

Running head: Longitudinal study of childhood listening difficulties

1 **Childhood listening and associated cognitive difficulties persist into adolescence**

Katsuaki Kojima^{1,2,3}, Li Lin³, Lauren Petley^{3,4}, Nathan Clevenger³, Audrey Perdew³, Mark Bodik^{3,5},
Chelsea M. Blankenship^{3,6}, Lina Motlagh Zadeh³, Lisa L. Hunter^{3,6,7,8}, and David R. Moore^{3,8,9}

1. Perinatal Institute, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039
2. Department of Pediatrics, University of Cincinnati College of Medicine, 3230 Eden Avenue, Cincinnati, OH 45267
3. Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, 240 Albert Sabin Way, Cincinnati, OH 45229
4. Department of Psychology, Clarkson University, 8 Clarkson Ave, Potsdam, NY 13699
5. Cornell University, Ithaca, NY 14850
6. Research in Patient Services, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039
7. College of Allied Health, Communication Sciences and Disorders, University of Cincinnati, 3225 Eden Ave, Cincinnati, OH 45267
- 2 8. Departments of Otolaryngology and Neuroscience, University of Cincinnati College of
3 Medicine, 260 Stetson St., Cincinnati, OH 45267
- 4 9. Manchester Centre for Audiology and Deafness, University of Manchester, Oxford Rd,
5 Manchester M13 9PL, UK
6

Keywords: hearing loss, speech-in-noise, listening difficulty, cognition, pediatric, longitudinal, auditory processing disorder, NIH Cognition Toolbox

Corresponding author: Katsuaki.kojima@cchmc.org

7
8 Data sharing statement: The data that support the findings of this study are available from the
9 corresponding author upon reasonable request.

10
11
12

1 **ACKNOWLEDGEMENTS**

2 This research was supported by grant 5R01DC014078 from the National Institute of Deafness and other
3 Communication Disorders (DRM), 2UL1TR001425-05A1 by the National Center for Advancing
4 Translational Sciences of the National Institutes of Health (KK), and by the Cincinnati Children's
5 Research Foundation. DRM received support from NIHR Manchester Biomedical Research Centre. We
6 would like to acknowledge that an earlier version of this manuscript was published as a preprint on
7 medRxiv (<https://doi.org/10.1101/2022.08.11.22278673>).

8
9 K.K. analyzed and interpreted data and cowrote the article. L.L. analyzed and interpreted data and
10 was involved in critical revisions. L.P. was involved in critical revisions. N.C. and A.P. collected data.
11 M.B. analyzed and interpreted data and was involved in critical revisions. C.M.B., L.M.Z, and L.L.H.
12 collected data and were involved in critical revisions. D.R.M. designed the study, analyzed and
13 interpreted data, and cowrote the article.

14

15 **Abstract (457 words /500 words)**

16 **Objective:** Listening difficulty (LiD) refers to the challenges individuals face when trying to hear and
17 comprehend speech and other sounds. LiD can arise from various sources, such as hearing sensitivity,
18 language comprehension, cognitive function, or auditory processing. Although some children with LiD
19 have hearing loss, many have clinically normal audiometric thresholds. To determine the impact of
20 hearing and cognitive factors on LiD in children with a clinically normal audiogram, we conducted a
21 longitudinal study. The Evaluation of Children's Listening & Processing Skills (ECLiPS), a validated and
22 standardized caregiver evaluation tool, was used to group participants as either LiD or typically
23 developing (TD). Our previous study aimed to characterize LiD in 6- to 13-year-old children during the
24 project's baseline, cross-sectional phase. We found that children with LiD needed a higher signal-to-noise
25 ratio during speech-in-speech tests and scored lower on all assessed components of the NIH Cognition
26 Toolbox than TD children. The primary goal of this study was to examine if the differences between LiD
27 and TD groups are temporary or enduring throughout childhood.

28

29 **Design:** This longitudinal study had three data collection waves for children with LiD and TD aged 6-13
30 years at Wave 1, followed by assessments at 2-year (Wave 2) and 4-year (Wave 3) intervals. Primary
31 analysis focused on data from Waves 1 and 2. Secondary analysis encompassed all three waves despite
32 high attrition at Wave 3. Caregivers completed the ECLiPS, while participants completed the Listening in

1 Spatialized Noise Sentences test (LiSN-S) and the NIH Toolbox Cognition Battery during each wave. The
2 analysis consisted of 1) examining longitudinal differences between TD and LiD groups in demographics,
3 listening, auditory, and cognitive function; 2) identifying functional domains contributing to LiD; and 3)
4 test-retest reliability of measures across waves. Mixed-effect models were employed to analyze
5 longitudinal data.

