ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Speech Categorization Consistency Predicts Language and Reading Abilities in Korean School-Age Children

Hyoju Kim,[a] Wi-Jiwoon Kim,[b] Bob McMurray,[a,c] and Dongsun Yim[b]

[a] Department of Psychological & Brain Sciences, The University of Iowa, Iowa City [b] Department of Communication Disorders, Ewha Womans University, Seoul, South Korea [c] Department of Communication Sciences and Disorders, The University of Iowa, Iowa City

ARTICLE INFO

Article History:
Received September 29, 2025
Revision received December 18, 2025
Accepted January 21, 2026

Editor-in-Chief: Jessica E. Huber
Editor: Emily Beth Myers

https://doi.org/10.1044/2026_JSLHR-25-00770

ABSTRACT

Purpose: Speech perception continues to develop throughout school age and plays a fundamental role in language and reading development. Recent findings in English-speaking children suggest that speech categorization consistency—the stability of a listener's percept across multiple encounters with a speech sound—predicts both language and reading abilities, with a particularly strong link between reading skills and vowel perception. One hypothesis is that this is due to complex grapheme–phoneme correspondences (GPCs) in English vowels. The present study tested (a) whether the relationship between categorization consistency and language/reading abilities extends to typologically different languages and (b) whether the vowel-specific link observed in English is shaped by GPC complexity, using data from Korean, a language with relatively transparent GPCs.
Method: Forty-four first-grade Korean-speaking children completed a visual analog scale task, in which they heard tokens from a speech continuum and rated the correspondence between the stimulus and each word on a continuous scale. Standardized assessments of language and word reading were also conducted.
Results: Children with poorer language/reading abilities exhibited lower categorization consistency, which is consistent with the findings from English-speaking children. In contrast to prior findings, however, this relationship was not specific to vowel perception but was held across all contrast types tested. Categorization gradiency (slope of the categorization function) was not significantly associated with any outcome.
Conclusions: These findings extend prior work by demonstrating that categorization consistency predicts language and reading abilities, even in a language with transparent GPCs. Importantly, this association was observed across all phonemic contrasts, not just vowels—suggesting that the previously observed vowel-specific link in English may stem from the greater GPC complexity of English vowels. Together, while categorization consistency appears to be a critical predictor of linguistic outcome, the specific pattern may vary across languages depending on the structure of their GPCs.
Supplemental Material: https://doi.org/10.23641/asha.31907296

A longstanding view in developmental psycholinguistics holds that speech categorization is largely established in infancy and largely through unsupervised learning mechanisms (Vallabha et al., 2007; Werker & Tees, 1984; for a review, see Werker, 2024). An infant's early commitment to native-language sound categories (and loss of ability to discriminate nonnative contrasts) has often been viewed as a major milestone—or even culmination—of a child learning the phonetic system of their language. However, a growing body of research challenges this perspective, revealing that speech categorization continues to develop well beyond infancy—through childhood, adolescence, and even across the adult lifespan (Hazan & Barrett, 2000; Kutlu

Correspondence to Hyoju Kim: hyoju-kim@uiowa.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of the publication.*

et al., 2026; McMurray et al., 2018; Nittrouer, 1992; Slawinski & Fitzgerald, 1998).

Understanding this developmental trajectory during school-age years is particularly critical, as this period is marked by rapid growth in language skills (e.g., vocabulary, phonological processing, and reading) and the increasing influence of sociocultural environments as children expand their social networks (Kutlu et al., 2024). Moreover, it is also the time when the reading instructions begin. For children learning an alphabetic language, stable and robust speech categorization may be critical for learning the mapping between sound and print, and conversely, the developmental pressures induced by the onset of reading education may prompt a reorganization of the perceptual system (particularly in alphabetic languages). Thus, a central question in developmental language science is how linguistic and extralinguistic factors jointly shape the trajectory of speech category development and, in turn, how these trajectories influence language and reading outcomes during this period.

Importantly, establishing robust speech categories is not a trivial task. Speech perception requires listeners to map a highly variable and often ambiguous speech signal onto stable phonological categories that support downstream processes like lexical access and sentence comprehension. Variability introduced by talker differences, coarticulation, and speech rate adds to this complexity (Holt & Lotto, 2010; McMurray & Jongman, 2011). These are sources of difficulty faced by adults with well-established categories—acquiring categories in the face of such variability may be doubly hard. As such, even small differences in how children navigate these complexities may have cascading effects on their broader language and reading development.

Given this complexity, it is perhaps unsurprising that speech perception difficulties are frequently observed in individuals with neurodevelopmental disorders, including developmental language disorder (DLD; Coady et al., 2005; Corriveau et al., 2007; Robertson et al., 2009; Ziegler et al., 2005) and dyslexia (or reading disorder; Robertson et al., 2009; Serniclaes et al., 2004; Werker & Tees, 1987), and individuals with brain damage (Blumstein et al., 1977; Dial et al., 2019; Wolmetz et al., 2011). Even among the neurotypical populations, speech perception abilities also vary as a function of typical development (Hazan & Barrett, 2000; Nittrouer, 2002), aging (DiNino, 2024; Toscano & Lansing, 2019), and multilingualism (Casillas, 2020; Pan et al., 2022). These underscore the broader relevance of speech perception as a core component of language processing, with implications across both typical and clinical populations.

This relevance is particularly pronounced in the context of DLD and dyslexia—two prevalent developmental disorders affecting an estimated 15%–20% of children (Adlof & Hogan, 2018; Bishop & Snowling, 2004). Although often diagnosed categorically, both are now widely viewed as reflecting the lower end of a continuous distribution of language or reading abilities (Leonard, 1987; Shaywitz et al., 1992; Tomblin & Zhang, 1999). From this perspective, an examination of how individual differences in speech perception predict language and reading outcomes can offer insight not only into the causes of these disorders but also into typical developmental trajectories and the mechanisms that support successful language and reading outcomes.

The Role of Speech Perception in Language/Reading Development

A prevailing hypothesis holds that deficits in low-level speech perception may contribute to both DLD and dyslexia, though potentially through distinct pathways. In the case of DLD, disruptions in speech categorization in the early stage may delay the development of broader language skills, such as vocabulary and sentence comprehension (Werker & Tees, 1984; Werker & Yeung, 2005). In contrast, dyslexia is often linked to difficulties in phonological processing, particularly the ability to form explicit representations of speech sounds, which is crucial for learning grapheme–phoneme correspondences (GPCs) in alphabetic languages (Ehri, 2022; Ehri et al., 2001). Here, if speech categories are not relatively robust at the onset of reading instruction, it may impede phonological processing and ultimately the ability to decode words. Thus, identifying the specific dimensions of speech perception that vary with language and reading ability may illuminate how perceptual mechanisms support different developmental pathways.

Despite extensive research, the role of speech perception in language and reading development (or DLD and dyslexia more narrowly) remains inconclusive. Studies on DLD, for example, have produced mixed results: Some report auditory or speech processing deficits (Bishop & McArthur, 2004; McArthur & Bishop, 2005), while others find no consistent impairments (e.g., Rosen, 2003). Similarly, findings on speech categorization also vary, with some studies showing group differences (e.g., Ziegler et al., 2005), others reporting null effects (e.g., Coady et al., 2005), and still others identifying deficits only for particular phonemic contrasts (e.g., Adlard & Hazan, 1998; Mody, 1993). More recent evidence suggests that difficulties in lexical processing may play a more central role in DLD than lower level perceptual impairments (Dollaghan, 2008; McMurray et al., 2014, 2019).

Parallel ambiguities exist in studies of reading ability. Some suggest that children with dyslexia exhibit speech perception deficits in quiet (e.g., Tallal, 1980) or in noise (e.g., Ziegler et al., 2009), while others find that these children struggle specifically with the categorization of ambiguous auditory input. For instance, dyslexic children do not differ in their ability to discriminate fine-grained acoustic differences but still show a shallower slope in standard speech categorization paradigms (e.g., Serniclaes et al., 2004). Yet, these effects are typically small or inconsistent (e.g., Manis et al., 1997; Werker & Tees, 1987; see Noordenbos & Serniclaes, 2015, for a review).

Conceptual and Methodological Challenges

Several factors may contribute to the inconsistencies in prior work. First, many studies often examine language or reading ability in isolation or focus on a single clinical group (DLD or dyslexia). However, reading and language are highly correlated, and DLD and dyslexia are comorbid (Catts et al., 1999, 2005; Tomblin et al., 2000). This separation limits our understanding of how speech categorization relates to both domains and can give rise to spurious inferences (e.g., a seeming effect of dyslexia that is actually driven by children with DLD). This underscores the need to examine language and reading in tandem.

Second, prior studies tended to focus on a narrow set of phonetic contrasts, such as stop voicing (e.g., /b/–/p/) and stop place (e.g., /b/–/d/), making it difficult to generalize their findings to other types of contrasts with fairly different acoustic cues, such as fricatives or vowels. Indeed, recent research suggests that individual differences in speech categorization profiles vary across phonetic contrasts (Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Myers et al., 2024). More importantly, in English, vowel perception may be particularly important for reading development due to the complex mappings between phonemes and letters. For example, the vowel /ʌ/ can be represented by multiple graphemes like O (*some*), OO (*flood*), OE (*does*), U (*mug*), and OU (*young*), and the same grapheme O corresponds to distinct vowel sounds like /oʊ/ in *comb*, /ɑ/ in *bomb*, and /u/ in *tomb*.

Finally, forced-choice tasks with speech continua, which have long been used to assess speech categorization, pose methodological and interpretive limitations. These tasks present listeners with a continuum of sounds ranging between two minimal pairs and ask them to categorize each token. They typically yield a psychometric function that features stable asymptotes at either end of the continuum (the unambiguous sounds) with a relatively rapid transition near the boundary.

Work applying such tasks to development and reading/language disorders has typically assumed—based on the outdated framework of categorical perception (Liberman et al., 1957)—that steeper categorization slopes reflect sharper, more discrete phoneme boundaries, while shallower slopes reflect increased noise or uncertainty in the system. From this perspective, children with shallow slopes are thought to have imprecise phonological categories. However, recent evidence now supports a gradient view of speech categories, in which less categorical responses may reflect flexible and adaptive processing rather than deficits (Fuhrmeister & Myers, 2021; Oden & Massaro, 1978; Schouten et al., 2003; Toscano et al., 2010; see McMurray, 2022b, for a review). This leads to a disconnect with the developmental and clinical literature, which assumes the opposite.

