered for further analysis.

Non-word repetition

Non-word repetition tasks involve phonological processing (segmentation of the input), phonological short-term memory, phonological assembly (motor planning in order to assemble the relevant speech units), and articulation skills. The child is asked to repeat a series of non-words (5 bisyllabic and 5 trisyllabic items), presented one by one (adapted in the BSEDS from Borel-Maisonny, 1951). 1 point is given per correct answer, for a total of 10 points. The sum of correct responses was used in further analyses.

Verbal short-term memory (forward digit span)

The forward repetition of digits allows evaluating a child phonological short-term memory (STM) span. The task involves a similar set of abilities as tapped by non-word repetition tasks, although most likely loading less heavily on phonological processing and assembly, as it uses largely familiar digit names. The child is asked to repeat series of digits (task adapted within BSEDS from Wechsler Intelligence Scale for Children, (Wechsler, 2003)). Series of increasing length are presented to the child, with a minimum of 2 and a maximum of 4 digits. The maximum length managed by the child is considered as his/her final score for further analysis.

Letter recognition

The letter recognition task within BSEDS (Borel-Maisonny, 1951) loads onto letter familiarity as well as visual-spatial attention. Single, lower-case letters are presented within a rectangle on a sheet of paper. Letter strings are printed elsewhere on the same sheet and shown to the child, one by one. For each, the child is asked to point to the lower-case letters within the aforementioned rectangle, following the order shown in the letter string (for example, for the letter string 'wulo', the child is asked to point to a single 'w' first, then to a 'u', then to an 'l' and finally to an 'o'). 5 letter strings (namely cpl, wulo, abld, clbpm, olqb) are shown in total. 1 point is given per correct answer (letter correctly

identified), with the maximum score being 20 points. The sum of correct responses was used in further analyses.

Visual search

Visuo-spatial attention was measured via a barrage test within BSEDS, known as the Bells test (Gauthier, Dehaut, & Joannette, 1989). The child is presented with a sheet of paper on which several objects are drawn. His/her task is to find and cross out as many targets (drawings of bells) as possible, within 60 seconds. The total number of targets identified was considered in further analyses.

Non-verbal reasoning

To assess children non-verbal reasoning abilities, the Matrices Reasoning test (WPPSI-III) was administered (Wechsler, 2002). The child is presented with an incomplete matrix of pictures or abstract drawings and is asked to select the missing one from 4 or 5 response options. Correct responses are counted until 4 errors (in 5 consecutive trials) are observed. Total correct responses (raw performance scores) were used in final analyses.

Audio-visual association task

A novel task probing the acquisition of audio-visual associations was administered. Visual stimuli consisted of symbols, adapted from the Bamum alphabet and unknown to the children. Auditory stimuli were easily discriminable yet unfamiliar environmental sounds, difficult to label verbally – similarly to the sounds used in previous audio-visual studies in adults (Seitz, Kim, van Wassenhove, & Shams, 2007). The duration of each was 1s. They were vaguely similar to a siren (Figure 1A), to the sound of a rolling object (Figure 1B) and to an uprising melody (Figure 1C), although not straightforwardly labelled. The three to-be-learned audio-visual pairs are depicted in Figure 1.

The audio-visual task was programmed using PsychoPy v1.82 (Peirce, 2007). It was administered on a 12" screen. It started with two brief familiarization phases in order to present the auditory and the visual stimuli in isolation, then followed by the association task. The overall duration of the familiarization phases plus the audio-visual associations learning phase was approximately 30 minutes.

During the auditory familiarization, the child passively listened to the three sounds, three times each. The sounds were presented one at the time, with 5 seconds inter-stimulus interval. Coherently with the storyline described below, the child was instructed to simply listen to the bears' strange names.

A 1-back task was used for the visual familiarization, in order to ensure that the child paid attention to the stimuli throughout. One bear appeared at the time, with one of the three symbols depicted on its chest (as can be seen on the bears depicted in Figure 2). The child was instructed to press a button when he/she realized that the currently shown bear was the same as the one presented right before. Each image was displayed for 2.5 seconds, with 1 second inter-stimulus interval. In total, there were 33 pictures displayed, with 6 1-back repetitions (2 of each symbol). The sequence of pictures was the same for all participants.

Following the two familiarization phases, children underwent the audio-visual association task. The task was computerized and contained a storyline, in which the child was requested to help the three bears build a balloon and start their journey around the world. In order to be able to collaborate with them, the child had to learn each bear's name, that is, which auditory stimulus was paired with each of them.

The task was divided into four runs, each run comprising 12 trials, with a total of 48 trials. Positive generic feedback was provided at the end of each run (i.e. "Well done! You have been helping the bears a great deal! Look at how far we got in building their balloon!"). In each trial, one of the three auditory stimuli was played (1 second) while a little musical note was displayed on the screen. This was followed by 1s of blank. The three response options (as shown in Figure 2A) were then displayed, until participant's response and for a maximum of 10s. The position of the bears changed from one trial to the next, to avoid spatial learning. Corrective feedback was provided on each trial (duration = 4s), with the correct bear displaying either joy or sadness for having been picked or not, respectively (see Figure 2B), while the corresponding sound was presented again.

The audio-visual pairings were fixed, such that all participants learnt the same associations. However the order of presentation of the audio-visual pairs was randomized and thus differed between children.

