Sound-Category Learning and Memory Skills in Neurotypical Adults and Adults with Language-Learning Disabilities

By Allison Fulcher

An undergraduate honors thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts in University Honors and Speech and Hearing Sciences

Dr. Carolyn Quam, Ph.D., Thesis Advisor

Portland State University

2023
Introduction

The current study uses sound-category learning to examine potential factors in generalized adult language learning difficulty and language learning disability in adults.

Adult-Second Language Learning

Attempting to learn a new language as an adult is a unique challenge, as language acquisition becomes progressively more difficult with age. Later onset of learning is associated with a substantially lower likelihood of achieving nativelike perception and production of speech sounds (Abrahamsson, 2012; Díaz et al., 2012; Flege 1995; Flege, 1999) and morphosyntax (Abrahamsson, 2012; Newport 1990). This diminished language-learning ability presents a problem for adults who need to learn a second language for survival following relocation, a pressing issue, as in 2013 the United Nations High Commissioner for Refugees, António Guterres, projected that 150-200 million people will be displaced due to climate change by 2050 (Kiang, 2013). It also creates an obstacle to academic success for college students, who often need to fulfill a multiyear second-language requirement as a part of their degree. At Portland State University, for example, students pursuing a Bachelor of Arts (B.A.), or Master of Arts (M.A.) are required to complete two years of foreign language study, or else show equivalent proficiency on a placement exam. This project concentrates on two potential causes of difficulty for adult language learners that may contribute to age-based differences second language (L2) learning outcomes: changes in reliance on different memory systems and strengthening of native-language (L1) biases.

While adult language learners have been shown to have a disadvantage in learning certain language structures when compared to their infant counterparts in experimental studies (Gerken
and Knight, 2015; Gerken et al., 2019; Moreton et al., 2015), the underlying neural cause of the divergence in language-learning abilities that occurs after infancy is debated. One explanation for the developmental change is shifts in the usage of different memory systems, as memory system usage differs between infants and adults, and each memory system is thought to contribute to different aspects of language learning. Changes in two systems—declarative memory and procedural memory—both of which contribute to long-term knowledge—are at the center of this theoretical basis for age-related differences in language learning outcomes.

Declarative memory is thought to underlie explicit learning, and is responsible for encoding, storing, and retrieving semantic and episodic knowledge (Eichenbaum, 2004; Squire, 2004). Evidence suggests that declarative memory underlies lexical knowledge (Ullman, 2004). A memory can be formed via the declarative system in a single exposure to target information but is strengthened through multiple exposures (Lum et al., 2012). Procedural memory underlies implicit learning and allows for the acquisition of skills and habits that are learned over time, such as navigation (Packard, 2009), sequencing (Fletcher et al., 2005; Willingham et al., 2002), and probabilistic categorization (Poldrack et al., 2001). Evidence suggests that procedural memory underlies the knowledge and usage of syntactic rules (Ullman 2001; Ullman 2004; Ullman & Pierpont, 2005). Implicit learning is slower than explicit learning, occurring gradually and through repetition (Lum et al., 2012). In addition to the contributions of declarative and procedural memory, working memory supports language learning and understanding by maintaining information in the short-term, creating a window of time for the information to be processed (Lum et al., 2012). It also engages a specialized function to aid in comprehending and learning new words’ phonological forms, through a rehearsal process—which maintains representations of novel phonological input—and a storage process—where novel sound-patterns
are held while longer-term memories are being created—known as a phonological loop (Baddeley et al., 1998).

While these memory systems are at least partially distinct, they are highly interconnected. There may be some system redundancy, as many functions subserved by procedural memory can also be subserved by declarative memory, albeit in differing ways (Ullman, 2004). There also is evidence that working memory may be closely connected to declarative memory, as the same prefrontal neural structures that underly retrieval from declarative memory also aid working memory processing (Buckner et al., 1999; Botvinik et al., 2001; Simon & Spiers, 2003).

Where explicit learning is thought to be dominant in adults, implicit learning is dominant in infants. The neural structures that enable procedural memory (and thus implicit learning) develop in gestation, earlier than the ones that aid declarative memory (and thus explicit learning) which mature more slowly, still developing in the period between birth and 10 months (Jones & Herbert, 2006; Richmond & Nelson, 2007). Consequentially, early memory formation and learning are both subconscious and implicit. Implicit learning may provide an advantage for some aspects of language learning (Quam et al., 2015). Statistical learning, thought to play a significant role in supporting early language learning, is believed to be a form of implicit learning (Gómez, 2016). As procedural/implicit learning is thought to underlie probabilistic categorization, it may support speech-category learning, which is a complex categorization problem, requiring the perception and mapping of inconsistent acoustic input onto corresponding phonemic concepts (Holt & Lotto, 2010).

In addition to stronger reliance on explicit learning, adults also differ from infants in having developed more native language (L1) bias. Adults’ ability to absorb and comprehend
second-language (L2) speech-sound contrasts is hindered by their experience with their native language, for contrasts found in L2 but not L1 (Flege, 1995, 1999; Best et al., 2001; Iverson et al., 2003; Lotto et al., 2004; Lim & Holt, 2011; Gabay & Holt, 2015; Gabay et al., 2015). This diminished capacity is thought to be a tradeoff for increased L1 processing efficiency (Zhang et al., 2005).

**Language-Learning Disability**

For some, learning a second language as an adult is complicated further by an underlying developmental impairment specific to language learning. Language-based learning disability (LLD) is a developmental disability of unknown etiology that causes a selective impairment to language development and processing, affecting the domains of morphology (van der Lely & Ullman, 2021), syntax (van der Lely, 2005), and phonology (Sundström et al., 2018). Developmental language disorder (DLD; one name for the condition in childhood)\(^1\) impacts 7%–12% of kindergarten children (Norbury et al., 2017; Tomblin et al., 1997). Differences between the language skills and language-learning abilities of adults with a history of DLD and neurotypical adults who had typical childhood language development often persist in adulthood (Clegg et al., 2005; Hall & Tomblin, 1978; Johnson et al., 2010). Johnson et al. (1999) reported that approximately 70% of adults who were identified as having a speech or language impairment at age 5 still demonstrated language impairment upon follow-up 14 years later. Despite the prevalence of LLD in adults, most existing research has focused on children.

Limited understanding, awareness, and identification of LLD in adults reduces the

---

\(^1\) DLD (developmental language disorder) and LLD refer to the same disability, but LLD is sometimes used for adults, whereas DLD encompasses all ages. Additionally, the term for a subtype of DLD, SLI (Specific Language Impairment) is often used when describing this population.
availability and effectiveness of evidence-based clinical interventions for this population and adults with LLD’s access relevant supports and resources (e.g., through a university disability resource center). Adults with LLD are entitled to education and workplace accommodations under the Americans with Disabilities Act (1991), but may be unaware of their eligibility for services, or otherwise unable to prove their eligibility for them if there is a lack of post-childhood treatment and documentation, or if they were never diagnosed as having LLD.

In comparing the number of children receiving services for DLD versus the estimated proportion of children who have them, it becomes clear that students with language learning disabilities are generally underserved (McGregor, 2020). In addition, inequities of access and bias in assessment mean that some students are less likely to be identified and provided with services in childhood than others. While DLD has been found to be more prevalent in boys than girls, the approximate male-to-female incidence ratio is 1.3 to 1 and the corresponding ratio for receiving services is 1.71 to 1 in the United States, making girls with DLD less likely to receive services (McGregor, 2020). Disparities exist for students from minoritized groups as well, with students from non-majority linguistic and ethnic backgrounds being less likely to receive services (McGregor, 2020). This difference is particularly striking for bilingual children, for whom historic overidentification (due to misinterpretation of developing English skills) has more recently swung toward under identification. In 2017, multilingual children were 50% less likely to be evaluated as eligible for services than English-monolingual children (McGregor, 2020). Students’ perceived behavior can also influence the frequency at which they receive services, as students with self-regulation difficulties and reported behavioral issues are more likely to receive services than students who are not deemed to be disruptive (McGregor, 2020).
Can the COVIS Model be Used to Test the PDH?