A critical problem is that the slope of the categorization function in a forced-choice task is inherently ambiguous (Apfelbaum et al., 2022; Honda et al., 2024). Imagine two listeners with a graded (more linear) mapping between the continuum steps and the underlying category. One might commit to a single category on every trial, always choosing whichever category is more likely on that trial. This would yield a steep (categorical) slope. In contrast, while another might vary their responses across trials, attempting to match the underlying probability (e.g., if they think a stimulus is 40% /p/-like, they choose /p/ on 40% of the trials). This would yield a shallow slope. Note that, in this case, both subjects have the same underlying perceptual mapping! Conversely, a listener with a discrete boundary could nonetheless show a shallow slope if trial-by-trial noise pushes their responses to the other side of the boundary. Thus, shallow slopes in a forced-choice task can emerge from multiple pathways, and the same perceptual system can give rise to very different slopes depending on response strategy or momentary noise.

This ambiguity is particularly problematic for developmental and clinical studies that rely heavily on slope as a diagnostic measure. If a shallow slope could reflect either adaptive gradiency or trial-by-trial noise, then interpreting slope alone risks conflating two very different phenomena. This is why it is critical to develop approaches that disentangle gradiency from noise, rather than treating slope as a unitary index.

Visual Analog Scale as an Alternative Approach

To address the conceptual and methodological limitations of forced-choice categorization tasks, researchers have increasingly turned to the visual analog scale (VAS) task (Apfelbaum et al., 2022; Kong & Edwards, 2016; Massaro & Cohen, 1983). Unlike traditional tasks that force listeners to choose between discrete categories (e.g., /b/ or /p/), the VAS task asks listeners to rate stimuli along a continuous visual scale anchored by prototypical category labels (e.g., *peach* on one end and *beach* on the other). Critically, this task distinguishes between the average psychometric function across trials (which captures something akin to the structure of long-term learned phonetic representations) and the variability around it (moment-to-moment processing stability). As such, the task offers a richer window into the cognitive and perceptual mechanisms that underlie speech categorization.

The VAS task offers several advantages. First, it is well aligned with the updated views that characterize speech categories as gradient rather than strictly discrete (McMurray, 2022b; Schouten et al., 2003; Toscano et al., 2010). Second, the continuous nature of the response format enhances psychometric sensitivity, capturing fine-grained distinctions in listeners' perception that are lost when responses are forcefully collapsed into discrete bins. Third, it yields multiple independent indices of categorization, enabling researchers to disentangle different sources of individual variability.

Three core measures can be derived from VAS data: slope, amplitude, and response variability. *Slope* of the mean categorization function reflects the steepness of the perceptual transition across the stimulus continuum. A steep slope suggests more categorical perception with a sharp boundary, while a shallower slope reflects a more gradual, gradient perceptual shift. This is thought to represent the long-term structure of the underlying phonetic categories. *Amplitude* refers to the separation between the lower and upper asymptotes of the categorization function. It reflects the extent to which a listener identifies unambiguous tokens at either end of the continuum. Lower amplitude may indicate difficulty in identifying clear category prototypes, even for acoustically extreme stimuli.

Perhaps the most theoretically novel index is *categorization consistency*. It captures the within-subject, trial-by-trial variability *around* the mean categorization function. Unlike slope or amplitude—indices that are computed *across* trials—this directly reflects the *stability* or *consistency* of categorization on a trial-by-trial basis. High variability indicates inconsistent processing of the same sound across trials, while low variability reflects stable categorization. That is, a listener with high trial-by-trial variability (low consistency) may hear the same stimulus on different trials but perceive and respond to it differently each time they hear it. This index is particularly useful for distinguishing between shallow slopes that result from internal perceptual noise versus those that reflect genuinely gradient, with stable representations.

The VAS task has demonstrated high test–retest reliability in adults (Kong & Edwards, 2016). It has been linked to a variety of downstream outcomes, including second language acquisition, speech-in-noise perception, and even neurocognitive organization (Fuhrmeister & Myers, 2021; Honda et al., 2024; Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; Myers et al., 2024). Despite its promise, the use of VAS in developmental research remains limited. Few studies have explored its application in children, and even fewer have examined its relevance for language acquisition and reading development.

Categorization Consistency as a Predictor of Language and Reading Outcomes

An expanding body of research with the VAS task suggests that categorization consistency is a robust predictor of linguistic outcomes over the traditional measure of slope. In adults, it is associated with speech perception in noise (Myers et al., 2024) and has emerged as a strong predictor of second-language learning success (Fuhrmeister et al., 2023; Honda et al., 2024; Wong et al., 2024). In adults, greater consistency is also linked to superior overall language functions (H. Kim et al., 2026), suggesting that stable perceptual encoding may be foundational not only for early development but also for lifelong language processing.

Recent evidence has further revealed that categorization consistency shows stronger correlations across different phonetic contrasts than slope, suggesting that it may reflect a more stable and generalizable individual trait in speech processing (H. Kim, McMurray, et al., 2025). This underscores the notion that capturing how consistently listeners process speech sounds may provide a more precise and developmentally relevant index of speech processing, particularly in studies focused on individual differences.

Importantly, categorization consistency has also demonstrated unique predictive power in children's language and reading abilities. In a recent large-scale study, H. Kim, Klein-Packard, et al. (2025) examined speech categorization using the VAS task alongside standardized assessments of reading, language, and cognitive regulation in 237 English-speaking school-aged children. Their findings showed that children with lower language or reading scores exhibited reduced categorization consistency, even after accounting for traditional predictors such as phonological processing and cognitive/attentional regulation (H. Kim, Klein-Packard, et al., 2025). In contrast, slope—long considered a key index of speech categorization—did not predict these outcomes. This could not be attributed to domain-general differences in attentional engagement, and the effect was not mediated by phonological processing, suggesting a rather specific locus in categorization consistency. These results suggest that categorization consistency, rather than the sharpness of category boundaries, plays a more central role in supporting successful language and reading development.

That same study also asked whether the link between categorization consistency and language/reading outcomes differed by phonetic contrasts. Interestingly, they found that reading ability was particularly associated with categorization consistency for vowels, whereas the associations of consistency with language ability were not contrast specific. This asymmetry raises important questions about the unique role of vowel perception in reading development, questions that motivate the present study.

The Unique Role of Vowel Perception in Reading

Why might stable vowel perception be particularly important for reading in English-speaking children? One possibility is that vowels pose greater perceptual challenges than consonants due to their high acoustic variability and the need to integrate multiple acoustic cues (e.g., formant structure, duration, spectral tilt), which vary across talkers and contexts. This variability may increase demands on listeners to form stable, robust processing—particularly in noisy or ambiguous environments. Under this view, early difficulties in vowel perception could contribute to downstream reading challenges. However, this account does not fully explain why vowel perception is linked specifically to reading rather than oral language skills.

Alternatively, vowels may be particularly relevant to early stages of reading due to the complexity of GPC rules. In English, GPCs are far more complex for vowels than for consonants. For example, the vowel /i/ can be spelled with EA (*team*), EE (*see*), IE (*chief*), or even E (*gene*), while the same grapheme may encode multiple vowel sounds (e.g., O in *comb, bomb, tomb*). In contrast, most consonants follow relatively straightforward one-to-one mappings (e.g., /p/ almost always corresponds to P). Learning to decode these complex vowel GPCs may require more stable perceptual representations, making vowel categorization consistency particularly important for reading performance in English-speaking children.

Cross-linguistic findings further support this view. For example, English-speaking children with dyslexia exhibit more severe reading impairments compared to their German-speaking counterparts, likely due to English's more inconsistent GPC rules (Landerl et al., 1997). If vowel perception indeed plays a critical role in reading within orthographically opaque languages, we would expect different patterns in languages with more transparent GPCs (e.g., Spanish, Italian, and Korean), where vowel spellings are more predictable. Such cross-linguistic variation underscores the need to investigate whether the relationship between vowel perception and reading development is universal or modulated by GPC complexity.

Expanding the Scope

Despite growing interest in perceptual predictors of reading, much of the current literature remains limited in two important ways. First, most studies examine reading as a categorical variable, distinguishing between clinical and nonclinical groups (e.g., dyslexic vs. typically developing rather than modeling it as a continuous outcome). However, mounting work acknowledges reading as a continuous dimension on which dyslexia or reading disorder represents the low end of the scale (Shaywitz et al., 1992). This dichotomous framing may obscure meaningful variability in how perceptual factors like categorization consistency contribute to reading skills across the full spectrum of ability.

Second, nearly all extant research has focused on languages that use Roman-based alphabetic scripts, such as English, German, French, Finnish, and Spanish. Far less attention has been devoted to language using non-Roman scripts (e.g., Chinese, Korean, Hindi, and Japanese). Understanding how perceptual consistency supports reading in these languages is critical for testing the generalizability of current models and for uncovering potential language-specific mechanisms. Finally, and most importantly, much of the work linking VAS performance to language and reading outcomes has been conducted on English, a phonologically opaque orthographic system.

The Present Study

The present study leverages the VAS task to examine speech categorization in first-grade Korean-speaking children. Building on prior work with English-speaking children, this study was designed to address the cross-linguistic generalizability of the categorization consistency effect, as well as to ask how the transparency of the GPC system shapes the relationship between speech perception and reading.

Korean provides a compelling test case for addressing these questions. Unlike English, which has a highly opaque orthographic system—particularly for vowels—Korean has a transparent GPC system. In the Korean *Hangeul* writing system, each grapheme typically maps onto a single phoneme in a systematic manner. Vowel and consonant letters are arranged into syllable blocks, but each element represents a discrete segment, and the relationships between the visual segments and phonemes (the Hangeul equivalent of a GPC) are highly predictable. This transparency offers a unique opportunity to test whether the association between reading ability and vowel categorization consistency observed in English-speaking children is driven by inherent perceptual difficulty in vowel processing or by the complexity of GPCs.

To this end, our study paralleled H. Kim, Klein-Packard, et al. (2025), testing an array of consonant and vowel continua in a VAS task and relating standard VAS indices to children's early reading and oral language skills and secondarily to nonverbal cognitive skills, which have not been assessed by prior studies. This was a convenience sample, building on existing work in Korea, so it does not offer a perfect, direct comparison to the prior work with English. However, by investigating a Korean sample using parallel VAS measures, we addressed the following research questions:

The Link Between Language/Reading and Categorization Consistency

Do children's language and reading abilities predict individual differences in speech categorization consistency in Korean, as observed in English? It is unknown whether the link between language/reading abilities and speech categorization consistency can be generalized across languages or if this pattern of effects is limited to English. By treating VAS indices as outcomes and modeling language and reading as predictors, we ask whether these effects generalize to a typologically distinct language with a transparent orthography and whether language and reading make unique versus shared contributions.