Data analyses

Statistical analyses were performed in R (R Core Team, 2016). Our final analysis sample was defined as follows. For all cognitive tests but the novel audio-visual association task, standard norms based on a population of 750 French-speaking children between 5 and 6 years of age were available (BSEDS, Azzano et al., 2011). We used those norms to exclude participants falling more than 2 standard deviations below the population mean in any of the tests, as potential outliers in our analyses. This procedure resulted in the exclusion of 5 children, for low performance in non-word repetition (N=1), verbal short-term memory (N=2) and RAN (N=2) tests. Finally, one participant had to be excluded because of missing data in a few cognitive tests. Our final sample thus comprised 76 non-readers (mean age: 68 months, sd: 4 months; 39 girls). For all analyses thereafter, the raw scores of all cognitive tests were used.

For the audio-visual associative task, we computed two separate and complementary measures. Given our interest in learning outcome, we computed performance over the final trials, but as sudden drops in performance may occur due to fatigue, we took the maximum performance in run 3 and run 4 for each child. Accuracy scores (as estimated in terms of proportion correct) were not normally distributed (see Suppl. figure 1). Thus, and in line with the recommendations of Vandierendonck (Vandierendonck, 2017), performance was estimated via a Rate Correct Score (RCS), that is the proportion of correct responses divided by the average of all reaction times (RTs). RCS was therefore computed over run 3 and run 4 separately, and the maximum for each subject was considered in all further analyses.

In addition, given our interest in estimating learning progression, we fitted linear functions to individual cumulative plots of correct responses and extracted the corresponding slopes to obtain an estimate of each child's learning rate, which we term here learning progression (Suppl. figure 2).

Stepwise multiple regressions were applied, using the MASS package in R (Venables & Ripley, 2002). In a first model, learning outcome of the audio-visual association task, as quantified by RCS, was defined as the dependent variable with age and cognitive tests scores as the independent variables. In a second model, learning progression in the audio-visual association task was considered as the dependent variable, with again age and scores in cognitive tests as the independent variables. For each model, we ensured

that the assumptions of normality, linearity, non-collinearity and homoscedasticity were all met.

In addition, elastic net regressions were performed, using the glmnet package in R (Friedman, Hastie, & Tibshirani, 2010), with the same predictors as above and with RCS as dependent variable in a first model, while learning progression was the dependent variable in a second model. In both cases, a 3-fold cross-validation procedure was applied for the tuning of parameters α and λ over a grid of values. The optimal values of these parameters were then used to fit our models.

RESULTS

Descriptive Statistics – Audio-visual Association Task

Descriptive statistics for the dependent variables can be found in Table 1. The audio-visual association task did result in inter-individual variability in our sample, with RCS ranging between 0 and 0.65, and learning slope between 0.07 and 1. Correlations between age, RCS and learning progression indicated weak correlations between age and each of our 2 dependent variables (RCS – $r=0.22$, $p = 0.06.$; learning progression – $r = 0.3$, $p = 0.009$), as well as an expected more robust correlation between our two measures, RCS and learning progression ($r = 0.6$, $p < 0.001$).

Descriptive Statistics – Cognitive Tests

Descriptive statistics for the 9 cognitive variables can be found in Table 1. As mentioned above, for all analyses presented below raw scores for each task were used.

All cognitive tasks resulted in satisfactory inter-individual variability in our sample, except for Non-word repetition and Verbal short-term memory tests that were respectively near ceiling and possibly lacked enough sensitivity to show variability within the narrow age range considered, although its mean value is in the expected range.

Correlations among cognitive tests are shown in Figure 3. As expected, the outcomes of some cognitive tests were correlated, yet none of the correlation coefficients exceeded ± 0.5 .

Correlates of individual differences in AV association task learning outcome

As we are interested in identifying the cognitive factors related with learning outcome in our novel audio-visual association task, we performed a stepwise regression with learning outcome, as measured by the dependent variable Rate Correct Score. Age was entered as a predictor in the model, as well as all cognitive tests, except Non-word repetition and Verbal short-term memory tests (as they did not show enough variability in the age range assessed). The initial model included all independent variables, while the final model was obtained through both forward and backward selection based on AIC (initial model AIC=-304.36, final model AIC=-310.26).

The corresponding correlation plots for significant predictors are reported in Figure 4. As can be observed in Table 2, the results highlighted a significant relationship between Rate Correct Score and phonological awareness ($p=0.03$), and between Rate Correct Score and oral comprehension ($p=0.02$). The adjusted R^2 was 0.15.

Stepwise regressions are widely applied in the literature, yet there are known statistical limitations to this approach (Derksen & Keselman, 1992; Harrell, 2001). Thus, we applied elastic net regression to the same set of data, in order to verify that the results of the two approaches converge.

Results of elastic net regressions (i.e. elastic net coefficients) are displayed in Table 3, confirming the significant predictors obtained through stepwise regression. Letter recognition also resulted as a predictor of audio-visual RCS following elastic net regression.

Adding verbal short-term memory and non-word repetition among independent variables in the previous regression analyses did not alter the results in any way.