Several accounts have been proposed to explain the core etiology of LLD/DLD, attributing the language impairments that people with LLD have to a variety of causes, including impaired phonological processing and mapping (Chiat, 2001), impaired working memory (Montgomery et al., 2010), and temporal auditory processing (Hill, P. R., Hogben, J. H., & Bishop, D. M. (2005). One account of the core etiology of language impairment is the Procedural Deficit Hypothesis (PDH), which theorizes that LLD can be primarily explained by impairments in procedural memory, caused by differences in the underlying brain structures (Ullman & Pierpont, 2005). The PDH predicts that the structures underlying declarative memory are not affected by LLD and therefore are able to partially compensate for impaired procedural memory function (Lum et al., 2012).

A theory that may provide a framework with which to test the PDH by tapping the memory systems for learning individually is the COVIS (Competition between Verbal and Implicit Systems) model (Ashby et al., 1998). Originating from the visual-category-learning literature, COVIS suggests that there are two learning systems, the reflective (verbal) system and the reflexive (implicit) system (Chandrasekaran et al., 2014). Reflective learning utilizes hypothesis testing, where the concept or skill being learned is refined based on self-corrective feedback stemming from external results (Chandrasekaran et al., 2014). Reflexive learning is done with less conscious awareness and self-corrective feedback than learning, and it may be difficult to verbalize the reasoning and basis of knowledge behind decisions made via reflexive learning (Chandrasekaran et al., 2014). The theory posits that rather than working collaboratively, the reflective (verbal) and reflexive (implicit) memory systems are in direct competition (Ashby et al., 1998; 2011). Despite the areas of cognition in which reflexive
learning strategies are thought to be more effective, neurotypical (NT) adults are biased to initially deploy the reflective system on a learning task, with a shift to reflexive learning requiring time and sometimes incentives (Ashby & Maddox, 2010). In addition to an inceptive bias towards reflexive learning, initial learning efforts begin with the development and testing of unidimensional rules, rather than more complex conjunctive ones that require integrating multiple dimensions (Ashby & Maddox, 2010).

This study was predicated on the notion that there might be theoretical overlap between the concepts of reflexive learning and implicit language learning, and the concepts of reflective learning and explicit language learning, and that these concepts, developed in distinct literatures, might map onto each other closely enough for the COVIS framework to be used to test the PDH’s prediction of a selective impairment to implicit learning in people with LLD. Therefore, terminology consistent with the PDH, “implicit” and “explicit” learning, will be used in place of “reflexive” and “reflective” learning from here on out. However, it is important to note that the COVIS model uses different terminology; we will revisit this discrepancy in the General Discussion.

Experiments testing the COVIS model have utilized two types of category structures, one rule-based and intended to engage more explicit learning, and the other requiring multidimensional integration (“information-integration”) and intended to engage more implicit learning (Chandrasekaran et al., 2014). The explicit rule-based category, as outlined by Chandrasekaran et al., consists of stimuli designed to be learned via hypothesis testing: consciously perceiving and applying rules that distinguish stimuli belonging to distinct categories.
The Present Study

This study aimed to assess the connections between individual memory systems, sound-discrimination skills, and sound-category learning in adults with LLD and NT adults. The study aimed to test the PDH by evaluating the contributions of different memory systems (procedural-memory, declarative-memory, and working-memory) to learning different category structures (implicit and explicit), while also assessing and accounting for the impact of baseline sound-discrimination ability on sound-categorization ability. To better simulate a second-language (L2) learning experience, we manipulated experimental sound-stimuli across two dimensions—one native and one non-native—testing participants’ ability to overcome their native language (L1) bias by incorporating both dimensions into their categorization decision-making, rather than only attending to the native-language dimension. We were interested in potential group differences in native-language bias, as well as if participants would be more successful in overcoming L1 bias in the more subconscious implicit-learning condition. We anticipated that participants with LLD may have greater difficulty shifting away from native-language bias in the implicit condition (Experiment 1 sound-category learning task) than NT group participants, as the impairment that the PDH predicts in LLD implicit learning may result in participants using compensatory explicit strategies that are ill-suited to the structure of the implicit task.

Experiment 1 utilized a sound-category-learning task with a category structure designed to be optimally learned implicitly, while Experiment 2 contained a parallel sound-category learning task with a category structure for which optimal learning would be explicit. Both category structures required integrating two sound dimensions for optimal learning, but for the structure designed to tap implicit learning, the integrative strategy was less easily verbalizable.
Methods

Participants

53 adult participants (including 10 men—1 of whom had LLD, 43 women—9 of whom had LLD, 0 who did not report) ranging in age from 19 to 53 ($M = 26.5$ years, $SD = 7.8$) completed the present study. Our original goal was to have twenty-four age and gender matched participants in each group (LLD and NT), but data collection was abruptly halted due to the onset of the COVID-19 pandemic. All study procedures for experiments 1 and 2 were approved by the Institutional Review Board (IRB) Committee at Portland State University and all participants gave written informed consent to participate.

All participants were required to pass a bilateral hearing screening at 500, 1000, 2000, and 4000 Hz at 25 dB HL and score above the range associated with intellectual disability (minimum index score of 70 + SEM) on the Test of Non-Verbal Intelligence (TONI-4) to meet the criteria for inclusion (Fidler et al., 2011). Participants filled out a questionnaire to collect demographic information and evaluate eligibility for inclusion. Participants were limited to native English speakers; bilingualism was not a basis for exclusion provided that English was learned from birth (Fidler et al., 2011; Quam et al., 2018). Participants with a history of diagnosed or suspected neurological injury or disorder were excluded (Fidler et al., 2011). Participants diagnosed with attention deficit disorder (ADD)/attention deficit hyperactive disorder (ADHD) were also excluded. Due to overlap with our population of interest, adults with LLD, participants with a history of using academic support services, receiving speech-language services in school, or having a suspected or diagnosed learning disability were not excluded. Participants with a family history of language learning disability were not excluded. Participants
who did not meet inclusion criteria on Day 1 were usually withdrawn from the study before participating in Day 2.

The test battery used to determine clinical group membership (LLD or NT)—based on the recommendations for identifying adults with LLD developed by Fidler et. al (2011) consisted of six assessments. Three subtests from the third edition of the Woodcock-Johnson Psychoeducational Battery (Woodcock, McGrew, & Mather, 2001) were used. The Letter-Word Identification subtest assessed word-identification through pronunciation. The Passage Comprehension subtest assessed reading comprehension through a fill-in-the-blank task. The Sentence Reading Fluency subtest assessed reading speed through a timed task, in which participants read basic sentences and answered a question about each sentence. The CELF-4 (Semel, Wiig, & Secord, 2003) was used to test semantic knowledge by having participants define words within varying contexts. A spelling test devised by Fidler et. al (2011) was used to assess ability to write orally presented words correctly. A modified version of the Token Test (Fidler et al., 2011)—using a prerecorded voice to produce commands, while physical tokens were employed for the physical component of the task—was used to test verbal comprehension based on ability to follow sequenced directions. The entire assessment battery took participants an estimated ninety minutes to complete and was administered in a single session. Participants were assessed individually in a quiet environment.

Fidler et. al (2011)’s methodology accomplishes participant group identification (LLD or NT) by applying a weighted scale to the results of the assessment battery. The variable weight of each assessment within the battery is proportional to the measure’s contribution to group differentiation, based on its sensitivity and specificity rates (Fidler et al., 2011). Participant results are calculated by multiplying the participant’s scores on each measure by the measure’s
associated weight and adding the constant variable (Fidler et al., 2011). A result with a positive net value identifies the participant as belonging to the LLD group, while a result with a negative net value identifies the participant as belonging to the NT group (Fidler et al., 2011). This method has been evaluated as evidence based (Nitido & Plante, 2020), as multiple studies utilizing the method have yielded satisfactory weighted composite sensitivity and specificity rates of 80% and 87% respectively (Hall, Owen Van Horne, and Farmer, 2019; McGregor et al., 2017; and Plante et al., 2017). Notably, one deviation between our method and Fidler’s is that while our recruitment materials welcomed participants with a history of language difficulty, we did not require participants to self-report a history of receiving disability services, as developmental language disability is underdiagnosed and underserved in children (McGregor, 2020), meaning that there may be many adults with LLD attending public universities who were never identified. This deviation in procedure has been used in previous work, with a similar participant population to the current study (Earle et al., 2021). Ten of the fifty-three participants who completed the assessment battery were determined to have LLD. See Table 1 for assessment scores by group.
### TABLE 1 | Assessment Battery Test Scores by Participant Group