The Role of GPC Complexity in Vowel Perception and Reading

Is the association between reading ability and vowel perception modulated by the transparency of GPC rules? If vowel *perception* is inherently more difficult and underlies reading impairments, we would expect Korean-speaking children with poorer reading abilities to also show greater inconsistency in vowel categorization relative to consonants. In contrast, if the English findings stem from the *orthographic complexity* of vowel spellings, then such a vowel-specific effect should not emerge in Korean. Instead, categorization consistency should relate more broadly to reading ability across all contrast types.

The Contribution of Nonverbal Cognitive Ability

An important question is whether links between speech categorization and language/reading reflect domain-specific linguistic processes or broader domain-general cognitive abilities. H. Kim, Klein-Packard, et al. (2025) reported that an index of children's cognitive regulation did not mediate this relationship; however, their measure was based on parent report questionnaires rather than direct assessments of children's cognitive abilities. To address this limitation, the present study incorporated a standardized assessment of nonverbal intelligence quotient (NVIQ). If the association between categorization consistency and language/reading is largely explained by NVIQ, this would suggest that the observed effects are driven by domain-general cognition. Conversely, if categorization consistency predicts unique variance in language or reading after accounting for NVIQ, this would provide stronger evidence for a domain-specific contribution of speech categorization to language/reading outcomes.

Method

Participants

Forty-four first-grade Korean-speaking children (24 boys, 20 girls; M_{age} = 87.3 months, SD = 3.7) participated in the study. The children were recruited from an elementary school in Seoul, South Korea. All participants were reported by their parents or teachers to have no linguistic, physical, sensory, neurological, academic, emotional, or behavioral difficulties. They were monolingual Korean children who primarily used Korean at home, in school, and in social settings. Because the study was designed to examine continuum-based individual differences in language and reading abilities within a typically developing population, the sample was not intended to diagnose or represent children with DLD or dyslexia (though we report the relevant numbers below). Written informed consent was obtained from parents or legal guardians of all participants, and verbal or written assent was obtained from the children prior to participation. All procedures were approved by the institutional review board of Ewha Womans University (Protocol #ewha-202406-0007-01).

Assessments

Language and reading abilities were evaluated using a comprehensive battery of standardized measures. Language was assessed through multiple tests, including the Receptive and Expressive Vocabulary Test (Y. Kim et al., 2009), the Following Directions subtest of the Korean Clinical Evaluation of Language Fundamentals (K-CELF; Pae et al., 2023), the Listening Comprehension subtest from the Korean Language-based Reading Assessment (KOLRA; Pae et al., 2015), and two subtests of the Word Recall task—ordered and random word presentation formats (Chun & Yim, 2017).

In the expressive vocabulary test, children were asked to verbally name pictures presented by the examiner. In the receptive vocabulary test, they were asked to choose one picture out of four that corresponded to a word verbally provided by the examiner. In the Following Directions subtest of the K-CELF, children were required to point to a shape or a sequence of shapes within a collection, following directions such as, "Point to the first square and then the last 'X' in order." In the Listening Comprehension subtest from KOLRA, children listened to recordings of three short paragraphs and answered six questions per paragraph, which included both factual and inferential questions. In the Word Recall task, children were verbally presented with a list of three, five, or seven words, either in syntactic or random order, and were asked to immediately recall the exact same words in the same order in which they were presented. Accuracy for each task was calculated, and standard scores from each subtest were *Z*-scored and averaged to create a composite measure of language ability.[1]

Reading (decoding) was assessed using the decoding subtest of the KOLRA, which measures children's word-level decoding abilities. In this task, children were required to accurately read two-syllable real words and nonwords written in Korean. Accuracy was calculated as the number of correctly decoded items, and the resulting score was *Z*-scored for use in subsequent analyses.

Nonverbal Cognition

In addition to language and reading assessments, children's *NVIQ* was assessed using the Korean version of the Kaufman Brief Intelligence Test–Second Edition (Kaufman & Kaufman, 2013; Moon, 2020). Standard scores were *Z*-scored for subsequent analyses.

Summary of the Sample

The descriptive statistics for the standard and raw scores are presented in Table 1. Figure 1A shows the distribution of the children's language and reading scores across the sample. As expected, language and reading abilities were significantly correlated (r = .54, p < .001; see Figure 1B). The individual measures contributing to language composites were strongly intercorrelated (see Table 2), supporting the validity of the composite for language. As shown in Table 2, NVIQ was only weakly correlated with a few language measures and showed no significant association with reading decoding ability.

Although our primary analyses treat language and reading skills as continuous dimensions, we also provide a supplementary description of how many children fall below a commonly used screening threshold to assist clinical readers in interpreting the lower end of the distribution. Using a threshold of −1.25 *SD*, which is often applied as an indicator of potential concern in school-age populations, two children (4.5%) showed reading-specific weakness, one child (2.3%) showed language-specific weakness, and two children (4.5%) demonstrated co-occurring weaknesses across both domains. The remaining 39 children (~89%) performed within the broadly typical range.

[1]We used raw scores for the Word Recall task, as the standard score is not available.

Table 1. Group-level scores of each assessment.

Assessment	Mean standard score (*SD*)
Receptive vocabulary	107.8 (32.7)
Expressive vocabulary	113.5 (26.52)
K-CELF-FD	11.8 (2.78)
L-Comp.	118.6 (13.86)
WR-Ordered (raw)	54 (8.2)
WR-Random (raw)	43.3 (7.23)
KOLRA: Decoding	104.2 (12.01)

Note. K-CELF-FD = Korean Clinical Evaluation of Language Fundamentals–Following Directions; L-Comp. = Listening Comprehension; WR = Word Recall; KOLRA = Korean Language-based Reading Assessment.

Assessing Speech Categorization

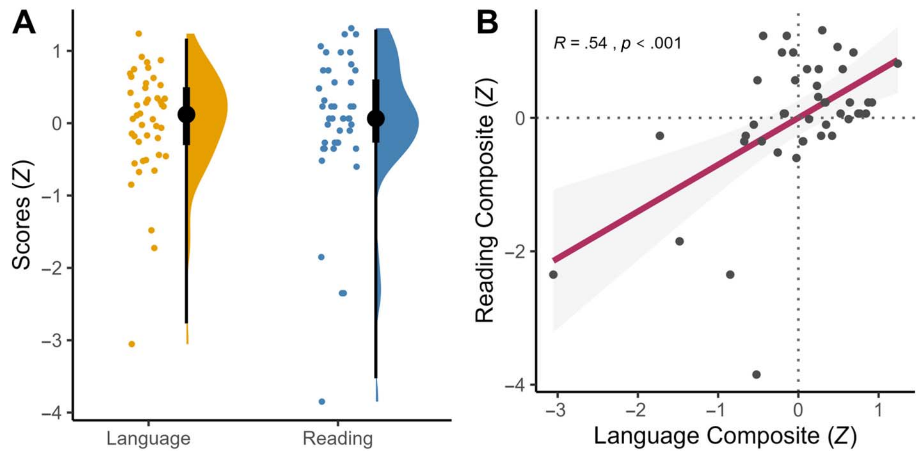
The children were tested on five speech continua, each varying between words forming a minimal pair. These included two vowel contrasts (/kʌm/ "sword"–/kɨm/ "gold"; /tal/ "moon"–/tɨl/ "field"), two stop place contrasts (/pom/ "Spring"–/kom/ "bear"; /pʰal/ "arm"–/tʰal/ "mask"), and one fricative contrast between aspirated (lenis) and unaspirated (fortis) fricatives (/sal/ "skin"–/s*al/ "rice").

Stimuli

Auditory stimuli were created from natural productions of the endpoint tokens, recorded by a native female speaker of Seoul Korean, and continua were constructed using a procedure similar to that of H. Kim, Klein-Packard, et al. (2025). Recorded tokens underwent noise reduction using Audacity and then were manipulated to form seven-step continua. For the stop place continua, we used TANDEM-STRAIGHT (Kawahara et al., 1999) to shift the formants at syllable onset—critical information to distinguish stop place—in small linear steps from one endpoint to the other. The same approach was taken for creating the vowel continua, with formants of the entire vowel portion shifted in small linear steps from one endpoint to the other.

The fricative continua varied in the duration of aspiration following the fricative and before the vowel (which is not phonemically contrastive in English but is in Korean). For this, we manipulated the relative duration of frication and aspiration using splicing. Specifically, we duplicated or deleted short segments of the frication or aspiration portions to adjust their relative lengths. In Korean, lenis fricative /s/ has a relatively long aspiration portion, whereas the fortis fricative /s*/ consists primarily of frication with minimal aspiration (Cho et al., 2002). We systematically manipulated this ratio across steps, such that the lenis fricative had a value of 1 and the fortis fricative had a value of 0. The complete set of auditory

Figure 1. Distributions and relationship between language and reading scores. (A) Violin plots showing the distribution of standardized (*Z*) scores for the language and reading composite measures. Each dot represents an individual participant; large black dots and bars represent the group mean and ± 1 *SD*, respectively. (B) Scatter plot illustrating the positive correlation between language and reading composite scores. The red line represents the linear regression line, with a shaded 95% confidence interval.



stimuli is available in the Open Science Framework repository for this study (https://osf.io/dbpc8/).

Procedure

The task was developed using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) and administered on a touchscreen tablet, Samsung Galaxy Tab S6 (Model: SM-T860), with an attached keyboard. On each trial, children saw a horizontal line centered on the screen with two endpoint words displayed at either end, presented both as written text and as corresponding clipart images (see Figure 2). After hearing an auditory stimulus over headphones, children selected a point along the line to indicate how closely the sound matched each endpoint. A vertical tick mark appeared at the selected location, and children could revise their responses before proceeding to the next trial.

Testing was blocked by a continuum, with endpoints fixed within each block to minimize the cognitive load from remapping response locations. Each continuum was tested in two sessions, allowing for counterbalancing of the image/text position across sessions within subjects. The order of the blocks and the trials within each block was randomized. In total, each child completed 140 trials: 7 steps × 5 continua × 2 repetitions × 2 sessions.

Quantifying Speech Categorization Indices From VAS

VAS data were modeled by fitting a psychometric function that related each continuum step to the participant's VAS response. The data were scaled such that higher continuum steps corresponded to higher VAS ratings, ensuring a rising logistic function.