Correlates of individual differences in AV association learning progression

In order to evaluate the cognitive factors related with learning progression in our audio-visual association task, we performed a stepwise regression with learning progression, as measured by the dependent variable Learning Slope. Age was entered as a predictor in the model, as well as all cognitive tests, except Non-word repetition and Verbal short-term memory tests (as clarified above). The initial model included all independent variables, while the final model was obtained through both forward and backward selection based on AIC (initial model AIC=-210.93, final model AIC=-219.6). A significant relationship between audio-visual association Learning Slope and phonological

awareness was observed ($p=0.05$). The adjusted R^2 was 0.17. The corresponding correlation plot is reported in Figure 5.

Again, to complement the stepwise regression analysis, we applied elastic net regression to this set of data. Elastic net coefficients are displayed in Table 3, again confirming the significant predictor obtained through stepwise regression, namely phonological awareness. Comprehension and letter recognition were also included as predictors.

Adding verbal short-term memory and nonword repetition among independent variables did not change the results for any of the two regressions.

DISCUSSION

To characterize the key cognitive factors related to audio-visual learning, we measured both learning outcome and learning progression in a novel non-linguistic audio-visual association task. We first showed that large inter-individual differences are found in these two distinct measures among 5-year-olds. Second, we explored the relations between children learning in our task and several cognitive factors estimated by standardized tests (phonological awareness, vocabulary, oral comprehension, RAN, non-word repetition, digit repetition, letter recognition, visual search and non-verbal reasoning). Phonological awareness abilities were associated with both learning outcome and learning progression in our task. Moreover, learning outcome was also found to be related to oral comprehension.

The observation that phonological awareness is linked to performance in our novel task reproduces an association, frequently reported in the literature, between phonological awareness and the acquisition of audio-visual pairings (de Jong et al., 2000; Ehm et al., 2019; Georgiou et al., 2017; Karipidis et al., 2017; Lervåg et al., 2009). Yet all cited studies in kindergartners measured audio-visual performance in tasks comprising linguistic stimuli. Better phonological awareness could thus have advantaged children by endowing them with better pre-existing representations of the unimodal, auditory stimuli upon which the associations had to be built. In the context of our task, which did *not* involve linguistic stimuli, the reported association between phonological awareness and audio-visual associations mastery cannot be due to the linguistic nature of the auditory signals employed.

One alternative interpretation is that phonological awareness and audio-visual associative learning may be related to a common, underlying factor. Among these, several possibilities are worth considering. Finer auditory skills could for instance confer an advantage in both phonological awareness tasks (Casini, Pech-Georgel, & Ziegler, 2018; Richardson, Thomson, Scott, & Goswami, 2004) and non-linguistic audio-visual association tasks (see Thomson & Goswami, 2010 for linguistic tasks). Similarly, greater auditory attention could positively influence both phonological awareness (Yoncheva, Maurer, Zevin, & McCandliss, 2014) and non-linguistic audio-visual association pairings (Chun & Turk-browne, 2007). While the present design does allow us to assess these possibilities, our task was built to minimize the role of fine auditory skills and fine auditory attention. Indeed, the discrimination of the auditory stimuli in the audio-visual task was easy, even for kindergartners. Other factors could include domain-general ones, such as sustained attention during the tasks, or the motivation to perform these tasks. Our finding of only a specific relation between the audio-visual task and the ones tapping into phonological awareness speaks against such domain general effects. Indeed, attention and motivation would be expected to similarly impact many, if not all, of the tasks we used. A further possibility is that our experiment uncovers another shared factor between better non-linguistic audio-visual learning skills and greater phonological awareness, that is better ability to integrate auditory and visual cues. It is plausible that better audio-visual integration skills would provide enhanced access to redundant auditory and visual information within speech during development, resulting in more efficient disambiguation of noise in the speech signal. This, in turn, could lead to more robust and/or better defined phonological representations, facilitating phonological awareness skills. Enhanced integration abilities of auditory and visual abilities would thus be related to both enhanced phonological awareness skills and better audio-visual associative learning in the context of our task.

Among the many factors that contribute to developing good phonological skills, children who better acquire and master audio-visual associations as they listen to and view a speaker may indeed have a crucial advantage. It is well known that visual speech (e.g. lip-reading), as a source of complementary and redundant linguistic information, contributes significantly to disambiguating vocalizations in noisy contexts (e.g., Schwartz, Berthommier, & Savariaux, 2004). Increased attention to the mouth of the speaker, and

better audio-visual binding and learning abilities in speech, may give children a greater ability to distinguish and manipulate phonemes (Bahrack, Lickliter, & Flom, 2004; Heikkilä, Lonka, Ahola, Meronen, & Tiippana, 2017; Teinonen, Aslin, Alku, & Csibra, 2008). Infants appear to be taking into account information from the speaker's mouth during the second semester of life, as they converge on the phonetic repertoire of their native language (Lewkowicz & Hansen-Tift, 2012). At later developmental stages, visual speech decoding abilities are related to greater receptive and expressive language (Altvater-mackensen & Grossmann, 2015; Young, Merin, Rogers, & Ozonoff, 2009). In addition, it has been suggested that children with language impairments have weaker audio-visual speech integration skills and are poorer at lip-reading, while presenting less efficient speech perception (Heikkilä et al., 2017; Meronen, Tiippana, Westerholm, & Ahonen, 2013; Norrix, Plante, Vance, & Boliek, 2007; Pons, Llorenç, Sanz-torrent, & Lewkowicz, 2013), and particularly less efficient phonology (Pons, Sanz-torrent, Ferinu, & Birul, 2018). Thus, it is possible that greater audio-visual integration abilities in speech might be related to the development of phonological awareness skills. Building on the evidence that pre-verbal infants can learn arbitrary audio-visual pairings (Mersad & Dehaene-Lambertz, submitted), including non-linguistic ones (Emberson, Richards, & Aslin, 2015; Kersey & Emberson, 2017), a goal for future research will be to longitudinally assess the causal links relating audio-visual learning skills in and outside the domain of language and phonological abilities during development. A related perspective will be to implement strategies for training audio-visual integration abilities and measure the effects of such training on phonological awareness skills, an area in which so far inconsistent results have been reported (Huotilainen et al., 2011; Kujala et al., 2011).