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>NT</th>
<th>Range</th>
<th>LLD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TONI-IV</strong></td>
<td></td>
<td>104.4 (9.7)</td>
<td>87.0 to 123.0</td>
<td>96.4 (8.7)</td>
<td>88.0 to 116.0</td>
</tr>
<tr>
<td><strong>Language Identification Battery</strong></td>
<td></td>
<td>-1.50 (0.6)</td>
<td>-2.44 to -0.47</td>
<td>0.91 (0.5)</td>
<td>0.27 to 1.62</td>
</tr>
<tr>
<td><strong>Broad Reading</strong></td>
<td></td>
<td>109.1 (8.4)</td>
<td>82.0 to 125.0</td>
<td>89.9 (7.5)</td>
<td>81.0 to 101.0</td>
</tr>
<tr>
<td><strong>W-J Passage Comprehension</strong></td>
<td></td>
<td>105.0 (8.5)</td>
<td>92.0 to 126.0</td>
<td>86.3 (7.5)</td>
<td>80.0 to 103.0</td>
</tr>
<tr>
<td><strong>W-J Letter-Word Identification</strong></td>
<td></td>
<td>103.2 (13.4)</td>
<td>&lt;40 to 125</td>
<td>87.4 (10.4)</td>
<td>70 to 99</td>
</tr>
<tr>
<td><strong>W-J Reading Fluency</strong></td>
<td></td>
<td>112.6 (7.5)</td>
<td>89.0 to 131.0</td>
<td>95.8 (14.2)</td>
<td>72.0 to 118.0</td>
</tr>
<tr>
<td><strong>CELF-IV</strong></td>
<td></td>
<td>14.6 (1.1)</td>
<td>12.0 to 16.0</td>
<td>12.0 (1.9)</td>
<td>9.0 to 16.0</td>
</tr>
<tr>
<td><strong>Spelling Test</strong></td>
<td></td>
<td>11.8 (2.7)</td>
<td>5.0 to 15.0</td>
<td>5.5 (3.8)</td>
<td>0.0 to 11.0</td>
</tr>
<tr>
<td><strong>Modified Token Test</strong></td>
<td></td>
<td>41.2 (2.6)</td>
<td>36.0 to 44.0</td>
<td>31.5 (5.9)</td>
<td>22.0 to 40.0</td>
</tr>
</tbody>
</table>

*Note:* Standard scores on the Test of Nonverbal Intelligence-Fourth Edition (TONI-IV) and Woodcock-Johnson (WJ) Broad Reading and subtests have a normative mean of 100 and an SD of 15. Scores on the Language Identification Battery (Fidler et al., 2011) are weighted, with negative numbers indicating neurotypical (NT) status and positive numbers indicating the presence of a language-learning disability (LLD). Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-IV) scores are scaled with a maximum score of 18. The Spelling Test scores are number correct out of 15. The Modified Token Test scores are number correct out of 44.

### Apparatus and Procedure

**Figure 1** depicts the timeline and order of events for participant testing for Experiment 1 (as well as Experiment 2). Participants completed the experiment over the course of three sessions. In the first session, the assessment battery was performed to determine group membership. In the second session, participants completed either the Experiment 1 or Experiment 2 sound-category learning task (order of task presentation was counterbalanced across participants) which took approximately thirty minutes to complete, followed by an assessment of procedural memory which took approximately forty-five minutes to complete. In the third session, participants completed whichever sound-category learning task they did not perform during the second session, followed by the first portion of a declarative memory
an assessment (which took an estimated five to ten minutes), an assessment of working memory (which took roughly fifteen to twenty minutes), the second portion of the declarative memory assessment (which took approximately five to ten minutes), and a sound discrimination task (which took an estimated five to ten minutes). The sound discrimination task was added to the experiment after participant data collection had already commenced (because effects of sound discrimination on category learning were found in another study with preschoolers; Quam et al., 2021) and was therefore placed at the end of the experiment to eliminate any potential impact the task may have on participant performance in the sound-categorization tasks based on increased stimuli familiarity.
The experimental tasks (sound categorization and sound discrimination) as well as the three memory assessments (procedural, declarative, and working memory) were administered via PsychoPy (Pierce, 2007) on a Mac Mini computer. All audio stimuli were presented to participants through headphones (Sennheiser HD 280 PRO). The audio volume of the computer was initially set to a standard level, but each participant was encouraged to adjust the volume to a level that they found comfortable. All tasks were completed individually in a quiet room.

**Stimuli**

Synthesized isolated vowels were used as auditory stimuli in the sound-category learning and sound discrimination tasks. The stimuli were created using the Klatt speech synthesizer (Klatt and Klatt, 1990) embedded in the Praat phonetic software program (version 5.3.43; Boersma and Weenink, 2008; Weenink, 2009). Two sets of forty-two sounds were created and utilized in the experimental tasks. The stimuli were synthesized to differ across two sound-dimensions, either varying in pitch (F0) and second-formant frequency (F2, or in duration (F0) and first-formant frequency (F1). Half of each participant group (NT and LLD) completed the sound-categorization task with stimuli differing in pitch and F2, while the other half completed the same task with stimuli differing in duration and F1.

By design, one dimension (F2 or F1) is phonetically contrastive in English (a native-language dimension), while the other (F0, perceived as pitch, or duration) is not. This was done to simulate L2 learning, where some L2 dimensions may be present in L1, while others may not be. A female native English speaker (the faculty advisor)’s vocal range was utilized to model the ranges of F0, F1, and F2 as well as the values of other formants. Both sets of stimuli were spaced equally across the Bark Scale (a logarithmic scale created to replicate the human auditory
system’s frequency encoding) on the F0 and F2 or F1 dimensions. The two categories were created to differ equally on both dimensions, with stimuli within each category ranging the entirety of the auditory dimension. For stimuli with differing pitch and F2, this resulted in one category that could be practically described as “high /i/” and another that could be practically described as “low /u/.” For stimuli with differing duration and F1, this resulted in one category that could be roughly described as “short /æ/” and another that could be roughly described as either “/I/” or “long /æ/”.

For stimuli manipulated along pitch and F2, pitch ranged from 8.826 to 15.177 barks. F2 values were created to range between a hyperarticulated /u/ and a hyperarticulated /i/. All sounds had a duration of 0.4 seconds and a uniform maximum amplitude of 70 dB SPL. For stimuli manipulated along duration and F1, duration ranged from 0.4 to 0.946 seconds, while F1 ranged from 5.368 to 7.277 barks. Two aspects of the stimuli were designed to make them sound less artificial—pitch declination was inserted between 0.25 and 0.3 seconds, with the pitch progressively decreasing to 96% of initial pitch height and subsequently remaining at that value for the final 0.1 second, and the ending of each sound featured an amplitude ramp (added utilizing a custom Matlab script created by Sarah Creel; Quam & Creel, 2017a,b). The amplitude ramp resulted in the amplitude declining linearly from 70 dB SPL to zero across ten milliseconds, preventing stimuli from being cut off at a higher amplitude.

**Sound Categorization Task**

Auditory stimuli were presented to participants in a random order within seven training blocks, each consisting of 36 stimuli. Each trial was comprised of the participant listening to the stimuli, pressing one of two response keys labeled with two unfamiliar symbols corresponding to
the two sound categories, and receiving positive or negative feedback depending on the success of their attempt at correctly categorizing the stimuli (see Figure 2).

The boundary between the two categories was set at a linear diagonal, requiring participants to integrate their knowledge of the parameters of both dimensional boundaries, with duration or pitch (F0) as the x-value and first-formant or second-formant frequency (F1 or F2) as the y-value (see Figure 3 for visual of sound-category structures and stimuli used). This multidimensional integration in category learning has been shown to be best learned implicitly (Chandrasekaran et al., 2014). Based on the work of Filoteo et al. (2010), feedback was presented immediately after participant responses and stayed on screen for five-hundred milliseconds. A large smiley face appeared if their response was correct, and a large frowny face appeared if their response was incorrect. The experiment was designed to allow for the measurement of learning outcomes across blocks, as the training trials consisted of judging individual stimuli category membership. This was done to maximize participant training time and increase the chances of participants adapting to utilize a more optimal information-integration strategy.

