**The measures included in LSI\_ClaytonWestHulmeLervag.dta**

The children were tested 4 times over a period of 14 months: a) September – December, (Reception Term 1); b) January – March, (Reception Term 2); c) May – July, (Reception Term 3) and d) September – November (Year 1 Term 1). At each time-point children were individually tested in two sessions each lasting approximately 30 minutes. All testing was completed in school. There was small amount of missing data where children were absent from school. In addition, some children did not complete all tasks at each time-point due to time constraints. However, as tasks were not administered in a fixed order, data can be considered to be missing completely at random (MCAR).

## Tests and Materials

The children completed an experimental task designed to measure automatic letter-sound integration and a range of measures assessing early reading and language skills.

Letter-sound priming task. This task involved the successive presentation of a visual letter prime and an auditory phoneme target. Children were required to decide on each trial whether the second stimulus (the ‘target’) was a ‘real’ speech-sound or not. Fifty percent of trials consisted of speech sounds; the other 50% of trials involved the presentation of a non-speech sound. Response time (RT) was measured to the auditory stimuli (speech/non-speech decision RT). Figure 1 details the trial structure across the three experimental conditions.

**Stimuli.** Stimuli in this task were recordings of the 5 phonemes /tə/ (293ms), /də/ (263ms), /və/ (428ms), /zə/ (413ms) and /dʒə/ (357ms). Non-speech versions of these stimuli were created in Matlab by randomly assembling 5ms segments of the original signal (Ellis, 2010). These non-speech sounds were identical in length, energy and spectral composition but sounded completely unlike speech. The lower case letters corresponding to the phonemes used were presented in Ariel font (approximately 23 x 20mm). On 50% of trials a letter was presented and on the other 50% of trials one of five novel letter-like forms (adapted from Taylor, Plunkett, & Nation, 2011) was presented.

**Apparatus.** Stimuli were presented and responses recorded (speed and accuracy) using E-Prime Software (version 2.0) using a Psychology Software Tools Serial Response Box (SRB; model 200a) and a laptop running Windows 7. Auditory stimuli were presented through headphones.

**Procedure.** Children were instructed to attend to both the letter and speech-sound and decide whether the sound was a ‘real’ speech-sound using “yes” and “no” response keys on the response box. Before the task began children were familiarized with the procedure in thirteen practice trials.

On each trial a centrally located fixation point was presented for 1000ms, followed by the letter or non-letter stimulus, presented in black and appearing on a white screen for 500ms. The auditory target was presented over headphones and its onset was synchronous with the offset of the visual letter. Each trial was followed by the visual prompt “Real sound?” Response times from the response box were recorded from the onset of the auditory target. The experimenter monitored the child’s performance, controlling the presentation of trials.

There were six conditions in the letter-sound priming task. In the congruent condition, the prime and target were the same letter/sound. In the incongruent condition the prime and target were not the same letter/sound. In the baseline condition, the prime was a novel letter and the target was a speech-sound. There were three additional control conditions to prevent children detecting the relationship between primes and targets and generating expectancies. In these control conditions the target was a non-speech sound. Novel symbols and scrambled speech-sounds were yoked to create pseudo baseline, congruent and incongruent control conditions.

The letter-sound priming task was completed across two sessions on consecutive days to reduce attentional demands. In total there were 20 congruent and 20 incongruent trials. In the congruent condition there were four trials of each pairing and in the incongruent condition each letter prime was presented once and paired with all other speech-sounds. There were 40 baseline trials to ensure equal probability of the presentation of a novel symbol relative to a real letter prime. This resulted in 180 trials in total, including 20 ‘catch’ trials to ensure children were attending to the screen. On catch trials the same letters were presented in a black and white animal print (for example, zebra stripes) and children were instructed to make a different response (using a button on the response box).

Letter-sound knowledge. Children completed the letter-sound knowledge (LSK) subtest from the York Assessment of Reading for Comprehension (YARC; Hulme et al., 2009). This test required children to say the sound corresponding to 32 letters and digraphs.

Reading. Children completed the Early Word Recognition (EWR) subtest from the YARC (Hulme et al., 2009). This test required children to read aloud a list of words of increasing difficulty without time pressure. The maximum possible score is 30.

Phoneme awareness. Children completed the sound deletion subtest from the YARC (Hulme et al., 2009). In this test children heard a word (and saw an accompanying picture) and were required to repeat it and then repeat it again after deleting a sound (for example “Can you say seesaw? Can you say it again but this time don’t say saw?”). Practice trials ensured children understood the instructions. There were 17 items of increasing difficulty and the number of items answered correctly was recorded.

Rapid Automatised Naming (RAN). Children completed RAN subtests (colours, and digits) from the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999). Each subtest required children to name a 9 x 4 array of stimuli as quickly and accurately as possible. The time taken to name all of the items was recorded as was the number of errors (incorrect naming and/or omission of an item). Testing was discontinued if the child made four or more errors.

Receptive vocabulary. Receptive vocabulary was assessed using the British Picture Vocabulary Scale 3 (BPVS III; Dunn, Dunn & Styles, 2009). Children heard a spoken word and were asked to select the correct corresponding picture from four possible options. Following practice trials a basal set was established whereby the children made no more than one error in a set (of twelve). The children continued through sets of increasing difficulty until eight or more errors had been made in a set (ceiling). A raw score was calculated (ceiling score minus errors).

Non-verbal reasoning. The Matrix Reasoning subtest from the Weschler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was used. This subtest asks children to study a picture and to select the correct missing piece from a selection of five possible answers. Testing was discontinued when the child made four consecutive errors or four errors on five consecutive trials.

Articulation. The Articulation measure from the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, Hua, Crosbie, Holm, & Ozanne, 2002 asks children to name a series of pictures including all the consonants in British English. This measure was re-administered at subsequent time-points if children scored less than 100%.