6
7 **Results:** The study enrolled 169 participants, with 147, 100, and 31 children completing the required
8 testing during Waves 1, 2, and 3, respectively. The mean ages at these waves were 9.5, 12.0, and 14.0
9 years. On average, children with LiD consistently underperformed TD children in auditory and cognitive
10 tasks across all waves. Maternal education, auditory and, especially, cognitive abilities independently
11 predicted caregiver-reported listening skills. Significant correlations between Waves 1 and 2 confirmed
12 high, long-term reliability. Secondary analysis of Wave 3 was consistent with the primary analyses of
13 Waves 1 and 2, reinforcing the enduring nature of listening difficulties.

14
15 **Conclusion:** Children with LiD and clinically normal audiograms experience persistent auditory,
16 listening, and cognitive challenges through at least adolescence. The degree of LiD can be independently
17 predicted by maternal education, cognitive processing, and spatial listening skills. This study underscores
18 the importance of early detection and intervention for childhood LiD and, for the first time, highlights the
19 role of socioeconomic factors as contributors to these challenges.

20
21
22
23

1 INTRODUCTION

2 Listening is commonly defined as an active form of hearing (Sweetow and Henderson-Sabes 2004). It
3 facilitates an accurate understanding of speech, a key component of everyday communication. Successful
4 communication allows healthy language development, as seen in improved language outcomes in children
5 with hearing loss who received early interventions (Pimperton and Kennedy 2012; Yoshinaga-Itano et al.
6 1998; Moeller 2000; Dettman et al. 2007). Problems with listening may similarly predispose children to
7 challenges in language acquisition. It is difficult, however, to determine the direction of the causal
8 pathway because children with primary language or other cognitive problems may have difficulties
9 comprehending speech that prevents them from active hearing (listening). Listening difficulty (LiD)
10 describes difficulty hearing and understanding speech and other suprathreshold sounds (Petley et al. 2021;
11 Dillon and Cameron 2021; Sharma et al. 2014). LiD is an umbrella term that describes symptoms; the
12 underlying causes of LiD may be problems in multiple domains, including hearing, language, cognition,
13 and auditory processing (Dillon and Cameron 2021; Sharma et al. 2009; Tomlin et al. 2015). Although
14 some children presenting with LiD have hearing impairments, as indicated by elevated audiometric
15 thresholds, other children with LiD have clinically normal audiometric sensitivity (Hind et al. 2011). To
16 clarify the contribution of hearing and cognitive factors to LiD for individuals with clinically normal
17 audiograms, we prospectively studied children with LiD, identified using a validated and standardized
18 caregiver questionnaire, Evaluation of Children's Listening & Processing Skills (ECLiPS) (Barry and
19 Moore 2021).

20 In the Sensitive Indicators of Childhood Listening Difficulties (SICLiD) study, we characterized LiD
21 in a cohort of 6- to 13-year-old children during a baseline, cross-sectional phase of the project, using
22 caregiver questionnaires and a battery of audiological, auditory processing, and cognitive tests (Petley et
23 al. 2021; Hunter et al. 2021). We found no group differences in audiological measures (pure-tone
24 audiometry, tympanometry, and middle ear muscle reflexes) between LiD and typically developing (TD)
25 children (Hunter et al. 2021). However, children with LiD required a higher signal-to-noise ratio during
26 speech hearing in noise testing (Listening in Spatialized Noise-Sentence, LiSN-S (Cameron and Dillon
27 2007)), had abnormal auditory processing measures tested using SCAN-3:C (Keith 2009), and lower
28 cognitive scores in all tested components of the NIH Cognition Toolbox (Weintraub et al. 2013b) relative
29 to their TD peers (Petley et al. 2021).

30 Whether LiD persists from childhood into adulthood or resolves over time remains unclear.
31 Answering this question is essential for informing clinical management and understanding the
32 mechanisms and consequences of LiD. If LiD is a persistent problem, more comprehensive assessment
33 and aggressive treatment may be warranted compared to a transient problem of childhood. Some evidence
34 suggests LiD may be persistent - young adults with a history of LiD in childhood reported ongoing

1 communication challenges nearly 10 years after initial diagnosis (Del Zoppo et al. 2015). However,
2 quantitative longitudinal data tracking changes in listening, auditory, and cognitive functioning in
3 children diagnosed with LiD earlier in life is lacking. As part of the SICLiD study, we performed
4 behavioral testing in the same cohort of children every two years to assess longitudinal changes.