The first block presented to participants, “block 0”, further promoted the utilization of the optimal strategy, as the stimuli in this block consisted of twelve stimuli that defined the edges of
the category boundaries. The goal in presenting only category-boundary-straddling stimuli to participants in the first training block was to boost their ability to recognize where the diagonal category boundary lay, promoting the development of an intuitive awareness of the outlying bounds of the categories and thus making the task of learning to integrate the respective dimensions easier. This implicit sound categorization task was previously used in the work of Quam et al., 2018, and will not be discussed further here.

FIGURE 3 | Synthesized speech stimuli varied in pitch (F0) and second-formant frequency (F2) (picted), or duration (F0) and first-formant frequency (F1). Block 0 contained 6 randomized stimuli on the categorical boundary (circled), while stimuli selection for blocks 1-6 was entirely random.

Reprinted with permission from Quam et al.

Procedural-Memory Assessment

Materials

Participants’ procedural memory abilities were assessed using a verbal adaptation of the serial-reaction-time task (SRT) (Misyak et al., 2010, a, b). Due to our interest in the impact of procedural memory on language learning, a linguistic form of the SRT was deemed more suitable than the standard visual SRT. In the linguistic version, series of three-non words derived from an artificial language created by Gómez (2002) were presented to participants both visually
and auditorily. A female English speaker’s production of auditory representations of the non-words (Gómez, 2002) were played, while written versions of the non-words were shown within a 2x3 grid on-screen. The non-word series followed a system where the beginning of non-words \((a, b, c)\) determined the end of non-words \((d, e, f)\). This dependent system maintained its non-adjacency through the insertion of a varying item \((X)\), creating the forms \(aXd, bXe,\) and \(cXf\). 24 bisyllabic options existed for the central \((X)\) item (e.g., \(wadim, hiftam, laeljeen\)), while monosyllabic non-words were used at the beginning and end, with a single option mapping onto \(a, b,\) and \(c\) (\(pel, dak,\) and \(vot\)) and \(d, e,\) and \(f\) (\(rud, jic,\) and \(tood\)), respectively. In each trial, the grid shown to participants contained a beginning non-word item \((a, b,\) or \(c)\) in the left column, an intervening non-word item \((X_1 - X_{24})\) in the middle column, and an ending non-word item in the right column \((d, e,\) or \(f)\). The on-screen visual of the grid was presented in tandem with auditory representations of the non-words being shown, prompting participants to match the auditory stimuli with its corresponding non-word by clicking on it. The vertical position in which the target and foil non-word stimuli were presented was pseudorandomized and counterbalanced, meaning that both kinds of stimuli were equally present in upper and lower column positions across trials. Figure 4 depicts an example trial, with one possible grid of non-words. The three target non-words that match the audio presentation have an added curser and underline to depict the process of the participant hearing the non-word and selecting the matching visual stimuli from the options presented in the grid.
The assessment procedure was based on that of Misyak et al. (2010, a, b). In the first phase of the assessment, participants were exposed to 432 novel strings of three non-words across six training blocks, with each training block containing 72 strings. Non-word strings contained one of the 24 X-items in the middle position and one of the three dependency-paired beginning and ending item options in the first and last positions. To ensure that participants would be unable to predict the placement of the target non-words based on stimuli placement within the columns (upper position vs. lower position), the frequency of target and foil non-word stimuli presentation was balanced between them. To be able to intuitively predict which non-word in the third column was the target stimulus, participants needed to pick up on the third non-word’s non-adjacent dependency on the first non-word. Participants were instructed to use their computer mouse to select the written non-word that corresponded to the auditorily presented non-word as quickly and accurately as possible. Each trial started with the 2×3 grid of non-words displayed on-screen and after 250 milliseconds (ms.), the first target non-word was presented to the participant via headphones. As soon as the participant selected the first non-word stimulus.
on-screen, the next auditory non-word stimuli was played. After the third and final selection, the screen went blank for 750 ms, before another grid of non-words was displayed.

In the second phase of the assessment, participants completed a test phase, in which 24 strings of non-words that violated the non-adjacent dependency rule of the previous training blocks were shown, followed by a recovery block of 72 strings that did follow the rule. Participants were not made aware of the shifts between initial training, rule-breaking, and recovery blocks. Finally, participants were given a final prediction task to gauge their non-adjacency pattern learning. In this task, participants were told that the sequencing of the target non-word stimuli was dictated by rules and subsequently completed 24 trials, in which they were provided with auditory representations of the target stimuli for the first two non-words within each stimuli sting and had to identify the third target non-word—between the target and foil non-words—without auditory cuing. Accuracy of performance on the prediction-task was measured based on percentage of trials with correct third non-word responses, making chance performance on the task 50%. Following the precedent of previous work (Misyak et al., 2010a), participant prediction-task accuracy was used to measure individual differences, serving as a predictor of procedural-memory ability for the statistical analysis in the Results.

**Declarative-Memory Assessment**

**Materials**

Participants’ declarative memory skills were evaluated using the logical-memory subtest of the Wechsler Memory Scale-4th edition (WMS-IV; Wechsler, 2009), which was purchased from Pearson-Clinical and modified for administration via computer. Participants were presented with two fake news stories that were three sentences in length, with 25 s of recorded audio
provided by a female native-English speaker for each one. The Weschler Memory Scale-4th edition contains transcriptions of the stories used. A series of “yes”/ “no” questions regarding the content of the stories and a scoring rubric for assessing the accuracy of story details recalled by the participants were utilized.

Procedure

Instructions to pay close attention to the auditory recording of a short news story were provided to participants prior to the recording being played through headphones. Directly after, participants completed an immediate-paragraph recall task, where they were shown a screen prompting them to type the story they had just heard as precisely as possible into a dialogue box. This process was then repeated for the second story. After the task had been performed for both stories, participants completed the working memory assessment (20-30 minutes in length based on individual speed), which functioned as an intervening task between immediate and delayed recall tasks. This insertion of an intervening task of varying length follows existing precedent (Gorwood et al., 2008). Following the working memory task, participants completed the declarative delayed paragraph-recall task, in which they typed their recollection of the stories into dialogue boxes without being provided with any reminders about the content of the stories. After delayed paragraph recall, participants answered a series of “yes”/ “no” questions concerning the stories to assess how accurately they were able to recall them.

The Weschler Memory Scale-4th edition response booklet from Pearson-Clinical provides a rubric to produce net quantitative scores based on the results of all three measures (immediate-paragraph recall task, declarative delayed paragraph-recall task, and “yes”/ “no” questions). However, due to technical difficulties that resulted in the loss of several participants’ typed
responses, we used “yes”/ “no” question response accuracy as the declarative-memory predictor in the statistical analyses reported in the Results. Previous work (Quam et al., 2018) has found a strong correlation between delayed paragraph recall and “yes”/ “no” question accuracy, $r(68)=0.60, p < .001$, suggesting they are both tapping the same delayed recall capacity.

**Working-Memory Assessment**

**Materials**

An auditory version of the reading-span test of working memory developed by Kane et al (2004) was used to as a quantitative measure of participants’ working memory (Daneman and Carpenter, 1980). In this methodology, participants complete two tasks, letter recall and semantic plausibility judgements, simultaneously. Participants were tasked with recalling letters that were presented auditorily in-between every sentence in the sentence-judgement task, in contrast to preceding reading-span tasks that required participants to recall an entire word. These adaptations—utilization of letter recall rather than word recall and usage of an auditory modality to present target information—were implemented to lessen the influence of literacy skills on task outcomes (Kane et al., 2004).

**Procedure**

Prior to the completion of the simultaneous tasks, participants engaged in training trials to familiarize themselves with the individual and combined task procedures. First, participants listened to strings of letters over headphones and were tasked with memorizing the sequence and typing the letters into a dialog box in the order in which they had been presented. The top of the dialog box displayed the possible letter options, including “NA” for forgotten letters. The letter
options were “h,” “j,” “k,” “l,” “m,” “n,” “p,” “q,” “r,” “s,” “t,” “y,” or “NA.” Following this, participants completed training trials in which they judged the semantic plausibility of auditory sentences. Participants were directed to respond “correct” or “incorrect” to each sentence with the highest speed and accuracy possible. For sentences that were semantically plausible (“correct”), participants pressed the up-arrow key, and for sentences that were semantically implausible (“incorrect”), participants pressed the down-arrow key. To motivate participants to keep their sentence-judgement accuracy above 80%, the participant’s current accuracy percentage was shown on the top-left corner of the screen. This was critical to guaranteeing that the letter-recall task succeeded in drawing on participants’ working memory, as the usage of dual-task paradigms in utilizing working memory capacity necessitates that the concurrent tasks receive balanced attention (Daneman and Carpenter, 1980; Kane et al, 2004). Finally, participants completed practice trials that incorporated both tasks simultaneously.