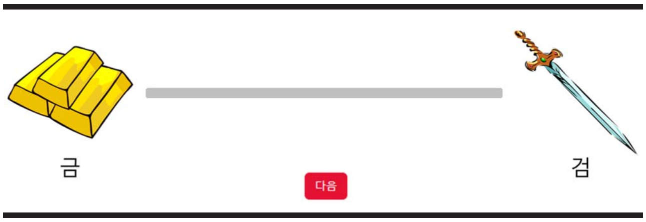
To estimate the key indices of interest, we fit a four-parameter logistic function to the VAS data separately for each subject × continuum. This function modeled the relationship between continuum step number (on the *x*-axis) and VAS response (on the *y*-axis) with four free parameters: the minimum and maximum asymptotes, the crossover

Table 2. Correlations between the assessments.

Variable		1	2	3	4	5	6	7
1	Receptive vocabulary							
2	Expressive vocabulary	*0.79*						
3	K-CELF-FD	*0.37*	0.26					
4	L-Comp.	*0.45*	*0.59*	*0.33*				
5	WR-Ordered	*0.52*	*0.44*	*0.51*	*0.55*			
6	WR-Random	*0.54*	*0.48*	*0.57*	*0.46*	*0.82*		
7	KOLRA: Decoding	*0.41*	*0.38*	*0.51*	*0.38*	*0.45*	*0.38*	
8	KBIT-2	*0.44*	0.19	*0.37*	0.26	*0.31*	0.25	0.11

Note. Values in italic indicate $p < .05$. K-CELF-FD = Korean Clinical Evaluation of Language Fundamentals–Following Directions; L-Comp. = Listening Comprehension; WR = Word Recall; KOLRA = Korean Language-based Reading Assessment; KBIT-2 = Kaufman Brief Intelligence Test–Second Edition.

Figure 2. Display screen used for the visual analog scale task.



point (i.e., the category boundary), and the slope at the crossover. Curve fitting was conducted in a custom MATLAB package (McMurray, 2017), which minimizes the least-squares error while enforcing several constraints (e.g., the fitted function must be bounded between 0 and 1, and the crossover must fall within the stimulus range). Functions were estimated individually for each subject and continuum, thereby capturing participant-specific categorization patterns.

From these fits, we derived three core indices for each participant: *slope* (as a measure of gradiency), *amplitude* (the difference between asymptotes), and *categorization consistency*. Categorization consistency was estimated by calculating the root-mean-square error between observed trial-by-trial responses and model predictions for that subject/item across continuum steps, reflecting trial-by-trial variability around the fitted function (cf. Apfelbaum et al., 2022). These values were inverted to indicate categorization consistency. These subject-level indices—gradiency, amplitude, and consistency—served as dependent variables in the subsequent analyses.

Statistical Analysis

All statistical analyses were conducted in R (Version 4.3.2; R Development Core Team, 2024). To determine how individual differences in language and reading abilities relate to speech categorization, we performed linear regressions predicting each VAS index from standardized composite scores for language and reading. Because language and reading are often correlated, we further employed a commonality analysis to partition the variance in VAS indices into portions uniquely explained by language, uniquely by reading, and jointly by both. This approach enabled us to isolate the unique contribution of each predictor, minimizing interpretive ambiguity. Commonality analysis was performed using the *yhat* package in R (Nimon et al., 2008), with VAS indices as dependent variables and language, reading, and their interaction as predictors.

To assess whether individual differences in VAS indices predict linguistic outcomes, we conducted additional regression models with language and reading as dependent variables. In these models, the three VAS indices served as predictors, allowing us to estimate the extent to which speech categorization dimensions account for variance in language or reading outcomes.

To determine whether the link between speech categorization and language or reading varies across different types of speech contrasts, we fit separate linear mixed-effects models for each predictor. In one model, categorization consistency was predicted by language ability, continuum type (fricative, stop place, vowel), and their interaction. In a separate model, consistency was predicted by reading ability, continuum type, and their interaction. Both models included random intercepts for subject and were implemented using the *lmer* function in the *lme4* package (Bates et al., 2015). These analyses allowed us to assess whether the influence of language or reading on categorization consistency was consistent across different types of phonemic contrasts or specific to certain continua.

Finally, to examine the contribution of NVIQ, we conducted multiple regression analyses with either language or reading ability as the dependent variable and included both categorization consistency and NVIQ as predictors. This approach allowed us to assess whether categorization consistency explained the unique variance in language or reading outcomes beyond domain-general cognitive ability.

Results

Overview of VAS Responses

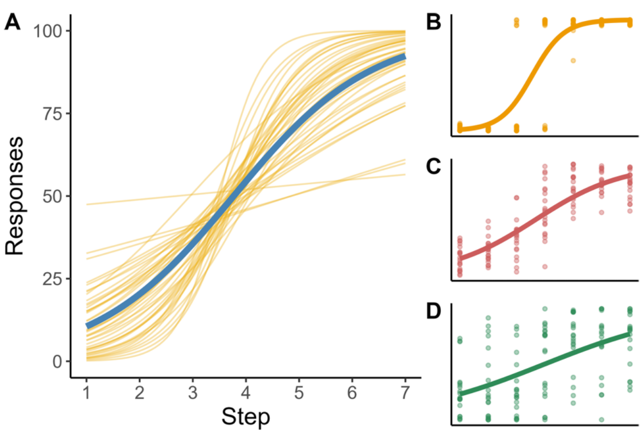
Figure 3A shows the mean (blue line) and individual (yellow lines) logistic curves of VAS ratings as a function of the continuum step. This shows the expected sigmoidal functions with low ratings (e.g., more /pom/-like) for lower continuum steps (stimuli that sound more like /pom/) and a transition to higher ratings (more /kom/-like) as the step increases.

However, these mean values overshadow subtler differences among individuals, particularly for differences in categorization consistency that cannot be seen in the means. As shown in Figures 3B–3D, individual listeners varied markedly in their gradiency and consistency. For instance, the child in Figure 3B showed a more categorical response using the extreme points of a continuous scale. The child in Figure 3C showed a gradient response similar to that in Figure 3D, but the subject in Figure 3C showed more consistent responses with ratings that more closely track the mean function.

The Effect of Language and Reading on Speech Categorization

To determine whether individual differences in speech categorization are associated with children's language and reading abilities, we conducted separate linear

Figure 3. (A) Visual analog scaling (VAS) ratings as a function of continuum step. Thin lines are logistic curves fit to individual children. A thick black line is the mean curve fit across subjects. (B) A subject who uses largely the endpoints, showing a high categorization consistency with a steep slope. (C) A subject whose consistency is high but with a shallow slope. (D) A subject who shows a similar gradiency to the subject in C but with lower consistency. For each individual plot, a line represents a logistic curve fit to a subject, and each dot indicates a trial-by-trial response.



regression models for each of the three VAS indices.[2] Each model predicted a VAS index (averaged across continua) from continuous language and reading scores (*Z*-scored) and their interaction. Summary statistics are provided in Table 3, and scatter plots of the observed relationships are shown in Figure 4.

For gradiency, the regression model did not reach statistical significance, $F(3, 40) = 1.17$, $p = .33$. Neither language ($\beta = .46$, $p = .096$) nor reading ($\beta = .06$, $p = .753$) emerged as a reliable predictor, and the interaction between the two was not significant ($\beta = .27$, $p = .15$). Similarly, for amplitude, the model failed to reach significance, $F(3, 40) = 1.61$, $p = .2$. Neither language ($\beta = -.38$, $p = .16$), reading ($\beta = .19$, $p = .33$), nor their interaction ($\beta = -.25$, $p = .19$) predicted amplitude. These findings suggest that individual differences in the overall gradiency (slope) or amplitude of categorization functions are not systematically related to children's language or reading abilities.

By contrast, the results for categorization consistency revealed a more complex and statistically significant pattern. The full model was significant, $F(3, 40) = 6.78$, $p < .001$, explaining 33.7% of the variance in categorization consistency. Neither language ($\beta = .01$, $p = .96$) nor reading ($\beta = .06$, $p = .72$) showed significant individual effects; however, their interaction was significant ($\beta = -.44$, $p < .01$). The magnitude of this interaction was moderate (partial $\eta^2 \approx .16$). Notably, addition of the interaction term increased the explained variance from 21% (main effects only) to 33.7%, indicating that the interaction accounted for roughly 12% of the variance. This unique variance indicates a moderation pattern, in which the relationship between one predictor (e.g., language) and categorization consistency depends on the level of the other predictor (e.g., reading). However, the fact that 21% of the variance is predicted by the main effects—even as neither was individually significant—suggests a pattern of suppression where the bulk of the variance in consistency is shared between language and reading.

We thus ran additional models with language and reading entered separately to further probe the overlap between the predictors. Both independently predicted categorization consistency—language alone explained 16.2% of variance in consistency ($\beta = .51$, $p < .01$), and reading alone explained 15.6% of variance in consistency ($\beta = .40$, $p < .01$). These findings suggest that each ability is associated with more stable (i.e., more consistent) speech categorization, but when both are included, their shared variance obscures these contributions.

To capture this shared variance between language and reading more precisely, we conducted a commonality analysis that decomposed the total variance explained by the model ($R^2 = .21$) into three components: variance uniquely attributable to language, uniquely attributable to reading, and jointly attributable to shared variance. The analysis showed that 4.5% was uniquely attributable to language, 5% to reading, and 11.2% to their overlap. More than half of the explained variance thus reflected shared influence, indicating the extent to which language and reading covary in shaping categorization consistency.

To further understand this interaction, we used a simple slopes analysis. For children with weaker reading skills (−1 *SD*), language ability significantly predicted categorization consistency ($\beta = .45$, $p < .05$). In contrast, the effect of language was negligible at average reading ability ($\beta = .01$, $p = .96$) and nonsignificant at higher reading ability (+1 *SD*; $\beta = -.43$, $p = .21$). A Johnson–Neyman analysis confirmed that language ability significantly predicted categorization consistency only when reading scores were below approximately −0.9 *SD*. This region reflects a relatively small portion of the sample, and we therefore interpret this moderation cautiously.

The results of this interaction analysis raise the possibility that our effects were driven by a small number of outliers. We addressed this in two ways. First, we were

[2]To examine the structure of the VAS-derived indices in our data set, we computed correlations among gradiency (slope), amplitude, and consistency. Gradiency and consistency were not significantly correlated ($r = -.006$), while amplitude showed moderate correlations with both gradiency ($r = -.60$, $p < .001$) and consistency ($r = .40$, $p < .01$). This pattern is consistent with prior reports showing that the degree of independence among VAS indices varies across contrasts and samples.

Table 3. Summary of coefficients in the regression models on visual analog scale (VAS) indices, with *Z*-scored language and reading scores as fixed effects.