5 Socioeconomic status has been associated with childhood cognitive performance (Christensen et al.
6 2014), language development (Pungello et al. 2009), mental health problems (Reiss 2013), and health
7 disparities (Fiscella and Williams 2004). However, the impact of social factors on LiD remains unclear.
8 Given the association between LiD and cognitive performance, we hypothesized that the degree of LiD
9 may also be associated with social factors. In a prior cross-sectional study, we developed a model
10 showing that SCAN-3:C, LiSN-S, and NIH Cognition Toolbox scores partially accounted for caregiver
11 reports of LiD (Petley et al. 2021). While social factors were evaluated, they were not incorporated into
12 the final model. Longitudinal data allowed us to assess social and other demographic factors across time
13 and with increased statistical power.

14 Evaluating longitudinal cognitive performance during development can be challenging because of the
15 different rates of development and potential practice effects (Sullivan et al. 2017). In the SICLiD study,
16 we used the NIH Cognition Toolbox to assess cognition. Weintraub et al. (2013b) showed that the
17 Toolbox test results had high reliability in pediatric samples with short testing intervals between 7 and 21
18 days, with intraclass correlations (ICCs) between 0.76 and 0.99. However, recent literature on the
19 reliability for longer intervals (> 1 year) has been mixed, with ICCs between 0.24 and 0.85 (Taylor et al.
20 2022) and 0.43 and 0.82 (Anokhin et al. 2022). The SICLiD study enabled evaluation of the reliability of
21 the NIH Cognition Toolbox over the years.

22 We had three aims in the current study. The primary aim was to evaluate longitudinal changes in
23 listening, auditory, and cognitive functions of children with LiD relative to control children. We
24 hypothesized that LiD would be long-lasting because LiD was seen across ages (between 6 and 13 years
25 old) in our Wave 1, cross-sectional data (Petley et al., 2021), and because of the general persistence
26 through adolescence and into adulthood of developmental problems associated with LiD, such as attention
27 (Hechtman et al. 2016; Roy et al. 2017), reading (Shaywitz et al. 1999; Lohvansuu et al. 2021), and
28 language (St Clair et al. 2011). Second, building on the previous model (Petley et al., 2021), we aimed to
29 develop an enhanced model of contributors to LiD using longitudinal data. Finally, we examined long-
30 term test-retest reliability of the NIH Cognition Toolbox, ECLiPS, and LiSN-S by using prospectively
31 collected data from children with and without LiD.

32

33 **MATERIALS AND METHODS**

1 The Institutional Review Board of Cincinnati Children’s Hospital (CCH) Research Foundation
2 approved this study.

3 **Participants**

4 LiD and TD children were 6 to 13 years old at enrollment. The LiD group was recruited from
5 children referred to CCH for audiological evaluation and from the community. All the participants in the
6 TD group were recruited from the community. Recruitment from the community involved flyers posted in
7 relevant CCH clinics (Audiology, Pediatrics, Speech-Language Pathology) and emails sent to all CCH
8 employees and families interested in research. Interested caregivers completed (i) a background
9 questionnaire and (ii) the ECLiPS (Barry and Moore 2021) to confirm eligibility for the LiD or TD group.
10 Total standardized ECLiPS scores of ≥ 7 defined the TD group, and scores < 7 defined the LiD group.
11 However, four children with LiD had a previous audiologic diagnosis of auditory processing disorder
12 (APD) but scored 7 - 9 on the ECLiPS at enrollment. APD is a hypothesized source of LiD (Dillon and
13 Cameron 2021). The inclusion criterion for both groups was English as the child’s native language.
14 Exclusion criteria for both groups were reported neurologic (e.g., epilepsy or stroke), psychiatric (e.g.,
15 attention deficits hyperactivity disorder (ADHD) requiring medical treatment), or intellectual conditions
16 that would prevent or restrict their ability to complete testing procedures. The TD group additionally
17 could not have any reported listening difficulty, or cognitive or learning disorder (e.g., developmental
18 delay, attention, language, reading, or high-functioning autistic diagnoses). Eligibility was determined
19 based on caregiver responses to the background questionnaire. Groups were age-matched by proportional
20 sampling (Table 1).

21 One hundred sixty-six participants (74 with LiD, 92 TD) were enrolled; 19 withdrew or otherwise
22 exited the study (6 with LiD, 13 TD). Reasons for withdrawal were schedule conflicts, transportation
23 availability, and becoming ineligible after enrollment. The remaining 147 participants (68 with LiD, 79
24 TD) completed the assessments at baseline (Wave 1). After the initial study visit, we contacted the
25 participants after ~2 years (Wave 2) and 4 years (Wave 3) to collect longitudinal data. A total of 100 (52
26 with LiD, 48 TD) and 31 (16 with LiD, 15 TD) children participated in Wave 2 and 3, respectively (Table
27 1). The sample size for Wave 3 was notably smaller than other waves, partly because of the COVID-19
28 pandemic. All participants had clinically normal hearing bilaterally, as detailed below.