In the main task, participants were exposed to ten sequences, each comprised of three sentences and a varying set size of 2-5 letters, followed by a single dialog box to type answers into. Across the assessment, the order in which letters and sentences were presented was randomized. To confirm that participants were simultaneously attending to both tasks, net sentence accuracy was calculated. A net sentence accuracy score below 70% is typically used as an exclusion criterion, as it may signify that one of the simultaneous tasks is being ignored in favor of the other, meaning that working memory is not being tapped. However, multiple participants in the LLD group had sentence-accuracy scores under 70% and similarly low letter-recall scores, leading us to believe that rather than focusing disproportionately on the letter-recall task at the expense of the sentence-recall task, participants with language impairment struggled with the language processing demand of the sentence-recall task. Therefore, we did not enact this
particular criterion. Letter-recall accuracy was calculated within each trial, meaning that for an individual trial to be correct, all three letters that had been presented needed to be entered accurately. Letter-recall accuracy was used as the working-memory predictor in the statistical analyses reported in the Results.

**Sound Discrimination Task**

Some participants (those tested after the task was added to the procedures) completed a same/different sound-discrimination task based on the one used by Quam et al., 2021. The task aimed to quantify their baseline sensitivity to the sound-dimension cues that were utilized in the sound-category learning tasks—pitch, F2, duration, and F1. This measure was added midway through data collection and was therefore placed at the end of the experimental procedures to prevent any confounding effects on participants’ sound-categorization outcomes. The task consisted of 94 trials: 6 sound-pairs presented 4 times each, for each of the 4 cues. Six possible trial orders existed per cue, with participants receiving different trial orders across cues. In each trial, participants were exposed to 2 sounds, played 1 second apart, and were instructed to listen to both and judge whether they were the same or different. Each trial consisted of a pair of sounds on the cue’s continuum (see **Figure 5**). One sound per set was taken from the end of the continuum, while the sound other was variable, either acoustically different by 1-5 steps, or identical to the preceding sound. There were four trials per sound pair with each possible variation of distance from each other (1-5 in steps, or identical).
D’ scores (see statistical design) that served as an index of participants’ sensitivity to differences between sounds were generated using participants’ trial-by-trial responses as to whether or not they perceived a difference between the sounds. Here, we were interested in the potential effects of group (NT vs. LLD) and cue (pitch vs. duration) on sound-discrimination skills, as well as the variable impact of distance between sounds on the continuum on participants’ sensitivity to acoustic difference.

**Statistical Design**

Planned statistical analyses included independent-sample t tests comparing the NT and LLD groups on the three memory tasks (procedural, declarative, and working memory), the implicit sound-category learning task, and sound-discrimination for each of four dimensions (pitch, F2, duration, and F1), Bonferroni-correcting within each domain and only assuming equal variances when Levene’s test was non-significant. Additionally, a multivariate analysis of covariance (MANCOVA) was run on sound-category-learning accuracy, with predictors being
training block (0-6), group membership (LLD or NT), cues (pitch+F2 or duration+F1), procedural prediction accuracy, working-memory letter recall, and declarative recall.

Model-based analyses were used to further examine participant learning processes and outcomes in the sound-category learning task. Computational models, used in several previous studies (e.g., Quam et al., 2018; Filoteo et al., 2010), provided the best-fitting category-learning strategy that each participant used in each of the training blocks (0-6). Further details regarding the modeling procedures beyond what is covered here are available in numerous previous papers (e.g., Maddox and Ashby 1993; Maddox, 1999; Maddox et al., 2016; Noh et. al., 2016). Each model was fit to each participant’s trial-by-trial responses within each block, and a best-fit model spanning the entire block was chosen using the Bayesian information criterion (BIC; Kass and Wasserman, 1995). BIC is defined as:

\[
BIC_i = 2 \ln L_i + \ln (n) V_i
\]

Wherein \( L_i \) equals model \( i \)'s likelihood, \( V_i \) is the number of free parameters in the model, and \( n \) is the number of trials per block. Note that BIC penalizes models with freer parameters. The best-fitting model meant the model with the smallest BIC value, as smaller BIC values equate to a better fit to the data.

Five types of possible models were included. The first, RR, inferred a random response strategy. The second, UDX, inferred the use of a unidimensional rule using the X-dimension (duration or pitch). The third, UDY, inferred the use of a unidimensional rule using the Y dimension (F1 or F2). The fourth, GLC (“general linear classifier”) or a sub-optimal linear category boundary, signified that the dimensions had been integrated but the placement of the category boundary was inaccurate. For the GLC model to be selected as best-fitting, the slope of
the linear boundary needed to be positive, as a negative slope indicated that the dimensions had been integrated backwards. The fifth, OPT (“optimal linear boundary”), meant that the integrations had been successfully integrated, resulting in a diagonal linear category boundary that had been placed accurately.

Of interest to us was quantifying participants’ use of linear strategies, which indicated successful integration of the two dimensions that the sound stimuli varied across. There were two linear category-learning strategies, the aforementioned GLC and OPT models. As a metric of successful dimensional integration and more optimal sound-categorization performance, we computed the “Number of Linear Blocks” (number of training blocks from blocks 0-6 that were GLC or OPT). An independent sample t-test was run comparing the Number of Linear Blocks across groups (LLD vs. NT).

Also of interest to us was potential differences in unidimensional strategy (UDX and UDY) usage, as the UDX dimensional cues (pitch or duration) are non-native to English, while the UDY dimensional cues (vowel quality—F2 and F1) are native to English. Therefore, disproportionate usage of the UDY strategy (as compared to UDX) was indicative of a native-language bias, resulting in dependency on the native-language dimension and a potential obstacle to successfully integrating the two dimensions. To quantify this effect, we computed participants’ “Native-Language Bias” (number of training blocks from 0-6 that were UDY minus number of training blocks from 0-6 that were UDX), with a positive value indicating the presence of a native-language bias in the participant’s sound-category learning task performance. An independent samples t-test was run comparing native-language-bias scores between groups, followed by a paired-samples t-test as a within-subjects comparison between participants’ native
language bias in the implicit sound-category learning task (Experiment 1) and the explicit sound-category learning task (Experiment 2).

For the sound-discrimination task data, the average accuracy of participants’ responses (either “same” or “different”) for each sound-dimension were translated into D’ scores to create a scaled D’ sensitivity index. The index was calculated as \( z(H) - z(F) \), where \( H \) (hits) represented the rate at which the participant correctly identified sounds being different, and \( F \) (false alarms) represented the rate at which participants inaccurately identified sounds as being different in trials where the sounds were identical. D’ scores were used in independent-sample \( t \)-tests comparing the groups across the four sound-dimension cues.

Results

Group Differences in Memory Skills

We first compared the two groups on their memory skills based on measures from the 3 memory assessments (procedural-prediction accuracy for procedural memory, letter-recall for working memory, and declarative recall for declarative memory) using unpaired \( t \)-tests, Bonferroni-corrected for 3 comparisons, with an adjusted \( p \)-value threshold of \( p = .0167 \).

Procedural and working memory skills differed significantly between groups. For procedural prediction-task accuracy, the NT group (\( M = 0.60, SD = 0.22 \)) scored significantly higher than the LLD group (\( M = 0.50, SD = 0.82 \)), \( t(39.6) = 2.89 \), two-sided \( p = .006 \). For working memory letter-recall, the NT group (\( M = 0.91, SD = 0.07 \)) also scored significantly higher than the LLD group (\( M = 0.65, SD = 0.17 \)), \( t(7.6) = 4.37 \), two-sided \( p = .003 \). There were no significant
differences between LLD and NT participants’ declarative memory.

**Group Differences in Sound-Category Learning**

Next, we compared the two groups on implicit sound-category learning task performance (averaged across blocks 0-6). Task accuracy differed between groups, with NT participants ($M = 0.68, SD = 0.07$) scoring significantly higher than their LLD counterparts ($M = 0.62, SD = 0.06$), $t(50) = 2.374$, two-sided $p = .021$. Figure 6 depicts block-by-block average sound-category learning task accuracy for both groups.