A few studies of audio-visual associations in kindergarten have evidenced links between the outcomes of learning and rapid automatized naming (Ehm et al., 2019; Georgiou et al., 2017; Lervåg et al., 2009), yet only when oral output (verbal production) was requested in the audio-visual task, and not otherwise (Karipidis et al., 2017). Coherently, in our own study, where oral output was not requested in the audio-visual task, such relationship was absent. In several studies, phonological awareness and RAN have been shown to predict performance in a key associative audio-visual skill, that is reading – although influencing reading at different stages. Indeed, phonological awareness predicts reading accuracy mainly during the first steps of reading acquisition, while RAN is mostly

related to reading speed and remains predictive of reading performance in older children (Moll et al., 2014; Su et al., 2017). Because RAN relies on the fast recovery of the sound patterns related to non-linguistic shapes (objects or colours), it has been hypothesized to partly target audio-visual integration. In this view, it is thus surprising not to find any significant relation between RAN and learning outcome in our audio-visual task. The fact that children were not speeded up during our audio-visual task might explain why this link is missing, but it could also underline that audio-visual learning might first need an effortful stage to accurately associate sound and shape, while RAN might better capture the speed of access to an already automatized audio-visual representation, which takes several months to develop in the case of reading (Froyen, Bonte, Atteveldt, & Blomert, 2008; van Atteveldt, Blau, Blomert, & Goebel, 2010).

In contrast with the observed link with phonological awareness, the finding of a relation between oral comprehension and learning outcome in the audio-visual associative task is novel. To the best of our knowledge, none of the aforementioned studies on the acquisition of audio-visual associations in kindergarteners included any measure of oral language comprehension. Interestingly however, a few recent studies have demonstrated a strong association between verbal abilities (in particular, oral comprehension) and the acquisition of audio-visual pairings, in the context of learning to read (letter-sound associations knowledge), in typical developing kindergarteners (Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017; Hjetland et al., 2018; Storch & Whitehurst, 2002) and in kindergarten children at-risk of learning difficulties (Hulme, Nash, Gooch, Lervåg, & Snowling, 2015). The interesting observation that this relation is maintained outside the realm of linguistic audio-visual pairings would benefit from further replication.

In the present study, neither learning outcome nor learning progression was related with executive functions. This could be considered as surprising, taking into account the emerging contribution of EFs in the learning literature (Blair & Razza, 2007; Cardoso-Leite & Bavelier, 2014; Lipsey et al., 2017). Yet previous reports on audio-visual association learning in childhood, which included measures of short-term memory, working memory and/or inhibition, have yielded rather inconsistent results. Some have identified positive relations between these measures and performance in audio-visual associative tasks (Ehm et al., 2019; Lervåg et al., 2009), while others have failed to uncover any (Windfuhr & Snowling, 2001). Given the limited range of useable EF scores

in the present study, we can only comment as to a very specific skill, visuo-spatial attention. To the best of our knowledge, it has never been used in the context of audio-visual associative learning in kindergartners. It is interesting to note that a standard visual search task like the barrage task employed here did not show any relation to either learning outcome or learning progression in our association task, in 5-years-old.

A secondary aim of this work was to present an audio-visual task that dissociates learning outcome from learning progression. We acknowledge that, in the present implementation, these were more related ($r=0.6$) than they could be theoretically, as most children failed to reach asymptotic performance, given that our task was short. Yet, it remains theoretically interesting in future studies to separate these, as they may provide a richer view of transfer. Indeed, as discussed earlier, transfer is typically assessed by first training individuals on a given task, and then asking whether the training produces benefits in the performance of a new, untrained task. A less common view of transfer entails facilitating the *learning* of a new task, in contrast to just its performance. By typically limiting transfer evaluation to a direct change in performance on a new transfer task, rather than a change in learning progression, the field of training and transfer may miss opportunities to fully characterize transfer and its determinants. To the best of our knowledge, the only audio-visual associative learning task in kindergartners that tracked learning progression in the one by Karipidis et al (2017), where the authors used training duration to derive a learning rate. The present study provides a similar opportunity to measure both learning outcome and progression in a *non-linguistic* audio-visual learning task and reveals that these processes at 5 years of age are related to verbal abilities, including phonological awareness and oral comprehension skills. Whether these same correlates of learning outcome and learning progression may be maintained across different developmental age ranges, or whether other correlates may be found at later developmental stages, is a question for future investigations. A relevant example is the relation between reading comprehension and decoding skills, which is strong in beginning readers and weakens as reading progresses, while conversely oral comprehension gains greater influence (e.g., Hjetland et al., 2018).