29 **Procedure**

30 All caregivers of eligible and interested participants reviewed the informed consent form with a
31 study staff member and discussed the purpose, procedures, risks and benefits, duration, and expectations
32 involved in the study. According to institutional policy, children aged 11 and above also assented using a
33 child-friendly version of the consent document. All participants received financial compensation for their
34 participation. Each participant completed audiometric and behavioral testing during the study visit, as

1 listed below. Several additional tests, including brainstem and cortical evoked responses and structural
2 and functional MRI, were also performed during the visit as part of the SICLiD study. The results of these
3 additional tests are beyond the scope and not included in the current manuscript. Some results of these
4 additional tests have been reported elsewhere (Hunter et al. 2021; Stewart et al. 2022; Hunter et al. 2023).

6 **Background Questionnaire**

7 Study staff developed a background questionnaire to collect information about the participants,
8 including demographics, medical and surgical history, and socioeconomic status. This questionnaire lacks
9 reliability or validity data because of its basic nature. Specific information collected by this questionnaire
10 was age, sex, race, ethnicity, maternal education (a proxy for socioeconomic status (Sherar et al. 2016)),
11 primary language used at home, and diagnosis and treatment received for otologic, neurologic, behavioral,
12 developmental, and psychiatric problems. The questionnaire included specific developmental conditions
13 as examples, such as ADHD, autism spectrum disorder, and dyslexia. Caregivers completed the same
14 background questionnaire at each Wave. Demographic variables at each time point are summarized in
15 Table 1.

17 **Caregiver Evaluation of Children's Listening**

18 The ECLiPS (Barry and Moore 2021) was administered at each wave to assess participants'
19 listening and communication abilities. This questionnaire consists of 38 straightforward statements (items)
20 describing behaviors commonly observed in children. Caregivers rated their child on a five-point scale,
21 ranging from strongly disagree to strongly agree, based on their agreement with each statement. These
22 ratings were averaged to derive age-scaled scores for five subscales: speech and auditory processing
23 (SAP), environmental and auditory sensitivity (EAS), language/literacy/laterality (L/L/L), memory and
24 attention (M&A), and pragmatic and social skills (PSS), each containing 6 to 9 distinct items. The
25 subscales were further grouped into Language, Listening, Social, and Total aggregate (composite) scores,
26 all of which were standardized with a population mean of 10 (SD = 3) based on British data (Barry et al.
27 2015). All ECLiPS subscales exhibit high test-retest reliability, with intraclass correlations (ICCs) above
28 0.8 (Barry and Moore 2021). Construct validity has been demonstrated through convergence with other
29 established tests measuring similar skills, including the Children's Auditory Processing Performance
30 Scale (CHAPS; (Smoski et al. 1998)), the CCC-2 (Bishop 2006), and the Social Communication
31 Questionnaire (Rutter et al. 2003). Although the ECLiPS provides scores for five latent factors related to
32 listening (SAP, EAS, L/L/L, M&A, and PSS), these scores are not validated for diagnosing specific
33 impairments in each area. Consequently, eligibility for inclusion in the LiD category was determined
34 based on the ECLiPS Total composite score (below 7, which defines a cutoff of < 1 SD from the

1 population mean). The Total composite score represents the overall degree of LiD (Barry and Moore
2 2021) and is appropriate for this purpose.

3 4 **Audiometry**

5 Otoscopy was completed, and if necessary, cerumen was removed before audiometry. All
6 audiometric tests were completed in a double-walled soundproof booth. Standard (0.25 - 8 kHz) and
7 extended high-frequency (10-16 kHz) thresholds were measured using the manual Hughson-Westlake
8 method for the range of 0.25 to 8 kHz at octave intervals and at four additional frequencies (10, 12.5, 14,
9 and 16 kHz) using an Equinox audiometer (Interacoustics Inc., Middelfart, Denmark) with Sennheiser
10 300 HDA headphones. Participants with elevated thresholds (>20 dB HL) in either ear and at any
11 standard clinical frequency were excluded and referred for further clinical care as appropriate.

12 13 **The Listening in Spatialized Noise-Sentences (LiSN-S) test**

14 The Listening in Spatialized Noise – Sentences (LiSN-S) test (Cameron and Dillon 2007; Brown
15 et al. 2010) measures the ability to listen and repeat simple, spoken sentences in the presence of
16 distracting sentences, developed using the same child-friendly criteria as the BKB sentences (Bench et al.
17 1979). We selected LiSN-S as the auditory assessment tool for the SICLiD study because of its
18 multifaceted approach to evaluating spatial and speech-hearing abilities, as well as its specific design for
19 use in pediatric populations. The LiSN-S (US Edition (Brown et al. 2010)) was administered at each wave
20 using a commercial CD played on a laptop, a task-specific soundcard, and Sennheiser HD 215
21 headphones. Participants were asked to repeat a series of target sentences, presented using virtual space
22 generic head-related transfer functions (Humanski and Butler 1988), directly in front (0°: diotic) while
23 ignoring two distracting speakers.