![Figure 6](image)

**FIGURE 6** Block-by-block sound-categorization attempt average accuracy (with standard error) of NT and LLD participant groups in the implicit condition.
Group Differences in Sound-Discrimination

We then compared NT and LLD groups’ baseline sound-discrimination abilities along each of the sound-dimensions that were used to manipulate sound-category stimuli between Experiment 1 (implicit sound-category learning task) and Experiment 2 (explicit sound-category learning task)—duration, first-formant frequency (F1), pitch, and second-formant frequency (F2). To do this, we used unpaired \( t \)-tests, Bonferroni corrected for 4 comparisons with an adjusted \( p \)-value threshold of \( p = .0125 \). Duration discrimination ability differed significantly between groups, with the NT group (\( M = 3.73, SD = 1.03 \)) scoring significantly higher than the LLD group (\( M = 2.18, SD = 1.46 \)), \( t(30) = 2.889, p = .007 \). LLD and NT participants did not differ significantly in pitch-discrimination ability, F2-discrimination, or F1-discrimination ability.

Impacts of Memory Skills on Implicit Sound-Category-Learning Accuracy

We ran a MANCOVA on implicit sound-category-learning accuracy to identify which factors impacted participants’ accuracy on the sound-category learning task, with predictors being training block (0-6), group membership (LLD or NT), cues (pitch+F2 or duration+F1), procedural prediction accuracy, working-memory letter recall, and declarative recall. A main effect of Declarative Recall was found, \( F(1,31) = 10.60, p = .003 \). There were no significant effects of Procedural Prediction Accuracy or Working-Memory Letter Recall. There were no effects of—or interactions with—group.

Usage of Integrative Strategies in Implicit Sound-Category Learning

On the implicit sound-category learning task, use of linear strategies indicated integration
of the two sound-dimensions that stimuli were manipulated across—making these strategies optimal for the experiment’s category structure, and a useful metric of sound-categorization performance. We computed participants’ “Number of linear blocks” (total number of blocks that were GLC or OPT—see Statistical Design for explanation of models) and ran an independent samples t-test to compare number of linear blocks between NT ($M = 2.40, SD = 1.61$) and LLD ($M = 1.77, SD = 0.67$) participant groups. Levene’s test was violated, $F(8.393), p = .006$, so equal variances were not assumed. There was a numerical trend towards lower LLD linear strategy usage that did not meet the threshold for statistical significance, $t(30.620) = 1.868$, two-sided $p = .071$. **Table 2** shows NT and LLD participants’ learning-strategy usage (best fit models) and accuracy of implemented strategies for the final training block (6). Note that while only block 6 is depicted in the table, the statistical analyses of sound-category learning task accuracy, learning strategy usage, and impact of memory skills on sound-category task performance included all blocks (0-7).
The implicit sound-category task stimuli were designed to differ across a native-dimension (F1 or F2) and a non-native-dimension (duration or pitch). In comparing participants’ usage of unidimensional strategies (focusing on one dimension for sound-categorization decision making, rather than integrating both) we are able to assess their native language (L1) bias based on how heavily they relied on the native dimensional strategy (UDX) over the non-native dimensional strategy (UDY). See Table 2 for breakdown of UDX and UDY learning-strategy usage and strategy accuracy between groups. Participants’ Native-Language-Bias scores were computed as number of blocks that were UDY minus number of blocks that were UDX, with a positive result representing a participant’s L1 bias. An independent samples t-test was run comparing Native-Language-Bias scores between groups. No significant differences were found.

<table>
<thead>
<tr>
<th>Model and measure</th>
<th>Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NT</td>
</tr>
<tr>
<td><strong>Optimal Integrative</strong></td>
<td>(Opt. CJ Model in Explicit Condition and Opt. Information-Integration Model in Implicit Condition)</td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.28</td>
<td>.11</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.81</td>
<td>.78</td>
</tr>
<tr>
<td><strong>Suboptimal Integrative</strong></td>
<td>(Subopt. CJ Model in Explicit Condition and Subopt. Information-Integration Model in Implicit Condition)</td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.12</td>
<td>.11</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.79</td>
<td>.67</td>
</tr>
<tr>
<td><strong>Unidimensional X</strong></td>
<td>Proportion of participants</td>
<td>.23</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.69</td>
<td>.63</td>
</tr>
<tr>
<td><strong>Unidimensional Y</strong></td>
<td>Proportion of participants</td>
<td>.35</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.74</td>
<td>.75</td>
</tr>
<tr>
<td><strong>Random Responding</strong></td>
<td>Proportion of participants</td>
<td>.02</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.42</td>
<td>.46</td>
</tr>
</tbody>
</table>

**Group Differences in Native-Language-Bias**
between the Native-Language Bias scores of NT ($M = 0.07, SD = 3.46$) and LLD ($M = 0.55, SD = 3.61$) participants on the implicit sound-category learning task $t(50) = -0.380$, two-sided $p = .705$, using the Bonferroni-corrected threshold of $p = .025$ (corrected for 2 comparisons).

**Discussion**

In line with the expectations of the procedural deficit hypothesis (PDH), there were significant differences in procedural memory and implicit sound-category-learning accuracy between NT and LLD groups, with NT participants scoring higher in both cases. Moreover, there were no group differences in declarative memory, supporting the PDH’s postulation that declarative memory is intact in LLD. NT participants also scored higher in duration-discrimination ability, which concurs with previous findings of higher overall discrimination in preschoolers with typical language development than preschoolers with developmental language disorder and NT preschoolers (Quam et al., 2020). We predicted that NT participants would achieve better dimensional integration on the implicit sound-category learning task than their LLD counterparts, resulting in group differences in linear (GLC and OPT) strategy usage, and while there was a numerical trend in the direction that we expected, it did not meet the threshold for statistical significance. Unexpectedly, NT participants also scored higher in working memory skills than the LLD group, going against the PDH’s prediction of a selective impairment to LLD procedural memory. Also unexpected was the main finding of Declarative Recall being significantly associated with accuracy on the implicit sound-category task, where we would have expected to see a main effect of Procedural Prediction Accuracy. We anticipated significant group differences in Native-Language Bias, predicting that the NT group would achieve better
dimensional-integration and thus, be more likely to let go of the unidimensional Y (native-language dimensional cue) strategy. However, there were no significant differences in Native-Language Bias in learning strategy usage between groups.

EXPERIMENT 2

Methods

Participants

The same set of participants from Experiment 1 took part in Experiment 2.

Sound-Category-Learning Task

As in Experiment 1, Experiment 2 utilized synthesized vowels as auditory stimuli. However, each participant was exposed to a different set of cues than in Experiment 1—if a participant completed the Experiment 1 task with stimuli varying along the pitch (F0) and second-formant frequency (F2) dimensions, they competed the Experiment 2 Task with stimuli varying along the duration (F0) and first-formant frequency (F1) dimensions, and vice versa. Cues were counter-balanced across Experiments (1 and 2) and groups (NT vs. LLD).

Participants completed six training blocks, each consisting of 42 sound-stimuli that were presented in a random order. The procedure of each individual trial was the same as that of Experiment 1. The category boundary between the two stimuli groups was roughly placed between stimuli in the left upper quadrant and the rest of the stimuli (see figure 7). The stimuli category in the upper left quadrant was either shorter in duration (F0) and higher in first-formant
frequency (F1), or lower in pitch (F0) and higher in second-formant frequency (F2). The category occupying the lower two quadrants and right upper quadrant spanned the entire range of F0 (duration or pitch), encompassing stimuli that was short in duration or low in pitch and low in F1 or F2, long in duration or high in pitch and low in F1 or F2, and long in duration or high in pitch and high in F1 or F2. This sound-category design, based on the one used in Filoteo et al., 2010, was designed to be learned best explicitly, as the rule that governed the sound-category boundary was verbalizable, meaning that a consciously mediated strategy could accurately categorize stimuli (e.g., “if it’s short in duration and sounds more like /æ/ than /ɪ/, then it goes in category A rather than category B”). Reaching optimal accuracy in categorizing the stimuli required participants to apply a conjunctive verbal rule, accounting for both manipulated dimensions, rather than a single unidimensional rule (focusing on either duration or pitch (X) or F1 and F2 (Y). According to the COVIS (Competition between Verbal and Implicit Systems) model, learners’ initial efforts in category learning favor testing verbal unidimensional rules (Bruner, Goodnow, & Austin, 1956; Shepard, Hovland, & Jenkins, 1961), with the complexity and weight of the rules being adjusted based on how effective they are at guiding accurate categorization, which is evaluated via explicit processing (Chandrasekaran et al., 2014). Therefore, optimal learning of the explicit sound-categories would entail moving away from using unidimensional rules, and instead refining a conjunctive rule to accurately place the category boundary.
Statistical Design