While addressing several limitations of the previous literature, the current study comes with its own imperfections. In particular, given the relatively small sample size, replication

would be beneficial. In addition, finer assessments of executive functions, including for instance inhibition, shifting and updating tasks, would be of interest. Overall, the study also opens to several perspectives, such as evaluating the relation between linguistic and non-linguistic audio-visual associations learning at various ages, and relating these processes to real-life associative experiences such as audio-visual speech and letter/sound acquisition in reading. Finally, tasks like the one designed here will offer the opportunity of testing audio-visual associative learning across different languages, while maintaining the exact same experimental conditions.

CONCLUSION

In summary, the new audio-visual association task presented in this work allows to track both the outcome and the progression in learning associations between non-linguistic auditory signals and novel visual shapes. This task successfully captures inter-individual variability in learning among typically developing 5-year-old children. Interestingly, we could show that despite our task is relying on non-linguistic stimuli, two verbal abilities – phonological awareness and to a lesser extent oral language comprehension – explained around 15% of the variance. Our study thus contributes to the understanding of the underlying mechanisms of associative audio-visual learning in childhood, a key process for the acquisition of many developmentally relevant competences.

REFERENCES

- Altwater-mackensen, N., & Grossmann, T. (2015). Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories. *Child Development, 86*(2), 362–378. <https://doi.org/10.1111/cdev.12320>
- Aravena, S., Snellings, P., & Tijms, J. (2013). A lab-controlled simulation of a letter – speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology, 115*, 691–707. <https://doi.org/10.1016/j.jecp.2013.03.009>
- Aravena, S., Tijms, J., Snellings, P., & Molen, M. W. Van Der. (2017). Predicting Individual Differences in Reading and Spelling Skill With Artificial Script – Based Letter – Speech Sound Training. *Journal of Learning Disabilities, 1–13*. <https://doi.org/10.1177/0022219417715407>
- Azzano, V., Jacquier-roux, M., Lepaul, D., Lequette, C., Pouget, G., & Zorman, M. (2011). *Bilan de Santé Évaluation du Développement pour la scolarité 5 à 6 ans (version 4.0)*.
- Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory Redundancy Guides the Development of Selective Attention , Perception , and Cognition in Infancy. *Current Directions in Psychological Science, 13*(3), 99–102.
- Bavelier, D., Green, C., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *Annual Review of Neuroscience, 35*, 391–416. <https://doi.org/10.1146/annurev-neuro-060909>
- Bengio, Y. (2012). Deep Learning of Representations for Unsupervised and Transfer Learning. *JMLR: Workshop and Conference Proceedings, 27*, 17–37. https://doi.org/10.1007/978-3-642-36657-4_1
- Bishop, D. V. M., & Hsu, H. J. (2015). The declarative system in children with specific language impairment : a comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology, 3*(3). <https://doi.org/10.1186/s40359-015-0062-7>
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647–663.
- Borel-Maisonny, S. (1951). Les troubles du langage dans les dyslexies et les dysorthographies. *Enfance, 4*(5), 400–444.

- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2008). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*(5), 905–921. <https://doi.org/10.1162/jocn.2009.21076>
- Cardoso-Leite, P., & Bavelier, D. (2014). Video game play, attention, and learning: how to shape the development of attention and influence learning? *Current Opinion in Neurology*, *27*(2), 185–191. <https://doi.org/10.1097/WCO.0000000000000077>
- Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*(2), e12530.
- Chun, M. M., & Turk-browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*, 177–184. <https://doi.org/10.1016/j.conb.2007.03.005>
- Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79–90. <https://doi.org/10.1016/j.ecresq.2015.12.009>
- de Jong, P. F., Seveke, M.-J., & van Veen, M. (2000). Phonological Sensitivity and the Acquisition of New Words in Children. *Journal of Experimental Child Psychology*, *76*, 275–301. <https://doi.org/10.1006/jecp.1999.2549>
- Deltour, J. J., & Hupkens, D. (1980). *Test de vocabulaire passif et actif pour enfants 5-8ans*. Paris: Editions EAP.
- Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479.
- Derksen, S., & Keselman, H. G. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology*, *45*(2), 265–282.
- Ehm, J., Lonnemann, J., Brandenburg, J., Huschka, S. S., Hasselhorn, M., & Lervåg, A. (2019). Exploring factors underlying children's acquisition and retrieval of sound – symbol association skills. *Journal of Experimental Child Psychology*, *177*, 86–99. <https://doi.org/10.1016/j.jecp.2018.07.006>
- Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences of the United States of America*,

112(31). <https://doi.org/10.1073/pnas.1510343112>

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22.

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The Sleeping Infant Brain Anticipates Development. *Current Biology*, 27(15), 2374–2380. <https://doi.org/10.1016/j.cub.2017.06.070>

Froyen, D. J. W., Bonte, M. L., Atteveldt, N. Van, & Blomert, L. (2008). The Long Road to Automation: Neurocognitive Development of Letter – Speech Sound Processing. *Journal of Cognitive Neuroscience*, 21(3), 567–580.

Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The bells test: a quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology*.

Georgiou, G., Liu, C., & Xu, S. (2017). Examining the direct and indirect effects of visual – verbal paired associate learning on Chinese word reading. *Journal of Experimental Child Psychology*, 160, 81–91. <https://doi.org/10.1016/j.jecp.2017.03.011>

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.

Heikkilä, J., Lonka, E., Ahola, S., Meronen, A., & Tiippana, K. (2017). Lipreading Ability and Its Cognitive Correlates in Typically Developing Children and Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 60, 485–493.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, 14, 1–156.