24 The LiSN-S has four listening conditions in which the distractors change voice (different or same
25 as target) and/or position (0° and 90°) (Humanski and Butler 1988). The LiSN-S is adaptive; the level of
26 the target speaker decreases or increases in SNR relative to the distracting speech if the listener responds
27 correctly or incorrectly during up to 30 sentences in each condition. The speech reception threshold (SRT)
28 represents 50% correct SNR for the condition. The Low Cue condition is where distractors are in the
29 same voice as the target and in front (0°) relative to the listener. Distractors are in different voices and
30 ±90° relative to the listener in the High Cue condition. Derived scores of the LiSN-S are the Talker
31 Advantage and Spatial Advantage, so-called because each is the difference between SRTs from two
32 conditions. This subtraction process should, to varying extent, separate auditory from cognitive influences
33 (Cameron and Dillon 2007; Petley et al. 2021). Spatial Advantage thus represents the participant's ability
34 to separate the spatial source of the speech using purely auditory cues, while Talker Advantage

1 demonstrates the participant's ability to distinguish individual talkers, which involves both acoustic and
2 linguistic information (Perrachione et al. 2011; Quinto et al. 2020). Test-retest comparisons for the four
3 listening conditions of the LiSN-S showed significant improvement from the first test to the second but
4 no differences in the advantage scores (Cameron et al. 2011).

5 A spatial "Pattern Score" is a quantitative clinical measure of the benefit of adding virtual spatial
6 cues to the information in the LiSN-S (i.e., target and distracting stimuli presented diotically) (Cameron
7 and Dillon 2011). The pattern score is calculated as follows.

$$Pattern\ Score = \{Spatial\ Advantage + (Total\ Advantage - Talker\ Advantage)\} / 2$$

8 The Pattern Score has been used as both a diagnostic score and a treatment outcome for a specific
9 auditory disorder termed Spatial Processing Disorder (SPD) (Cameron et al. 2011, 2012; Cameron et al.
10 2014). The age-specific cutoff value of the pattern score is determined as follows.

$$Cutoff = 7.1208 + (0.0561 * Age) + (-1.96 * 1.51)$$

11
12 **Cognition (National Institutes of Health, NIH Toolbox)**

13 Participants' cognitive abilities were evaluated using the NIH Toolbox for the Assessment of
14 Neurological and Behavioral Function, Cognition Domain (Weintraub et al. 2013a; Weintraub et al.
15 2013b). Participants completed testing either online or via an iPad app in a private sound-attenuated booth
16 or a quiet room at each wave. The precise composition of the testing battery depended on participant age.
17 Up to eight standardized cognitive instruments measuring different aspects of fluid or crystallized
18 reasoning were utilized.

19 All participants in this study completed the Picture Vocabulary Test (PVT), Flanker Inhibitory
20 Control and Attention Test (Flanker), Dimensional Change Card Sort test (DCCS), and Picture Sequence
21 Memory test (PSMT). Each test produced a US age-corrected standardized score. The scores from all four
22 tests were combined to calculate a single Early Childhood Composite (ECC), which provides a measure
23 of overall cognitive skill for children aged six years and older. The PVT is an adaptive test that evaluates
24 the participants' vocabulary skills. Participants listen to an audio recording of a word and select one of
25 four pictures that closely matches the word's meaning. The Flanker assesses the participant's ability to
26 inhibit responses and focus their attention. Participants report the direction of a central visual stimulus in
27 a series of five similar, flanking stimuli, where the flanking stimuli may either be congruent or
28 incongruent. The DCCS tests cognitive flexibility (attention switching). Target and test card stimuli vary
29 along two dimensions, shape, and color. Participants are asked to match test cards to the target card
30 according to a specified dimension that varies for each trial. The PVT and DCCS scores both measure
31 accuracy and reaction time. The PSMT assesses episodic memory by presenting illustrated objects and

1 activities along with audio-recorded descriptive phrases. Participants are scored on the number of adjacent
2 picture pairs correctly remembered over two learning trials. The length of the picture sequences varied
3 from 6 to 18 pictures based on the participant's age. The ECC score derived from these tests allows for a
4 comprehensive evaluation of overall cognitive performance in early childhood.