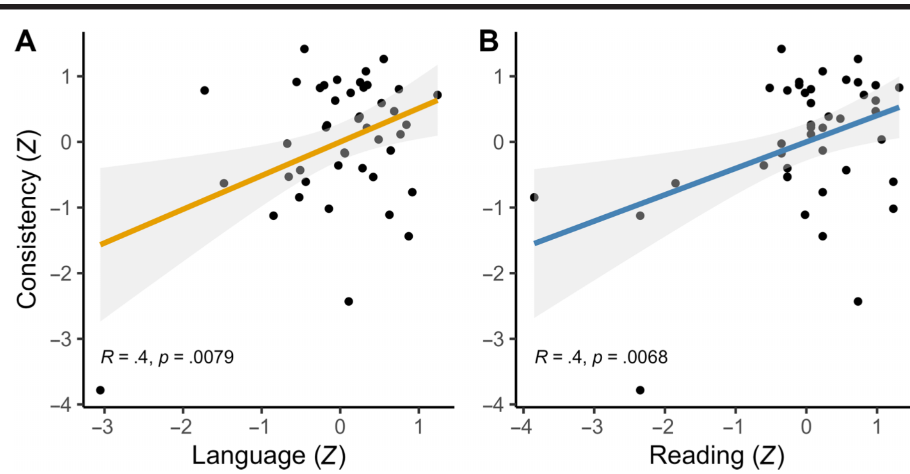
Gradiency				
Fixed effects	**Estimate**	***SE***	***t***	***Pr*(>|t|)**
(Intercept)	−0.110	0.168	−0.657	.515
Language	0.458	0.269	1.703	.096
Reading	0.063	0.199	0.317	.753
Language × Reading	0.271	0.186	1.460	.152
Consistency				
Fixed effects	**Estimate**	***SE***	***t***	***Pr*(>|t|)**
(Intercept)	0.179	0.143	1.262	.214
Language	0.011	0.228	0.049	.961
Reading	0.062	0.169	0.364	.718
Language × Reading	−0.442	0.158	−2.805	< .01
Amplitude				
Fixed effects	**Estimate**	***SE***	***t***	***Pr*(>|t|)**
(Intercept)	0.100	0.165	0.605	.549
Language	−0.381	0.265	−1.440	.158
Reading	0.194	0.197	0.987	.330
Language × Reading	−0.246	0.183	−1.344	.186

Note. R syntax: lm(VAS_Index.Z ~ Language.Z × Reading.Z). *SE* = standard error.

concerned that these low scores may represent measurement noise (in which these children should be excluded as outliers). However, as detailed in Supplemental Material S1, the children in this lower reading region show consistent multidomain weaknesses across language and reading subtests, indicating that these observations represent genuine low-ability cases rather than statistical outliers. Second, we conducted nonparametric Spearman and Kendall correlations (see Supplemental Material S2), which are robust to distributional assumptions and extreme values; these analyses showed the same directional pattern as the linear models, providing additional support for the robustness of the observed associations. Nonetheless, given their small number, replication in samples with a more balanced ability distribution will be important for confirming the robustness of this moderation, a point we return to in the general discussion.

Taken together, these results suggest that language supports consistent speech categorization primarily among children with weaker reading skills. For children with average or stronger reading skills, language ability does not explain the additional variance in categorization consistency. In summary, categorization consistency was the only VAS index systematically linked to language and reading abilities. The significant interaction shows that poorer

Figure 4. Relationship between categorization consistency and language or reading ability. (A) Greater language ability was associated with higher categorization consistency. (B) A similar positive association was observed between reading ability and consistency. All variables were *Z*-scored. Shaded regions indicate 95% confidence intervals for the regression lines.



language and reading skills predicted greater inconsistency, and this was largely shared variance between the predictors. Moreover, the contribution of language depends on reading level: Stronger language predicts stability only among children with weaker reading skills. This suggests that oral language skills may serve as a compensatory resource for stabilizing speech categorization when reading skills are still developing. Together, these findings indicate that language and reading jointly support the stability of speech categorization, but their relative contributions may shift with developmental level: Language plays a critical role when reading is weak, whereas the two skills increasingly overlap as reading develops.

The Effect of Speech Categorization on Language/Reading Abilities

To determine the degree to which the VAS indices explain individual differences in children's language and reading outcomes, we conducted regression models predicting language and reading abilities from three VAS indices (gradiency, amplitude, and categorization consistency). To maintain an adequate ratio of predictors to sample size, interaction terms were not included.

The model predicting language scores (see Table 4) was significant, $F(3, 41) = 3.38$, $p < .05$, accounting for 19.8% of the variance. Consistency emerged as the only reliable predictor ($\beta = .33$, $p < .05$), indicating that children with more stable categorization tended to have stronger language skills. Neither gradiency ($\beta = .09$, $p = .54$) nor amplitude ($\beta = -.07$, $p = .64$) was significant.

The model predicting reading ability (see Table 4) was also significant, $F(3, 40) = 2.97$, $p < .05$, explaining 18.2% of the variance. Consistency showed a marginal effect ($\beta = .33$, $p = .05$), suggesting a trend toward more consistent categorization supporting better reading outcomes. Gradiency ($\beta = .17$, $p = .39$) and amplitude ($\beta = .19$, $p = .36$) did not contribute significantly.

In summary, categorization consistency was the only VAS index that systematically predicted children's language or reading skills, emerging as a significant predictor of language ability and a marginally significant predictor of reading. Gradiency and amplitude were unrelated to either outcome. These findings converge with earlier analyses in identifying categorization consistency as the most robust index of individual differences in speech perception relevant to language and literacy development.

Continuum-Specific Effects

Figure 5 provides a visual summary of the correlations between categorization consistency and individual abilities by contrast type. For language ability (see Figure 5A), the correlations with consistency were strongest for stop place contrasts ($r = .39$, $p < .01$) and vowels ($r = .34$, $p < .05$), while the association for fricatives was weaker and not statistically significant ($r = .22$, $p = .15$). In contrast, the relationship between reading ability and categorization consistency was statistically robust across all continuum types (see Figure 5B). The strongest correlation was observed for fricatives ($r = .41$, $p < .01$), followed by vowels ($r = .34$, $p < .05$) and stop place contrasts ($r = .32$, $p < .05$).

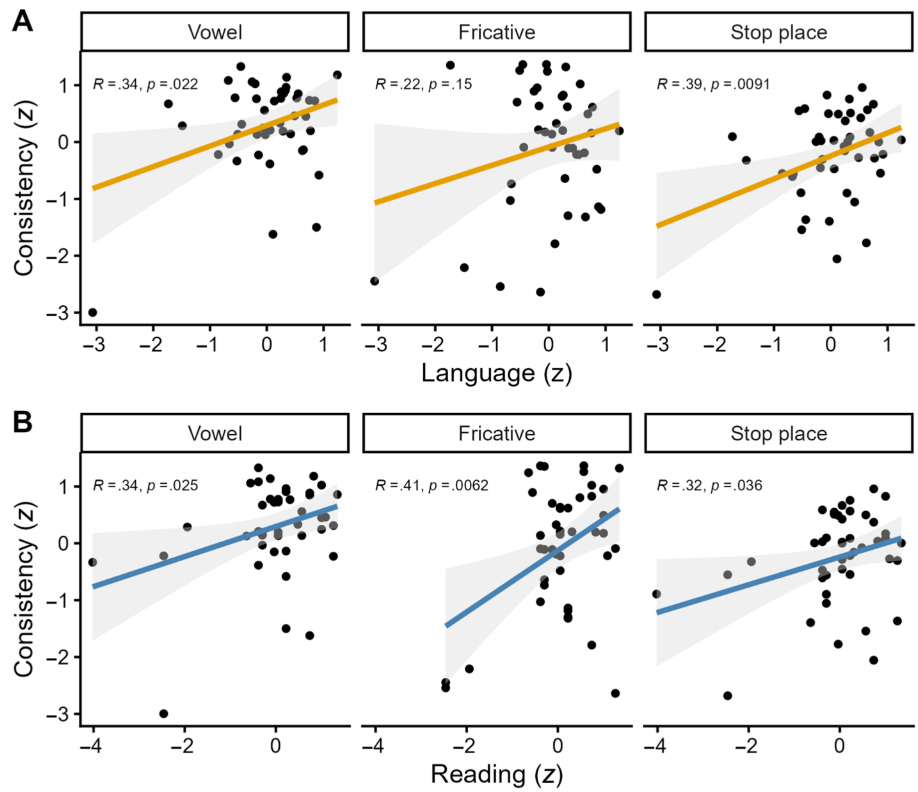
To characterize this statistically, we conducted a series of linear mixed-effects models. Individuals' categorization consistency served as the dependent variable, with either language or reading ability entered as a continuous predictor in separate models. Continuum type was modeled as a within-subject factor using dummy coding, with vowel continua serving as the reference level. The continua were coded to fricative (/s/–/s*/), stop place contrasts (/p^h/–/t^h/, /p/–/k/), and vowels (/a/–/ɨ/, /ʌ/–/ɨ/). Critically,

Table 4. Summary of coefficients in the regression models on language, reading, with gradiency, consistency, amplitude, and their interactions as fixed effects.

Language				
Fixed effects	**Estimate**	***SE***	***t***	***Pr*(>|t|)**
(Intercept)	–0.000	0.108	0.000	1.000
Gradiency	0.089	0.145	0.612	.544
Consistency	0.335	0.126	2.654	< .05
Amplitude	–0.074	0.158	–0.467	.642
Reading				
(Intercept)	–0.000	0.141	0.000	1.000
Gradiency	0.165	0.189	0.875	.387
Consistency	0.327	0.164	1.989	.054
Amplitude	0.191	0.206	0.930	.358

Note. R syntax for Language: lm(Language.Z ~ Gradiency.Z + Consistency.Z + Amplitude.Z); Reading: lm(Reading.Z ~ Gradiency.Z + Consistency.Z + Amplitude.Z). *SE* = standard error.

Figure 5. Associations between speech categorization consistency and individual differences in language and reading abilities as a function of continuum type. (A) Correlations between categorization consistency and language scores for vowel, fricative, and stop-place continua. (B) Correlations between consistency and reading scores for each continuum type. All measures are *Z*-scored. Shaded regions indicate 95% confidence intervals.



the continuum type was allowed to interact with reading (or language) to ask if any specific continuum was more (or less) related to that skill. All models included by-subject random intercepts to account for repeated observations across the continua.

The results of the mixed-effects models are summarized in Table 5. Both models revealed a significant main effect of continuum type. Compared to vowel contrasts, participants showed significantly lower consistency when categorizing fricative (language model: $\beta = -.385$, $p < .01$; reading model: $\beta = -.413$, $p < .01$) and stop contrasts (language: $\beta = -.535$, $p < .001$; reading: $\beta = -.535$, $p < .001$). These findings indicate that vowel contrasts elicited the most consistent responses, followed by fricatives and then stops, suggesting that categorization stability varies systematically by contrast type, possibly reflecting differences in acoustic properties or cue variability.