Hjetland, H. N., Hagtvet, B., Lyster, S. A. H., Hulme, C., Lervåg, A., & Melby-Lervåg, M. (2018). Pathways to Reading Comprehension: A Longitudinal Study from 4 to 9 Years of Age. *Journal of Educational Psychology*.

Horbach, J., Scharke, W., Cröll, J., Heim, S., & Günther, T. (2015). Kindergarteners' performance in a sound – symbol paradigm predicts early reading. *Journal of Experimental Child Psychology*, 139, 256–264. <https://doi.org/10.1016/j.jecp.2015.06.007>

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia.

Psychological Science, 26(12), 1877–1896.

Huotilainen, M., Lovio, R., Kujala, T., Tommiska, V., Karma, K., & Fellman, V. (2011). Could audiovisual training be used to improve cognition in extremely low birth weight children? *Acta Paediatrica, International Journal of Paediatrics*, 100(11), 1489–1494. <https://doi.org/10.1111/j.1651-2227.2011.02345.x>

Karipidis, I. I., Pleisch, G., Martina, R., Hofstetter, C., Dornbierer, D., St, P., & Brem, S. (2017). Neural Initialization of Audiovisual Integration in Prereaders at Varying Risk for Developmental Dyslexia. *Human Brain Mapping*, 38, 1038–1055. <https://doi.org/10.1002/hbm.23437>

Kersey, A. J., & Emberson, L. L. (2017). Tracing trajectories of audio-visual learning in the infant brain. *Developmental Science*, 20(e12480), 1–13. <https://doi.org/10.1111/desc.12480>

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., & Naatanen, R. (2011). Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), 10509–10514. <https://doi.org/10.1073/pnas.181589198>

Law, J. M., De Vos, A., Vanderauwera, J., Wouters, J., Ghesquière, P., & Vandermosten, M. (2018). Grapheme-Phoneme Learning in an Unknown Orthography: A Study in Typical Reading and Dyslexic Children. *Frontiers in Psychology*, 9, 1–10. <https://doi.org/10.3389/fpsyg.2018.01393>

Lecocq, P. (1996). *L'E.CO.S.SE*. Villeneuve d'Ascq: Septentrion Presse Universitaire.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The Cognitive and Linguistic Foundations of Early Reading Development: A Norwegian Latent Variable Longitudinal Study. *Developmental Psychology*, 45(3), 764–781. <https://doi.org/10.1037/a0014132>

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), 1431–1436.

Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevance of selected measures. *Journal of Educational Psychology*, 109(8), 1084.

- Litt, R. A., de Jong, P. F., Bergen, E. Van, & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL – reading relationship? *Journal of Experimental Child Psychology*, 115(1), 137–149. <https://doi.org/10.1016/j.jecp.2012.11.012>
- Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, 56, 211–221.
- Mersad, K., & Dehaene-Lambertz, G. (n.d.). A precursor of reading capacity? 3-month-old infants easily learn to pair a phoneme with a visual shape. *Submitted*.
- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., ... Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, 29, 65–77. <https://doi.org/10.1016/j.learninstruc.2013.09.003>
- Norrix, L., Plante, E., Vance, R., & Boliek, C. (2007). Auditory-visual integration of speech in children with and without specific language impairment. *Journal of Speech, Language and Hearing Research*, 50, 1639–51.
- Peirce, J. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin and Review*, 21(1), 178–185. <https://doi.org/10.3758/s13423-013-0466-4>
- Pons, F., Llorenç, A., Sanz-torrent, M., & Lewkowicz, D. J. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *Journal of Child Language*, 40, 687–700. <https://doi.org/10.1017/S0305000912000189>
- Pons, F., Sanz-torrent, M., Ferinu, L., & Birul, J. (2018). Children With SLI Can Exhibit Reduced Attention to a Talker's Mouth. *Language Learning*, 68(s1), 180–192. <https://doi.org/10.1111/lang.12276>
- R Core Team, . (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing

- skills and phonological representation in dyslexic children. *Dyslexia*, 10(3), 215–233.
- Sasser, T. R., Bierman, K. L., & Heinrichs, B. (2015). Executive functioning and school adjustment: The mediational role of pre-kindergarten learning-related behaviors. *Early Childhood Research Quarterly*, 30, 70–79. <https://doi.org/10.1016/j.ecresq.2014.09.001>
- Schwartz, J., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93, 69–78. <https://doi.org/10.1016/j.cognition.2004.01.006>
- Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, 36(10), 1445–1453. <https://doi.org/10.1068/p5843>
- Soto-faraco, S., Calabresi, M., Navarra, J., Werker, J. F., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In *Multisensory Development* (pp. 207–228).
- Storch, S. A., & Whitehurst, G. J. (2002). Oral Language and Code-Related Precursors to Reading: Evidence From a Longitudinal Structural Model. *Developmental Psychology*, 38(6), 934–947. <https://doi.org/10.1037//0012-1649.38.6.934>
- Su, M., Peyre, H., Song, S., McBride, C., Tardif, T., Li, H., ... Shu, H. (2017). The influence of early linguistic skills and family factors on literacy acquisition in Chinese children: Follow-up from age 3 to age 11. *Learning and Instruction*, 49, 54–63. <https://doi.org/10.1016/j.learninstruc.2016.12.003>
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108(3), 850–855.
- Thomson, J. M., & Goswami, U. (2010). Learning novel phonological representations in developmental dyslexia: Associations with basic auditory processing of rise time and phonological awareness. *Reading and Writing*, 23(5), 453–473.
- van Atteveldt, N. M., Blau, V. C., Blomert, L., & Goebel, R. (2010). fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neuroscience*, 11, 11. <https://doi.org/10.1186/1471-2202-11-11>
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: a rejoinder on the binning procedure. *Behavioral*

Research, 49, 653–673. <https://doi.org/10.3758/s13428-016-0721-5>

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. (Fourth). New York: Springer.