5 Children at least 8 years of age qualify for additional subtests from the Toolbox, contributing to a
6 Total Composite (TC) measure of general cognitive skill. These additional tests were administered to all
7 children 8 years of age and older (i.e., all children for Waves 2 and 3). Fluid Composite measures include
8 the list sorting working memory (LSWM), the pattern comparison processing speed (PCPS) tests, and the
9 DCCS, Flanker, and PSMT. The LSWM subtest evaluates working memory by requiring participants to
10 arrange visually and auditorily presented objects (food and animals) in order of size. In the PCPS subtest,
11 participants were instructed to indicate as quickly as possible whether two visually presented cards were
12 the same or different. Participants completed the Reading Recognition (RR) test in addition to the PVT
13 for the Crystallized Composite measure. The RR requires participants to read words and letters aloud.

14

15 **Analysis**

16 The analysis of the study was divided into three distinct parts, each corresponding to a specific
17 aim of the research. In the first part, we evaluated the longitudinal changes in listening, auditory, and
18 cognitive functions of children with LiD in comparison to TD children. In the second part, we developed
19 a model to explain childhood listening difficulty using longitudinal data. Finally, the third part aimed to
20 establish the long-term test-retest reliabilities of the NIH Cognition Toolbox, ECLiPS, and LiSN-S, using
21 prospectively collected data.

22 In Part 1 of the analysis, we investigated differences between two groups, TD and LiD, on their
23 demographics, auditory, and cognitive function. Demographic information included age, sex, race,
24 ethnicity, and maternal education. Test scores of ECLiPS, LiSN-S, and NIH Cognition Toolbox were
25 compared. We used mixed-effect models (Verbeke 1997; Fitzmaurice et al. 2012) to analyze the
26 longitudinal data. These models can handle repeated measures and are flexible on missing data. We
27 assumed the data were missing at random (MAR) and used Restricted Maximum Likelihood (REML)
28 estimation. We tested the MAR assumption with Little's test (Little 1988). A mixed-effect model has
29 fixed effects and random effects. Fixed effects are variables that have a consistent effect on the outcome
30 variable, while random effects are variables that have a varying effect on the outcome variable depending
31 on the level of another variable. Mixed-effect models are useful for analyzing data that have repeated
32 measurements or hierarchical structures. Mixed-effect models can use both dependent and independent
33 variables of multiple time points by modeling the change in the outcome variable over time as a function

1 of fixed and random effects. This allows accounting for within-subject correlation and between-subject
2 variations in the data.

3 For each demographic characteristic, we created a mixed-effect model using group (TD vs. LiD),
4 wave (Wave 1, 2, and 3), and interaction (Group x Wave) as fixed effects and participant as a random
5 effect. Maternal education was collapsed into four groups: up to high school education (High school),
6 some college education (College), Bachelor's degree (Bachelor), and post-graduate education (Post-grad).
7 Race was collapsed into two groups: white and non-white. Ethnicity was collapsed into two groups:
8 Hispanic or Latino and others. Demographic characteristics that are statistically different between groups
9 ($p < .05$) may confound the statistical analysis that compared behavioral testing results between TD and
10 LiD. Therefore, such demographic variables were included in the analyses to control potential
11 confounding.

12 Next, we evaluated longitudinal changes in listening and cognitive skills of TD and LiD children
13 using the ECLiPS sub-scores and Total Scaled score, LiSN-S Advantage, Cue and Pattern scores, and
14 NIH Cognition Toolbox sub-tests and composite scores. We constructed a mixed-effect model for each
15 test, incorporating fixed effects for group (TD vs. LiD), wave (Wave 1 vs. 2), interaction (Group x Wave),
16 and maternal education (a potential confounding variable), and participant as a random effect (Part 1a
17 model). All analyses were conducted using standardized scores adjusted for age on the test date. We
18 applied Benjamini and Hochberg (B-H) adjustment (Benjamini and Hochberg 1995) to the p-values for
19 multiple comparisons, accounting for eight ECLiPS sub-scores, four LiSN-S scores, seven NIH Cognition
20 Toolbox sub-scores, and four NIH Cognition Toolbox composite scores with a false discovery rate (FDR)
21 of less than 5%. Wave 3 data were not included in these mixed-effect models due to the smaller sample
22 size compared to other waves. To examine further longitudinal changes in Wave 3, we created separate
23 mixed-effect models using data only from participants who completed all three waves (Part 1b model).

24 In Part 2 of the analysis, our objective was to identify the functional domains that contributed to
25 LiD. We first tested whether a model developed in our prior cross-sectional study (Petley et al. 2021)
26 could predict the ECLiPS Total Scaled score at Wave 2. The original model included the SCAN-3:C
27 Composite Score, the LiSN-S Talker Advantage score, and the NIH PVT and DCCS scores as predictors
28 of ECLiPS Total Scaled scores at Wave 1. Since the SCAN-3:C was not administered in Wave 2, this
29 model included only the LiSN-S Talker Advantage score and NIH PVT and DCCS scores from Wave 2
30 as predictors. Next, we developed a new mixed-effect model using the longitudinal data to predict
31 ECLiPS Total Scaled scores. This longitudinal model included both predictor variables and outcome
32 variables from Waves 1 and 2 in a single mixed-effect model. Candidate predictors were maternal
33 education, five LiSN-S scores (Cue, Advantage, and Pattern scores), and measures of cognitive function
34 (either the seven NIH Cognition Toolbox subtests or the four NIH Toolbox composite scores), and the