Importantly, the interaction terms between individual ability and continuum type were not significant in either model (all *p*s > .1), indicating that the relationship between consistency and language or reading skill was stable across the contrast types. Collectively, these results suggest that while the overall level of categorization consistency varies across contrast types, its association with language and reading skills is generalizable and not tied to a specific type of contrast.

The Role of Nonverbal IQ in Predicting Language/Reading Abilities

A key question in understanding the mechanisms underlying VAS responses is whether speech categorization consistency reflects domain-specific linguistic processes or more general cognitive abilities. To address this, we additionally examined the role of NVIQ in predicting language and reading outcomes and whether categorization consistency uniquely contributes to these abilities above and beyond NVIQ. As a first step, we assessed the relationship between categorization consistency and NVIQ and found that the two measures were not correlated ($r = .12$, $p = .42$), indicating that they capture largely independent sources of variance.

We then constructed a series of multiple linear regression models with either language or reading ability

Table 5. Summary of coefficients in the regression models on consistency, with language, continuum, and their interaction as fixed effects, and with reading, continuum, and their interaction as fixed effects.

Language				
Fixed effects	**Estimate**	***SE***	***t***	***Pr*(>|t|)**
(Intercept)	0.291	0.134	2.180	< .05
Language	0.362	0.175	2.073	< .05
Continuum: Fricative	−0.385	0.138	−2.785	< .01
Continuum: Stop Place	−0.535	0.137	−3.897	< .001
Language × Continuum: Fricative	−0.035	0.180	−0.195	.846
Language × Continuum: Stop Place	0.041	0.179	0.231	.818
Reading				
(Intercept)	0.295	0.130	2.265	< .05
Reading	0.264	0.127	2.083	< .05
Continuum: Fricative	−0.413	0.137	−3.021	< .01
Continuum: Stop Place	−0.535	0.135	−3.960	< .001
Reading × Continuum: Fricative	0.245	0.153	1.599	.113
Reading × Continuum: Stop Place	−0.020	0.131	−0.152	.879

Note. R syntax for Language: lmer(Consistency.Z ~ Language.Z × Continuum.Type + (1|subject)); Reading: lmer(Consistency.Z ~ Reading.Z × Continuum.Type + (1|subject)). *SE* = standard error.

as the dependent variable and with both categorization consistency (*Z*-scored) and NVIQ as predictors. This enabled us to assess the unique contribution of each predictor as well as the degree of shared variance. These models were complemented by an additional analysis in which NVIQ was added directly as a covariate to the primary regression predicting consistency from language and reading.

Language Ability

The model predicting language ability from NVIQ and categorization consistency was significant, $R^2 = .245$, $F(2, 41) = 6.66$, $p < .01$, indicating that these predictors together accounted for approximately 24.5% of the variance in language. Importantly, both predictors made statistically significant independent contributions to the model: categorization consistency: $\beta = .238$, $p < .05$, and NVIQ: $\beta = .239$, $p < .05$. A comparison with a reduced model containing only NVIQ as a predictor revealed a significant improvement in model fit when categorization consistency was added, $\Delta R^2 = .089$, $F(1, 41) = 4.83$, $p < .05$. This suggests that speech categorization consistency explains variance in language ability that cannot be fully attributed to nonverbal cognitive skills, indicating a domain-specific component to the categorization–language link.

Reading Ability

The model with both NVIQ and categorization consistency was significant, $R^2 = .162$, $F(2, 41) = 3.96$, $p < .05$. However, only categorization consistency emerged as a significant predictor, $\beta = .402$, $p < .01$, whereas NVIQ was not statistically significant, $\beta = .001$, $p = .99$. This indicates that categorization consistency is a robust and specific predictor of reading ability, independent of domain-general cognitive ability. Unlike in the language model, NVIQ did not explain any unique variance in reading performance.

Consistency as a Function of Language, Reading, and NVIQ

To further evaluate whether general cognitive ability influenced the primary analyses, we re-estimated the main regression predicting categorization consistency with language, reading, their interaction, and NVIQ included as predictors. This model remained significant, $R^2 = .378$, $F(4, 39) = 5.93$, $p < .001$, and critically, NVIQ did not significantly predict consistency ($\beta = .22$, $p = .12$). The Language × Reading interaction remained significant ($\beta = -.47$, $p < .01$), replicating the pattern observed in the original model. Thus, including NVIQ as a covariate does not alter the fundamental pattern of findings: Categorization consistency is associated with the joint profile of reading and language abilities, and this association is not reducible to general cognitive ability.

Discussion

This study asked how individual differences in speech categorization, as measured with the VAS task, relate to language and reading abilities in first-grade Korean-speaking children. Our central goal was to determine whether categorization consistency predicts language and reading outcomes in a language with transparent GPCs. Several key findings have emerged. First, our results largely replicated the findings with English-speaking children (H. Kim, Klein-Packard, et al., 2025)—categorization consistency, but not slope (gradiency) or amplitude, was systematically related to children's language and reading outcomes. Second, the link between language and categorization consistency depended on children's reading skill: Stronger language predicted more consistent categorization, primarily among weaker readers, suggesting a compensatory role for oral language when decoding skills are still immature. Third, the association between consistency and language/reading was robust across phonetic contrast types, indicating that the observed relationship was not restricted to a specific type of speech contrast. Finally, consistency explained unique variance in both language and reading after controlling for NVIQ, supporting the interpretation that consistency reflects domain-specific linguistic processes rather than domain-general cognitive ability.

Cross-Linguistic Generalization of Categorization Consistency

This study provides the first evidence that categorization consistency in speech perception is robustly associated with individual differences in language and reading abilities in Korean-speaking children. These findings directly extend prior work with English-speaking children (H. Kim, Klein-Packard, et al., 2025), suggesting that the link between categorization consistency and linguistic skills is not confined to a single language or an orthographic system. Instead, this relationship appears to reflect a broader processing trait that can be generalized across typologically and orthographically distinct languages.

This cross-linguistic replication is thus theoretically significant. While English has often been the focus of research on speech perception and literacy, its highly opaque GPCs—particularly for vowels—raise questions about the generalizability of findings to other languages. Korean, by contrast, offers a transparent and systematic orthography, particularly in the representation of vowels. The present findings thus confirm that categorization consistency is not simply a byproduct of the perceptual challenges imposed by English orthography but is likely a more fundamental component of speech processing that underpins language and reading outcomes across languages.

Moreover, the predictive relationship between categorization consistency and language/reading abilities in Korean-speaking children remained significant even after controlling for NVIQ. This extends prior work by supporting the view that speech categorization consistency reflects domain-specific aspects of linguistic processing rather than merely indexing general cognitive ability. While the NVIQ did predict language skill to a comparable degree, it explained no unique variance in reading, underscoring that categorization consistency is especially critical for reading development.

Vowel-Specific Effects Reflect the Complexity of the GPC System

A key contribution of this study is our analysis asking whether the association between categorization consistency and linguistic ability varied across vowels, stops, and fricatives. Overall, categorization consistency was the highest (best) for vowel contrasts, followed by fricatives and then stops. This is not consistent with claims that vowels are more perceptually challenging in general. Importantly, the strength of the association between consistency and language or reading ability did not differ across the contrast types. That is, although categorization consistency varied as a function of phonetic contrast, its relationship to individual differences in language and reading was stable and generalizable across contrasts.

This finding is particularly important in light of earlier research suggesting that certain speech contrasts may be more sensitive indicators of language-related deficits (e.g., Adlard & Hazan, 1998; Ziegler et al., 2005). While some prior studies observed stronger group differences for specific phoneme types (e.g., fricatives or vowels), the present data suggest that categorization consistency manifests across contrast types and is not driven solely by idiosyncratic perceptual properties of specific phonemes, and if anything, this may differ cross-linguistically.

An additional cross-linguistic factor is the size of the vowel inventory itself. Korean has only seven monophthongs, compared to the much larger and more complex vowel system of English. Korean-speaking children, therefore, confront a relatively simpler phonological system, on top of the transparent GPCs, which may further reduce the perceptual and orthographic demands of vowel processing. This simpler vowel system likely contributes to the absence of vowel-specific effects in Korean.

Critically, the pattern of results across continua is most theoretically informative when contrasted with our prior research on English-speaking children (H. Kim, Klein-Packard, et al., 2025). Here, we preface this discussion by noting that the samples differed in age (this sample tested first graders, while the English study examined third graders), in the nature of the sample (particularly the distribution of children in the lower ends of the scale), and the specific speech continua tested (since the continua tested in English were not phonologically attested in Korean). Nonetheless, with these caveats in mind, the contrast is informative.

In the English study, categorization consistency was lowest for vowel contrasts, and vowel consistency showed the strongest association with reading, raising the possibility that vowels may play a unique role in reading development in this sample. It was not clear whether that was due to the phonetic/phonological properties of vowels or the fact that these are particularly challenging in the GPC system of English. However, the current study found no evidence of a vowel-specific association: In Korean, reading ability was equally related to consistency across all contrast types, and categorization of vowels was, in fact, more consistent than other contrasts. This raises the possibility that the locus of the robust effects found in our earlier study might derive from the opacity of vowels in the reading system.

Because the present Korean study and the English data set differ in grade range and sampling context and do not include matched cross-linguistic materials, we treat this comparison as conceptual rather than as a direct test of equivalence. Nevertheless, to establish the robustness of these cross-study differences (which could derive from any of these factors), we conducted an exploratory combined analysis comparing the English and Korean data sets (see Supplemental Material S3). Using a linear mixed-effects model with reading, contrast type, and study (English vs. Korean) as predictors, we found that reading ability robustly predicted categorization consistency in both data sets, replicating the core relationship observed separately in each study. However, the overall pattern of contrast-type effects differed markedly across languages. English-speaking children showed significantly lower consistency for vowels relative to stops and fricatives, whereas Korean-speaking children showed no reliable vowel disadvantage. Furthermore, simple slope analyses indicated that, in English, reading ability predicted consistency across all contrast types, while in Korean, reading predicted consistency only for fricatives, with weaker and nonsignificant trends for vowels and stops. These cross-linguistic patterns directly demonstrate that the vowel-specific effect observed in English does not generalize to Korean.