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, 160, 103–109. <https://doi.org/10.1016/j.cognition.2017.01.002>

Wechsler, D. (2002). *The Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler intelligence scale for children-WISC-IV*. Psychological Corporation.

Windfuhr, K. L., & Snowling, M. J. (2001). The Relationship between Paired Associate Learning and Phonological Skills in Normally Developing Readers. *Journal of Experimental Child Psychology*, 80, 160–173. <https://doi.org/10.1006/jecp.2000.2625>

Yoncheva, Y., Maurer, U., Zevin, J., & McCandliss, B. (2014). Selective attention to phonology dynamically modulates initial encoding of auditory words within the left hemisphere. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2014.04.006>

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, 12(5), 798–814.

Zorman, M., & Jacquier-Roux, M. (1999). Évaluation de la conscience phonologique et entraînement des capacités phonologiques en grande section de maternelle. *Rééducation Orthophonique*, 197, 139–157.

Table 1. Means, standard deviations, min, max, skewness and kurtosis for all variables.

	Possible scores range	Mean	SD	Min	Max	Skewness	Kurtosis
AV association – Rate Correct Score		0.33	0.14	0	0.65	0.04	2.78
AV association – Learning Slope	0-1	0.61	0.25	0.07	1	-0.31	1.91
Phonological awareness	0-18	12.12	4.54	3	18	-0.59	2.23
Vocabulary	0-30	20.04	4.14	9	29	-0.44	2.86
Comprehension	0-10	7.24	1.79	2	10	-0.78	3.51
RAN (seconds)		63.83	14.68	40	102	0.57	2.73
Non-word repetition	0-10	9.63	0.7	7	10	-2.02	6.82
Verbal short-term memory	0-4	3.85	0.35	3	4	-1.98	5.08
Letter recognition	0-20	17.42	2.03	12	20	-0.61	2.76
Visual search	0-35	14.25	4.09	6	24	0.42	2.76
Non-verbal reasoning	0-29	15.61	3.08	11	27	1.22	4.79

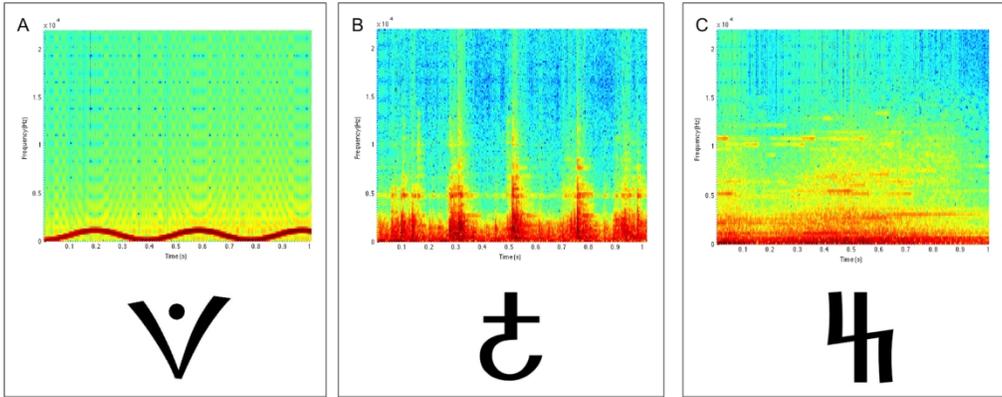
Table 2. Stepwise regressions results, showing for each predictor the standardized estimate (Stand. β), simple estimates (B), standard error in estimate (SE B) and p-values.

<i>Dependent Variable:</i>					
----------------------------	--	--	--	--	--

AV Rate Correct Score	Stand. β	B	SE B	p-value
<i>Phonological awareness</i>	0.266	0.008	0.003	.03
<i>Comprehension</i>	0.292	0.022	0.009	.02
<i>Vocabulary</i>	-0.224	-0.007	0.004	n.s.
Dependent Variable:				
AV Learning Slope				
<i>Phonological awareness</i>	0.226	0.012	0.006	.05
<i>Age</i>	0.208	0.013	0.007	n.s.
<i>Comprehension</i>	0.194	0.027	0.016	n.s.

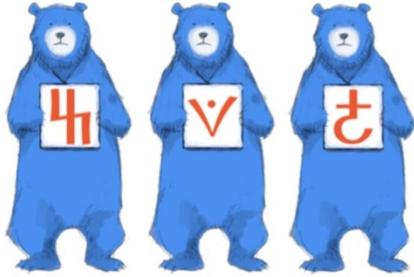
Table 3. Elastic net regressions results, showing for each predictor the corresponding coefficient (threshold > 0.001).