The statistical design that was implemented for Experiment 2 was identical to that of Experiment 1, barring one exception. The best-fit sound-category learning strategy models for Experiment 2 differed from those of Experiment 1, as the designs of the category structures were different. Three of the same models were used—random response (RR), unidimensional Y (UDY), and unidimensional X (UDX). However, the linear diagonal strategies (GLC and OPT) that indicated successful-informational integration in the implicit condition were not relevant in the explicit condition. Instead, there were different ideal strategy models in the explicit condition. The fourth model was a suboptimal conjunctive top-left (CJ TL) model, representing a rule-based strategy that accounted for both manipulated dimensions, but was not optimal in its placement of the categorical boundary. The fifth model was an optimal conjunctive top-left model (labeled OPT, but not to be confused with the optimal linear diagonal strategy of the implicit condition), representing an integrative rule-based strategy with an optimal category boundary.
Akin to the “Number of Linear Blocks” metric in Experiment 1, we planned to compute “Number of Conjunctive Blocks” (number of training blocks from 1-6 that were CJ TL or OPT) as a metric of successfully incorporating both dimensions into a rule-based strategy, and consequentially achieving more optimal sound-categorization outcomes. However, participants minimally adopted these integrative strategies in the explicit condition (see Table 3 in Results and Discussion), creating strong floor effects that made this metric unusable.

Results

Group Differences in Sound-Category Learning Task

**Figure 8** depicts block-by-block average sound-category learning task accuracy for both groups. As in Experiment 1, we compared the two groups on explicit sound-category learning task performance (averaged across blocks 0-6). Sound-category learning task performance (accuracy averaged across blocks 1-6) did not differ significantly between NT ($M = 0.63, SD = 0.08$) and LLD ($M = 0.61, SD = 0.08$) groups, $t(51)= 1.225$, two-sided $p = .226$. 
We ran a MANCOVA on explicit sound-category-learning accuracy to identify which factors impacted participants’ accuracy on the sound-category learning task, with predictors being training block (0-6), group membership (LLD or NT), cues (pitch+F2 or duration+F1), procedural prediction accuracy, working-memory letter recall, and declarative recall. A significant main effect of Working-Memory Letter Recall was found, $F(1,31) = 7.85, p = .009$. The effect of Procedural Prediction Accuracy did not meet the threshold for statistical significance, $F(1,31) = 3.79, p = .061$. There were no significant effects of Declarative Recall, group membership, sound-discrimination of cues, or block.
As was the case in Experiment 1’s implicit sound-category learning task, the stimuli used for Experiment 2’s explicit sound-category learning task was designed to differ across a native-dimension (F1 or F2—UDY) and a non-native-dimension (duration or pitch—UDX). Comparison of participants’ usage of unidimensional strategies (using one dimension for sound-categorization decision making, rather than both) allowed us to assess participants’ Native-Language Bias (computed for each participant as number of UDY blocks minus number of UDX blocks). See Table 3 for breakdown of UDX and UDY learning-strategy usage and strategy accuracy between groups. Unpaired t-tests, Bonferroni-corrected for 2 comparisons, with an adjusted p-value threshold of $p = .025$, found no significant differences were found between the

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Proportion of Participants by Group Whose Data Were Best Fit by Each Model and Corresponding Accuracy Levels (Percentage Correct) for Block 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model and measure</td>
<td>Group</td>
</tr>
<tr>
<td></td>
<td>NT</td>
</tr>
<tr>
<td><strong>Optimal Integrative</strong>&lt;br&gt; <em>(Opt. CJ Model in Explicit Condition and Opt. Information-Integration Model in Implicit Condition)</em></td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>—</td>
</tr>
<tr>
<td>Accuracy</td>
<td>—</td>
</tr>
<tr>
<td><strong>Suboptimal Integrative</strong>&lt;br&gt; <em>(Subopt. CJ Model in Explicit Condition and Subopt. Information-Integration Model in Implicit Condition)</em></td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.05</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.76</td>
</tr>
<tr>
<td><strong>Unidimensional X</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.21</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.60</td>
</tr>
<tr>
<td><strong>Unidimensional Y</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.70</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.71</td>
</tr>
<tr>
<td><strong>Random Responding</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion of participants</td>
<td>.05</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.46</td>
</tr>
</tbody>
</table>

**Group & Task Differences Native-Language Bias**
Native-Language Bias scores of NT ($M = 1.54$, $SD = 3.74$) and LLD ($M = 2$, $SD = 4.14$) participants on the explicit sound-category learning task, $t(51) = -0.348$, two-sided $p = .730$.

Given a lack of group differences in both Experiments 1 and 2, groups (LLD vs. NT) were collapsed for a paired-samples $t$-test comparing participants’ Native-Language Bias in Experiment 1 (implicit sound-category learning task) vs. Experiment 2 (explicit sound-category learning task). Participants showed significantly more Native-Language Bias in the explicit sound-category learning task (Experiment 2; $M = 1.62$, $SD = 3.79$) than in the implicit sound-category learning task (Experiment 1; $M = 0.15$, $SD = 3.46$), $t(51) = 2.378$, two-sided $p = .021$.

**Discussion**

As predicted by the procedural deficit hypothesis (PDH), performance on the explicit sound-category learning task did not differ between groups. Unexpectedly, Working-Memory Letter Recall was a significant predictor of accuracy on the explicit sound-category learning task, where we would have expected to see a main effect of Declarative Recall, which in turn had no significant effect. The lack of group differences in Native-Language Bias in sound-category learning strategy selection align with our prediction that NT and LLD participants would show equal native language bias on the explicit sound-category learning task, as the PDH posits that declarative memory and explicit learning are intact in LLD. The significant difference found between native language bias on the implicit sound-category learning task (Experiment 1) and the explicit sound-category learning task (Experiment 2)—where groups were collapsed—indicates that learners, regardless of LLD or NT status, are better able to overcome Native-Language Bias in implicit learning conditions, rather than explicit ones. This is consistent with
our predictions for the NT group, but it was held for both groups.

**General Discussion**

The goals of this study were to examine group differences in speech-sound learning and how those differences connect to differences in participants’ memory and sound discrimination abilities. We also aimed to examine factors that contributed to overall success in an artificial adult second-language learning scenario, with an interest in the role of memory systems and effects of native language (L1). Over two experiments, adult participants with and without LLD learned two artificial sound categories that differed in structure, based on the learning system each was designed to tap, and completed assessments of baseline sound-discrimination ability (for F1, F2, duration, and pitch) and memory skills (procedural, declarative, and working memory).

Our experimental design tested the Procedural Deficit Hypothesis (PDH) (Ullman & Pierpont, 2005; Lum et al., 2012), which predicted that participants with LLD would show selective impairments to procedural memory and implicit learning, but not to declarative memory and explicit learning. This study aimed to see if the framework of the COVIS (Competition between Verbal and Implicit Systems) model of category-learning (Ashby et al., 1998; 2011) could be used to test the PDH. COVIS theorizes that the verbal (“reflective” or explicit) and implicit (“reflexive”) systems are in competition to mediate control of category learning, and that in adults there is an initial bias towards the explicit system (Chandrasekaran et al., 2014). Although the two theories were developed within two distinctly separate literatures, both assert distinctions in the neurological bases of implicit and explicit learning, with
fundamental overlap between the proposed functions of each learning system (Chandrasekaran et al., 2014; Lum et al., 2012). However, it is important to note that there are some theoretical and terminological distinctions between the PDH and COVIS model that must be considered when discussing our findings.