1 outcome variable was the ECLiPS Total Scaled score in both Waves 1 and 2. By modeling the trajectories
2 of ECLiPS scores over time as a function of the predictors, this approach accounts for both within-subject
3 correlations across waves and between-subject variations. The random effects of the model capture
4 individual differences in trajectories, while the fixed effects estimate overall trends. Separate models were
5 tested using the NIH Cognition Toolbox subtests or composite scores as candidate predictors to avoid
6 including duplicate variables in the model. Owing to its smaller sample size, Wave 3 was omitted from
7 this longitudinal model.

8 All candidate predictors were initially examined against the ECLiPS Total Scaled score through
9 univariate analyses using a mixed-effect model, with participants used as a random effect and predictor
10 variables and the interaction term (predictor x wave) as fixed effects. Only variables with $p < 0.1$ in the
11 univariate analysis were included in the backward selection. Pairwise comparisons among the levels of
12 significant variables were adjusted for using Tukey-Kramer multiple adjustment. Beta weights for each
13 predictor variable were calculated to indicate the strength of their effects on the ECLiPS Total Scaled
14 score.

15 The aim of Part 3 of the analysis was to evaluate the test-retest reliability of the measures. To
16 achieve this, Intra-class correlation coefficients (ICC) and Pearson correlation coefficients were computed
17 to investigate the consistency of the test outcomes between Waves 1 and 2 for all participants. Wave 3
18 data were excluded from the analysis due to the limited sample size. Furthermore, we calculated test-
19 retest differences as another relevant metric. For each measure, we subtracted the Wave 2 score from the
20 Wave 1 score for each participant. We then calculated the mean and standard deviation of these difference
21 scores. To determine whether there were any variations in the correlation coefficients between the TD and
22 LiD groups, the coefficients were calculated separately for each group, and z-tests were performed. To
23 avoid type I errors, the p-values were adjusted for multiple comparisons using the B-H adjustment method
24 (Benjamini and Hochberg 1995) with an FDR of less than 5%. B-H adjustment was separately applied to
25 four LiSN-S scores, seven NIH Cognition Toolbox sub-scores, four NIH Cognition Toolbox composite
26 scores, and eight ECLiPS sub-scores. All analyses used SAS version 9.4 (SAS Institute, Cary, N.C.),
27 setting the two-sided significance level at 0.05.

28

29 **RESULTS**

30 **LiD group associated with lower maternal education**

31 Table 1 shows the number and demographics of participants in each group and wave. As Part 1 of our
32 analysis, we compared the TD and LiD groups on their demographics, auditory, and cognitive function.
33 Among the demographic variables, only maternal education differed significantly between the two groups,
34 with a significant Group x Wave interaction observed (see Table 1, Supp. Table 1). Specifically, the TD

1 group consistently had higher maternal education than the LiD group across all waves, and this difference
2 increased from Wave 1 to Wave 3. To address the potential confounding effect of maternal education, we
3 included it as a covariate in the subsequent analyses.

5 **Group differences in caregiver-reported listening skills were maintained at Wave 2**

6 By design, the LiD group had lower ECLiPS scores than the TD group at Wave 1 (Figure 1A and Table 2)
7 (Petley et al. 2021). This difference between the groups persisted at Wave 2 (Analysis Part 1a model).
8 There was no significant change in ECLiPS scores across the two waves for either group. Furthermore,
9 there was no significant Group x Wave interaction effect, indicating that the listening skills of both
10 groups remained stable over time (Supp. Tables 2, 3).

12 **LiD group had consistently lower cognitive scores**

13 The LiD group scored significantly lower than the TD group on all seven sub-tests and four composite
14 scores of the NIH Cognition Toolbox across Waves 1 and 2 (Figure 1B, Table 3, Supp. Tables 4, 5).
15 Scores did not change significantly over time for either group as indicated by non-significant Wave or
16 Group x Wave interaction effects.

18 **LiD group had consistently lower skills in segregating sentences**

19 The LiD group had lower LiSN-S z-scores than the TD group across waves (Figure 1C and Table 2, Supp.
20 Table 5, 6). Spatial Advantage scores increased significantly from Wave 1 to Wave 2 for both groups.
21 Other LiSN-S z-scores did not show any significant change over time. Spatial Pattern Scores were similar
22 between groups and waves. Notably, fewer children met the diagnostic criteria for SPD based on Pattern
23 Scores at Wave 2 than at Wave 1: only three children (3 LiD, 0 TD) compared to thirteen children at
24 Wave 1 (7 LiD, 6 TD).