Dependent Variable:	Elastic net coefficients
AV Rate Correct Score	
<i>Phonological awareness</i>	0.002
<i>Comprehension</i>	0.007
<i>Letter recognition</i>	0.002
Dependent Variable:	
AV Learning Slope	
<i>Phonological awareness</i>	0.007
<i>Age</i>	0.006
<i>Comprehension</i>	0.016
<i>Letter recognition</i>	0.003

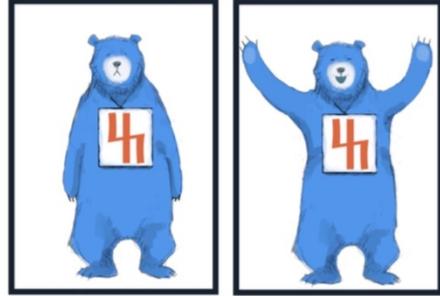


desc_12913_f1.jpg

A.



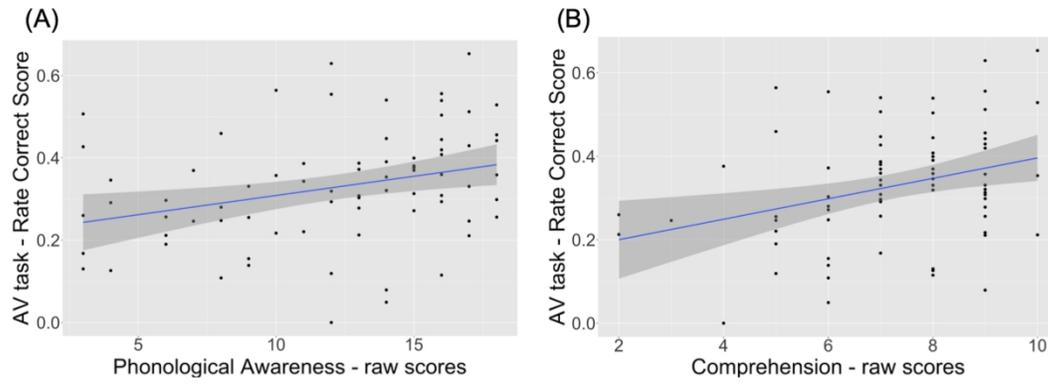
B.



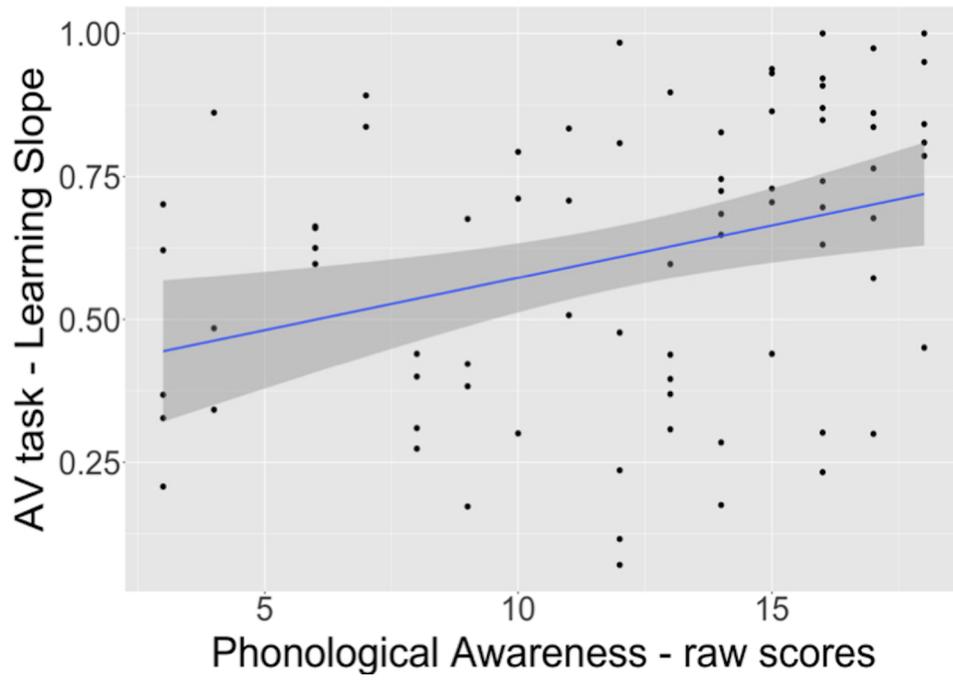
desc_12913_f2.jpg

	Age	Letter recognition	Phono Awareness	Vocabulary	Comprehension	Non-word repetition score	RAN	Non-verbal reasoning	Verbal STM	Visual search
Age	1	0.27	0.14	0.19	0.29	-0.07	-0.27	0.23	0.27	0.24
Letter recognition		1	0.26	0.09	0.39	-0.02	-0.31	0.24	0.31	0.36
Phono Awareness			1	0.31	0.38	0.15	-0.46	0.4	0.28	0.13
Vocabulary				1	0.33	0.05	-0.19	0.43	0.21	0.2
Comprehension					1	0.06	-0.26	0.31	0.45	0.2
Non-word repetition score						1	-0.06	0.01	0.1	0.04
RAN							1	-0.2	-0.2	-0.17
Non-verbal reasoning								1	0.26	0.2
Verbal STM									1	0.28
Visual search										1

desc_12913_f3.jpg



desc_12913_f4.jpg



desc_12913_f5.jpg