Based on the COVIS model, we anticipated that the structure of the implicit sound-category learning task (Experiment 1) would tap unconscious learning, mediated by the procedural memory system and that of the explicit sound-category learning task (Experiment 2) would tap conscious, verbalizable, hypothesis-testing driven learning (Chandrasekaran et al., 2014). We predicted that for NT participants, there would be less interference of the native language dimension in the implicit sound-category learning condition than in the explicit sound-category learning condition, due in part because of the COVIS model’s prediction of initial bias towards unidimensional strategies in the explicit condition (Filoteo et al., 2010), and because the implicit task’s linear structure is difficult to apply verbalizable rules to (Chandrasekaran et al., 2014), making it more difficult for participants to consciously “grab onto” the native dimension as a strategy for categorizing stimuli. We predicted that participants with LLD would not see the same gains in shifting away from native-language bias in the implicit condition that the NT group participants did, as the PDH predicted that participants with LLD would have a selective impairment to implicit learning that they may try to compensate for using explicit strategies (Lum et al., 2012), which are ill-suited to the structure of the implicit task. Based on the Procedural Deficit Hypothesis (Ullman & Pierpont, 2005; Lum et al., 2012), we anticipated that participants with LLD would show selective impairments to procedural memory and implicit learning, but not to declarative memory and explicit learning.

In line with our predictions, NT participants achieved better dimensional integration
(more usage of integrative models and less L1 bias in categorization attempts) in the implicit learning condition (Experiment 1) than the explicit learning condition (Experiment 2).

Furthermore, comparison between native-language bias on the implicit and explicit sound-category learning tasks (with groups collapsed due to no group level differences), found significantly more influence of native language bias in the explicit condition than in the implicit condition. Notably, usage of strategies that considered both dimensions in categorization decision-making—integrative strategies—occurred almost entirely within the implicit condition. This may be explained by multiple factors, owing to both differences in the explicit and implicit learning systems and in facets of the sound-category designs. The implicit task’s non-verbalizable structure may have decreased participants’ conscious awareness of the native-language dimension, making them more adaptable their sound-categorization strategy.

The lack of adoption of conjunctive strategies in the explicit condition—with the suboptimal conjunctive strategy being used 5% of the time by NT participants and 0% of the time in block 6, and the optimal conjunctive strategy not receiving any use by either group in block 6—may also be attributed to the influence of native language bias. The strength of participants’ gravitation toward the L1 contrastive dimension (F1 or F2) over the non-L1 contrastive dimension (duration or pitch) may have hindered participants from overcoming their initial bias towards unidimensional rule testing that is predicted in the explicit condition (Filoteo et al., 2010), making them less likely to make the leap to conjunctive rule testing. Additionally, participants may not have been properly motivated to shift strategies within the task’s limited timeframe and trials, given that the unidimensional strategies (X and Y) both yielded a maximum accuracy of 83% on the explicit sound-category learning task, well above the aforementioned averages that participants were achieving by training block 6. It is possible that if the task had a
longer duration and included more trials—particularly if those trials were split between different session dates—that participants would have reached the ceiling of unidimensional strategy accuracy, and consequently shifted to a conjunctive strategy to continue to see improvement.

Group outcomes on the implicit (Experiment 1) and explicit (Experiment 2) sound-category-learning tasks, as well as the assessments of procedural and declarative memory, align with the PDH, as NT participants scored significantly higher in implicit sound-category-learning accuracy and procedural memory than LLD participants, but the groups did not differ in explicit sound-category-learning accuracy and declarative memory. Additional significant group differences were in duration discrimination and working-memory ability, with NT participants scoring higher than LLD participants on both measures. This difference in group-level working-memory ability supports the idea that LLD may be caused by an impairment in explicit learning due to impairments to multiple underlying memory systems.

The level of separation that exists between memory systems, their effects on one another, and their relative contributions to implicit and explicit learning are debated within the current category-learning literature. Some studies have found that multiple memory skills provide positive contributions to category learning, with Craig & Lewandowsky (2012) finding that category-learning speed was predicted by working-memory skill (regardless of implicit or explicit learning strategy), while Lewandowsky et al. (2012) found that working-memory skills were a predictor of category learning ability for both implicit and explicit based learning strategies. The study that preceded the current study (Quam et al., 2018), tested the implicit sound-category-learning ability of neurotypical adults (the current study’s Experiment 1 sound-category structure is the same as the one utilized in Experiment 2 of Quam et al.) and found that working memory was a significant predictor of implicit sound-category-learning accuracy. fMRI
imaging has shown that a significant amount of the brain regions employed for explicit tasks also aid in implicit tasks, and vice versa, such as the hippocampus and medial temporal lobes (Carpenter et al., 2016). These mixed results have drawn more attention to the role of system-level interactions, with more recent studies within the COVIS framework positing that increased flexibility to switch between category-learning strategies may lead to improved learning outcomes (Ashby & Maddox, 2011).

Another point of interest in evaluating the contributions of memory systems to sound-category learning outcomes was the associations between participants’ memory skills and their accuracy on the sound-category learning tasks. Surprisingly, Declarative Recall was a predictor of accuracy on the implicit sound-category task, instead of Procedural Prediction Accuracy and Working-Memory Letter Recall was a significant predictor of accuracy on the explicit sound-category learning task, instead of expected effects of Declarative Recall.

Theoretical differences between the PDH and COVIS may partially help account for these discrepancies. The PDH focuses more narrowly on the neurological bases of declarative memory—medial temporal lobe structures (Ullman & Pierpont, 2005)—and of procedural memory—portions of frontal/basal ganglia-circuits (Ullman & Pierpont, 2005)—as the neurological bases of explicit and implicit learning. The COVIS model considers explicit learning to be subserved by frontal brain structures that underlie both declarative and working memory, as the hypothesis-testing that mediates explicit learning is heavily reliant on executive function (Chiu & Yantis, 2009), in line with some evidence that both memory functions are mediated by overlapping structures (Buckner et al., 1999; Botvinik et al., 2001; Simon & Spiers, 2003). Further work is required to better understand the joint and individual functions of these memory systems in facilitating both implicit and explicit sound-category learning. Interestingly,
the COVIS model and PDH also disagree on the role of working-memory in regard to implicit learning. The PDH contends that while it is not a primary impairment, individuals with LLD may have impairments to working-memory as well as procedural memory, as working-memory and procedural memory may be subserved by anatomically proximate and parallel frontal/basal ganglia circuits (Ullman & Pierpont, 2005). Conversely, COVIS contends that the executive attention and working-memory are not relied upon for information-integration (implicit) tasks, as they are subserved by the posterior caudate, putamen, and supplementary area (Chandrasekaran et al., 2014).

In discussing the current study’s results, the small LLD sample size—and consequentially limited statistical power with which to analyze group level differences—must be acknowledged. 10 participants with LLD completed Experiment 2 and 9 of those participants completed Experiment 1. It is possible that we did not find significant effects in areas where we expected to because of this. One factor limiting the comparability between sound-category designs is that the maximum accuracy of both unidimensional strategies (X and Y) was lower on the implicit sound-category, at 79%. This difference in the potential effectiveness of unidimensional strategies between learning conditions may have contributed to the differences in participants’ adoption of integrative (conjunctive or linear) strategies between the explicit and implicit sound-category learning tasks. This disparity could be controlled in future replication studies. However, the 4% difference between 79% and 83% is small, making it unlikely to be a full explanation for the striking differences in dimensional integration in the two cases. Replicating this study with a larger sample size and modifications to ensure equivalence between the implicit and explicit sound-category structures will help validate our current findings. Additionally, testing the practical applications of tapping implicit learning to improve L2 learning outcomes for both NT
adults and adults with LLD is merited, given that participants in both groups showed less language bias in the implicit condition than the explicit condition.

In conclusion, our findings support the procedural deficit hypothesis predictions that procedural memory and implicit learning are impaired in adults with LLD, while declarative memory and explicit learning are not. More optimal dimensional integration in the implicit sound-category learning task compared to the explicit sound-category learning task supports the COVIS model prediction that learning requiring the integration of multiple dimensions is best-suited to an implicit learning condition. However, the presence of an additional impairment to working memory in LLD participants, as well as unexpected connections found between memory systems and implicit and explicit learning, as well as a lack of group differences in native language bias, indicate that the interactions between memory systems, and their subsequent impact on category learning outcomes, do not straightforwardly match the predictions of the PDH. Promisingly, these results suggest that some of the difficulty that adults face in second-language learning, owing to difficulty comprehending and learning non-native language dimensions, can be ameliorated by strategically tapping implicit learning, regardless of LLD.
Citations