Putting these pieces together, these findings suggest that the vowel-specific effect reported in English-speaking children might not be due to intrinsic perceptual difficulty with vowels. Rather, it likely arises from the opaque and irregular nature of vowel spellings in English. In contrast, Korean's GPCs—where vowels and consonants are represented with comparable transparency—do not place additional perceptual load on vowel processing in the service of decoding. This interpretation is consistent with cross-linguistic studies showing that children learning transparent orthographies (e.g., Finnish, Spanish) acquire decoding skills more rapidly than children learning opaque systems like English (Seymour et al., 2003). Our findings extend this work by showing that the impact of perceptual consistency on reading may be modulated by orthographic transparency, particularly for vowel processing.

The Unique Role of Consistency

The use of the VAS task allowed us to decompose categorization performance into multiple indices—gradiency (slope), amplitude, and categorization consistency. Only categorization consistency was significantly associated with linguistic abilities, echoing prior findings that categorization consistency captures individual differences in speech categorization better than traditional slope measures (H. Kim et al., 2026; H. Kim, Klein-Packard, et al., 2025; H. Kim, McMurray, et al., 2025). Importantly, this relationship held both when consistency was used to predict language/reading scores (even controlling for nonverbal IQ) and when linguistic skills were used to predict consistency, with both types of models explaining a comparable proportion of variance. This reciprocal pattern highlights that categorization consistency is not merely an outcome of language experience but may itself serve as a key perceptual precursor to language and reading development.

The absence of effects for slope and amplitude further supports growing concerns about the interpretive limitations of these traditional measures in individual differences studies (Apfelbaum et al., 2022; H. Kim et al., 2026; H. Kim, McMurray, et al., 2025). These measures, derived from across-trial averages, may obscure meaningful moment-to-moment variability that reflects the stability and robustness of underlying phonetic representations. In contrast, consistency measures directly capture how reliably a listener maps similar acoustic input onto the same phonological category in the moment.

Mechanistic Accounts of Categorization Consistency

Given that consistency appears to be a unique predictor of language/reading skills, the next question is why some children have a more consistent speech categorization. While this question remains largely unanswered, categorization consistency may reflect a balance between bottom-up encoding fidelity and top-down stabilizing mechanisms. One possible source of inconsistency is neural noise in early auditory processing, which introduces variability in how speech cues are encoded. For instance, electrophysiological studies have shown that neural response consistency—the similarity of evoked responses to repeated presentations of the same sound—indexes the stability of auditory encoding. Reduced neural consistency reflects noisier neural signals; crucially, such neural noise has been consistently linked to reading outcomes in school-age children (Hornickel & Kraus, 2013; Lam et al., 2017).

Concurrently, perceptual stability may also depend on higher level mechanisms, which exploit lexical, phonological, and orthographic knowledge to reduce noise and stabilize perceptual outcomes. Thus, consistency may emerge not only from the robustness of sensory encoding but also from the effectiveness of feedback processes that "clean up" the signal (Magnuson et al., 2024).

These clean-up mechanisms may speak to the *directionality* of the relationship between speech perception and reading. On one hand, unstable speech categorization may hinder the acquisition of the GPC system, depriving children of stable phonological anchors for decoding. On the other hand, orthographic learning may itself act as a stabilizing force: Once reliable GPC mappings are established, orthography can feed back to reinforce phonological categories, effectively serving as a clean-up mechanism.

The contrast between English and Korean illustrates this dynamic. In opaque orthographies such as English, vowel spellings are highly irregular, and thus orthography provides little stabilizing support. For struggling readers, this lack of orthographic scaffolding may exacerbate inconsistency for vowels, leading to the vowel-specific effects observed in English-speaking children. In contrast, Korean offers uniformly transparent GPCs and a smaller vowel inventory. These properties together reduce the perceptual and orthographic demands of vowel processing, which likely explains why Korean children showed no vowel-specific effects.

The Role of Nonverbal Cognitive Ability

A longstanding question is whether links between speech perception and linguistic outcomes reflect domain-specific linguistic processes or a broader domain-general cognition. The present study addressed this by directly measuring NVIQ. The results showed that both categorization consistency and NVIQ predicted language ability, whereas only consistency predicted reading ability. This pattern indicates that while both linguistic and cognitive skills contribute to oral language outcomes, reading depends more specifically on consistent phonetic encoding. These results extend prior work that relied on parent-report measures of cognitive regulation (H. Kim, Klein-Packard, et al., 2025) by showing that even when NVIQ is directly measured, speech categorization contributes unique explanatory power—particularly for reading.

Implications for Development and Individual Differences

Our work is consistent with an emerging picture that suggests that the development of speech perception is protracted throughout the school-age years (Hazan & Barrett, 2000; McMurray, 2022a; McMurray et al., 2018) and thus may critically intersect with reading. This particular dimension we highlight—categorization consistency—is all the more important given recent evidence that it is one of the critical aspects of speech categorization that develops during this time (Kutlu et al., 2026) and our prior work linking it to reading and language outcomes (H. Kim, Klein-Packard, et al., 2025). In that light, the present data offer a number of potential new insights.

Before drawing broader conclusions, however, we must consider the possibility that our effects are driven by the small number of lower performing children in our sample, and thus any developmental interpretation must be considered tentative. It was clear from our supplementary analysis (see Supplemental Material S1) that these children's low scores do not represent measurement noise, so they are not mere outliers. Without more children in the low-to-intermediate ability range, our results are consistent with several possible forms of the relationship between language/reading and speech categorization consistency.

First, it remains possible that these effects are spurious. Although this seems unlikely given robust evidence for a similar effect in English-speaking children with a much larger sample (H. Kim, Klein-Packard, et al., 2025), we cannot rule out this possibility at present. Second, the effect of language and reading may show influences over the entire scale, and we simply did not sample enough children with intermediate abilities to detect this more continuous relationship. Third, the relationship may be curvilinear—relatively flat within the typical range and falling off more steeply as children reach more extreme levels of impaired reading and language. H. Kim, Klein-Packard, et al. (2025) provide some evidence for this, as they found larger effect sizes for group-based effects (groups approximating DLD and dyslexia) than in a continuous analysis. Because our sample includes relatively few children in the intermediate-to-low range, both of the latter interpretations remain plausible. Future work should deliberately recruit a larger proportion of children with weaker literacy and language skills to better assess the stability and shape of these effects across the full ability distribution.

If we take these effects at face value, however, the moderation effect we observed—that language predicts categorization consistency primarily among weaker readers—offers suggestive implications for developmental mechanisms. One possible interpretation is that strong oral language skills may act as a compensatory resource during the early stages of reading development, scaffolding perceptual stability while decoding skills are still emerging. Once reading skills strengthen, however, language ability no longer explains additional variance in consistency, which is consistent with a developmental model in which reading gradually overlaps with and supports speech processing (Chung et al., 2024; Muter et al., 2004; Nation, 2019; Nation & Snowling, 2004). This dynamic interplay highlights the importance of studying language and reading jointly, rather than in isolation, particularly when examining their links to speech perception. Longitudinal and clinical group studies will be essential for clarifying how language and reading development jointly shape speech categorization development.

Limitations and Future Directions

Several limitations should be noted. First, as we have already noted, a critical limitation concerns the distribution of ability levels in our sample. Although our follow-up analyses (see Supplemental Materials S1 and S2) indicate that the lower performing children are not anomalous outliers but instead reflect meaningful variation at the lower end of the ability continuum, there were only a small number of such children; they could have played an outsized role in our results, and the lack of intermediate abilities in the sample makes it difficult to fully characterize the shape of the analysis.

Second, and relatedly, while the sample size was sufficient for the analyses conducted, it was modest. Larger samples are necessary to confirm the generalizability of the moderation effect and to test for more nuanced interactions between predictors. Third, the cross-sectional design limits causal inference. Longitudinal studies are essential to determine whether consistency predicts later growth in language and reading or whether these skills reciprocally shape perceptual stability over time (e.g., Kutlu et al., 2026).

Additionally, extending this work cross-linguistically remains a critical goal. The Korean findings extend the results from English, showing that categorization consistency is also a key factor in a language with transparent GPCs. However, it remains an open question of how this relationship operates in logographic systems (like Chinese) and syllabary systems (like Japanese), where the mappings between speech and text differ fundamentally from an alphabetic system. Such cross-linguistic work is essential for understanding the comprehensive aspects of how speech processing is associated with reading development across diverse writing systems.

Conclusions

The present findings provide converging evidence that speech categorization consistency is a developmentally and cross-linguistically robust predictor of language and reading outcomes. Using the VAS task, we showed that consistency—but not slope or amplitude—was systematically associated with linguistic abilities in Korean-speaking children, even after accounting for nonverbal intelligence. At the same time, cross-linguistic comparisons suggest that orthographic complexity modulates the strength and specificity of these effects. In opaque systems such as English, complex GPCs may amplify instability in categorization—particularly for vowels—making consistency especially critical for reading acquisition. By contrast, in transparent systems like Korean, consistency supports reading more broadly across contrasts. Together, these findings highlight perceptual stability as a foundational mechanism in language and reading development and underscore the value of the VAS task, with categorization consistency emerging as the most informative index for capturing individual differences.

Ethics Statement

All procedures performed in the present study were approved by the institutional research committee at Ewha Womans University (ewha-202406-0007-01).

Data Availability Statement

Data and codes for formal analyses are available at the Open Science Framework site for this project (https://osf.io/dbpc8/).

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A01096102) awarded to Dongsun Yim and the National Institutes of Health (DC008089) awarded to Bob McMurray. Part of the data presented here were presented at the 32nd Annual Meeting of the Society for the Scientific Study of Reading (SSSR 2025) in Calgary, Canada. The authors would like to thank the 14 research assistants of the Ehwa Child Language Lab Team for data collection of this study.

References

Adlard, A., & Hazan, V. (1998). Speech perception in children with specific reading difficulties (dyslexia). *The Quarterly Journal of Experimental Psychology Section A, 51*(1), 153–177. https://doi.org/10.1080/713755750

Adlof, S. M., & Hogan, T. P. (2018). Understanding dyslexia in the context of developmental language disorders. *Language, Speech, and Hearing Services in Schools, 49*(4), 762–773. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0049

Anwyl-Irvine, A. L., Massonnie, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical*

Society of America, 152(6), Article 3728. https://doi.org/10.1121/10.0015201

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bishop, D. V., & McArthur, G. M. (2004). Immature cortical responses to auditory stimuli in specific language impairment: Evidence from ERPs to rapid tone sequences. *Developmental Science, 7*(4), F11–18. https://doi.org/10.1111/j.1467-7687.2004.00356.x

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychology Bulletin, 130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15*(3), 371–383. https://doi.org/10.1016/0028-3932(77)90089-6

Casillas, J. V. (2020). Phonetic category formation is perceptually driven during the early stages of adult L2 development. *Language and Speech, 63*(3), 550–581. https://doi.org/10.1177/0023830919866225

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331–361. https://doi.org/10.1207/s1532799xssr0304_2

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193–228. https://doi.org/10.1006/jpho.2001.0153

Chun, S., & Yim, D. (2017). A comparative study of chunking mechanism in children with and without language delay. *Communication Sciences & Disorders, 22*(2), 233–244. https://doi.org/10.12963/csd.17374

Chung, H., Kim, W.-J., Yoon, Y. L., & Yim, D. (2024). Language skills and predictors of reading comprehension in first grade: Comparing good and poor comprehenders. *Journal of Speech, 33*(3), 069–080. https://doi.org/10.15724/jslhd.2024.33.3.069

Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research, 48*(4), 944–959. https://doi.org/10.1044/1092-4388(2005/065)

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research, 50*(3), 647–666. https://doi.org/10.1044/1092-4388(2007/046)

Dial, H. R., McMurray, B., & Martin, R. C. (2019). Lexical processing depends on sublexical processing: Evidence from the visual world paradigm and aphasia. *Attention, Perception, and Psychophysics, 81*(4), 1047–1064. https://doi.org/10.3758/s13414-019-01718-3

DiNino, M. (2024). Age and masking effects on acoustic cues for vowel categorization. *JASA Express Letters, 4*(6). https://doi.org/10.1121/10.0026371

Dollaghan, C. (2008). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics, 19*(2), 193–207. https://doi.org/10.1017/s0142716400010031

Ehri, L. C. (2022). What teachers need to know and do to teach letter–sounds, phonemic awareness, word reading, and phonics. *The Reading Teacher, 76*(1), 53–61. https://doi.org/10.1002/trtr.2095

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287. https://doi.org/10.1598/rrq.36.3.2

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language, 215,* Article 104919. https://doi.org/10.1016/j.bandl.2021.104919

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(7), 1161–1175. https://doi.org/10.1037/xlm0001213

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, and Psychophysics, 72*(5), 1218–1227. https://doi.org/10.3758/APP.72.5.1218

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 50*(4), 370–394. https://doi.org/10.1037/xhp0001191

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *The Journal of Neuroscience, 33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

Kapnoula, E. C., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise. *Journal of Experimental Psychology: Human Perception and Performance, 47*(4), 578–595. https://doi.org/10.1037/xhp0000900

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language, 223,* Article 105031. https://doi.org/10.1016/j.bandl.2021.105031

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594–1611. https://doi.org/10.1037/xhp0000410

Kaufman, A. S., & Kaufman, N. L. (2013). Kaufman Brief Intelligence Test, Second Edition. In *Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals*. Wiley. https://doi.org/10.1002/9781118660584.ese1325

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication, 27*(3), 187–207. https://doi.org/10.1016/S0167-6393(98)00085-5

Kim, H., Klein-Packard, J., Sorensen, E., Oleson, J., Tomblin, B., & McMurray, B. (2025). Speech categorization consistency is associated with language and reading abilities in school-age children: Implications for language and reading disorders. *Cognition, 263,* Article 106194. https://doi.org/10.1016/j.cognition.2025.106194

Kim, H., McMurray, B., Sorensen, E., & Oleson, J. (2025). The consistency of categorization-consistency in speech perception. *Psychonomic Bulletin & Review, 32*(5), 2246–2258. https://doi.org/10.3758/s13423-025-02700-x

Kim, H., Tomblin, B., & McMurray, B. (2026). *Speech categorization consistency predicts overall language abilities*. PsyArXiv. https://doi.org/10.31234/osf.io/u46pj

Kim, Y., Lee, J., Hong, G., Kim, K., & Jang, H. (2009). Validity of the receptive and expressive vocabulary test on preschool children with language delay. *Journal of Speech-Language & Hearing Disorders, 18*(1), 57–72. https://doi.org/10.15724/jslhd.2009.18.1.004

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59,* 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kutlu, E., Baxelbaum, K., Sorensen, E., Oleson, J., & McMurray, B. (2024). Linguistic diversity shapes flexible speech perception in school age children. *Scientific Reports, 14*(1), Article 28825. https://doi.org/10.1038/s41598-024-80430-1

Kutlu, E., Kim, H., & McMurray, B. (2026). Longitudinal changes in the structure of speech categorization across school age years: Children become more gradient and more consistent. *Developmental Science, 29*(1), Article e70085. https://doi.org/10.1111/desc.70085

Lam, S. S., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia, 103,* 162–167. https://doi.org/10.1016/j.neuropsychologia.2017.07.023

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition, 63*(3), 315–334. https://doi.org/10.1016/S0010-0277(97)00005-X

Leonard, L. B. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics, Vol. 1: Disorders of first-language development* (pp. 1–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511571374.002

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. https://doi.org/10.1037/h0044417

Magnuson, J. S., Crinnion, A. M., Luthra, S., Gaston, P., & Grubb, S. (2024). Contra assertions, feedback improves word recognition: How feedback and lateral inhibition sharpen signals over noise. *Cognition, 242,* Article 105661. https://doi.org/10.1016/j.cognition.2023.105661

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology, 66*(2), 211–235. https://doi.org/10.1006/jecp.1997.2383

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication, 2,* 15–35. https://doi.org/10.1016/0167-6393(83)90061-4

McArthur, G. M., & Bishop, D. V. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language, 94*(3), 260–273. https://doi.org/10.1016/j.bandl.2005.01.002

McMurray, B. (2017). *Nonlinear curvefitting for psycholinguistics (and other) data* (Version 46). https://osf.io/4atgv/overview

McMurray, B. (2022a). The acquisition of speech categories: Beyond perceptual narrowing, beyond unsupervised learning and beyond infancy. *Language, Cognition and Neuroscience, 38*(4), 419–445. https://doi.org/10.1080/23273798.2022.2105367

McMurray, B. (2022b). The myth of categorical perception. *The Journal of the Acoustical Society of America, 152*(6), 3819–3842. https://doi.org/10.1121/10.0016614

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118*(2), 219–246. https://doi.org/10.1037/a0022325

McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition, 191,* Article 104000. https://doi.org/10.1016/j.cognition.2019.06.012

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 57*(4), 1344–1362. https://doi.org/10.1044/2014_JSLHR-L-13-0196

Mody, M. (1993). *Bases of reading impairment in speech perception: A deficit in rate of auditory processing or in phonological coding?* (Publication no. 9325132) [Doctoral dissertation, City University of New York]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/304033779?sourcetype=Dissertations&Theses

Moon, S. B. (2020). *Korean Kaufman Brief Intelligence Test–Second Edition (KBIT 2)*. Hakjisa.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Myers, E., Phillips, M., & Skoe, E. (2024). Individual differences in the perception of phonetic category structure predict speech-in-noise performance. *The Journal of the Acoustical Society of America, 156*(3), 1707–1719. https://doi.org/10.1121/10.0028583

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods, 40*(2), 457–466. https://doi.org/10.3758/brm.40.2.457

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics, 20*(3), 351–382. https://doi.org/10.1016/S0095-4470(19)30639-4

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America, 112*(2), 711–719. https://doi.org/10.1121/1.1496082

Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-analysis. *Scientific Studies of Reading, 19*(5), 340–359. https://doi.org/10.1080/10888438.2015.1052455

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*(3), 172–191. https://www.ncbi.nlm.nih.gov/pubmed/663005

Pae, S., Kim, M., Yoon, H., & Jang, S. (2015). *Korean Language-based Reading Assessment (KOLRA)*. Hakjisa.

Pae, S., Yoon, H., Seol, A., & Jahng, S. (2023). *Korean Version of Clinical Evaluation of Language Fundamentals*. Hakjisa.

Pan, L., Ke, H., & Styles, S. J. (2022). Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *Scientific Reports, 12*(1), Article 4703. https://doi.org/10.1038/s41598-022-08557-7

R Development Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.2) [Computer software]. R Foundation for Statistical Computing.

Robertson, E., Joanisse, M. F., Desroches, A., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science, 12*(5), 753–767. https://doi.org/10.1111/j.1467-7687.2009.00806.x

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics, 31*(3–4), 509–527. https://doi.org/10.1016/s0095-4470(03)00046-9

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*(1), 71–80. https://doi.org/10.1016/s0167-6393(02)00094-8

Serniclaes, W., Van Heghe, S., Mousty, P., Carre, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology, 87*(4), 336–361. https://doi.org/10.1016/j.jecp.2004.02.001

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(Pt. 2), 143–174. https://doi.org/10.1348/000712603321661859

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326*(3), 145–150. https://doi.org/10.1056/nejm199201163260301

Slawinski, E. B., & Fitzgerald, L. K. (1998). Perceptual development of the categorization of the /r-w/ contrast in normal children. *Journal of Phonetics, 26*(1), 27–43. https://doi.org/10.1006/jpho.1997.0057

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9*(2), 182–198. https://doi.org/10.1016/0093-934X(80)90139-X

Tomblin, J. B., & Zhang, X. (1999). Are children with SLI a unique group of language learners? In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders: Contributions to a new framework from the cognitive neurosciences* (pp. 361–382). MIT Press.

Tomblin, J. B., Zhang, X., Buckwalter, P., & Catts, H. (2000). The association of reading disability, behavioral disorders, and language impairment among second-grade children. *The Journal of Child Psychology and Psychiatry, 41*(4), 473–482. https://doi.org/10.1111/1469-7610.00632

Toscano, J. C., & Lansing, C. R. (2019). Age-related changes in temporal and spectral cue weights in speech. *Language and Speech, 62*(1), 61–79. https://doi.org/10.1177/0023830917737112

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences, 104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Werker, J. F. (2024). Phonetic perceptual reorganization across the first year of life: Looking back. *Infant Behavior and Development, 75,* Article 101935. https://doi.org/10.1016/j.infbeh.2024.101935

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*(1), 48–61. https://doi.org/10.1037/h0084150

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences, 9*(11), 519–527. https://doi.org/10.1016/j.tics.2005.09.003

Wolmetz, M., Poeppel, D., & Rapp, B. (2011). What does the right hemisphere know about phoneme categories? *Journal of Cognitive Neuroscience, 23*(3), 552–569. https://doi.org/10.1162/jocn.2010.21495

Wong, B., Samuel, A. G., & Kapnoula, E. C. (2024). *Language-specific or universal? The nature and roles of consistency and gradiency in speech perception*. PsyArXiv. https://doi.org/10.31234/osf.io/t5zbc

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences, 102*(39), 14110–14115. https://doi.org/10.1073/pnas.0504446102

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science, 12*(5), 732–745. https://doi.org/10.1111/j.1467-7687.2009.00817.x