26 **Differences between the TD and LiD groups persisted through adolescence**

27 We examined longitudinal changes in listening skills, speech-in-speech skills, and cognitive function
28 across three waves based on 31 participants who completed all three waves (Figure 2). The TD group had
29 higher ECLiPS sub-scores and Total Scaled scores than the LiD group (Analysis Part 1b model, Supp.
30 Table 7), as in the Part 1a analysis of Waves 1 and 2 data. However, both groups had improved listening
31 skills over time relative to the age-standardized scores. ECLiPS speech and auditory processing subscale
32 score and composite Listening score were significantly higher in Wave 2 than Wave 1 across both groups.
33 Moreover, ECLiPS Total Scaled scores in Wave 3 were significantly higher than Wave 1. In contrast to
34 the listening skills, there was no significant change in cognitive function or speech-in-speech skills over

1 time for either group. The TD group had significantly higher NIH Cognition Toolbox scores and two
2 LiSN-S z scores (Low Cue and Talker Advantage) than the LiD group. Wave 1 participants who did not
3 participate in Wave 2 dropped out randomly (Little's $p = 0.74$). However, those who did not participate in
4 Wave 3 were non-random (Little's $p = 0.01$), suggesting systematic reasons for attrition. Maternal
5 education was not included in the model due to the small sample size.

7 **Cognitive, sensory, and socioeconomic factors independently contributed to parent-reported** 8 **listening skills**

9 For Part 2 of the analysis, we compared two models to predict ECLiPS Total Scaled scores: one based on
10 Wave 2 data only (Petley et al. 2021), and another based on longitudinal data from both Waves 1 and 2.
11 The original model explained 42% of the variance of ECLiPS Total Scaled scores at Wave 2. Our new
12 model explained 54.8% of the variance ($\chi^2(2) = 48.21, p < 0.0001$). The new model included three
13 significant predictors: maternal education ($\beta = .70, p = .0038$), NIH total composite score ($\beta = .49, p$
14 $< .0001$), and LiSN-S Spatial Advantage ($\beta = .16, p = .0023$). These results indicate that children's
15 listening skills during Waves 1 and 2 were independently influenced by their cognitive function, speech-
16 in-speech skills, and maternal education level. Children whose mothers had a Bachelor's degree scored
17 higher on the ECLiPS than those whose mothers had a High school diploma by 3.1 points.

19 **ECLiPS, LiSN-S, and NIH Toolbox scores demonstrated reliability between Waves 1 and 2**

20 As Part 3 of the analysis, we evaluated the test-retest reliability of the measures (Figure 3, Table 4, Supp.
21 Table 8). The ECLiPS Total Scaled scores showed good reliability across Waves 1 and 2 with no
22 differences between TD and LiD groups. The mean test-retest difference on the ECLiPS was 0.10, and the
23 SD was 2.33. The NIH Toolbox sub-tests and composite scores also had significant and consistent
24 correlations across waves (ICCs for each sub-test ranging from 0.48 to 0.82) and groups (p values from z -
25 tests ranging from 0.07 to 0.92; Supp. Table 8). The mean test-retest differences ranged from -5.59 to
26 4.58, with the SDs ranging from 9.66 to 27.02 compared to the normalized mean score of 100. LiSN-S
27 scores revealed a moderate but significant agreement between waves, with ICCs ranging from 0.28 to
28 0.42. The mean test-retest differences in standard deviation units were between 0.18 and 0.41, with SD
29 between 1.26 and 2.13.

31 **DISCUSSION**

32 **Impact of study**

33 We found that ECLiPS, LiSN-S, and NIH Toolbox scores were stable over time. Children with
34 LiD, identified by the ECLiPS caregiver questionnaire, consistently performed more poorly than TD

1 children on auditory and cognitive tasks across multiple years. Maternal education was a significant
2 predictor of caregiver-reported listening skills, independent of auditory and cognitive skills.
3 Socioeconomic factors may, therefore, play a crucial role in this complex condition. Aligning with our
4 prior cross-sectional study, cognitive skills played a relatively larger role in the parent report of listening
5 difficulty than auditory skills. The enduring nature of LiD has clinical implications for early detection
6 and intervention for childhood LiD, which may have long-term effects on affected individuals.

8 **Socioeconomic factors in listening difficulties**

9 This study is the first to show how socioeconomic factors affect reported listening skills in
10 children with LiD after partially controlling for auditory and cognitive skills measured by the LiSN-S test
11 and NIH Toolbox scores. Previous studies have linked maternal education to various aspects of child
12 development, such as academic outcomes (Harding et al. 2015), language skills (Magnuson et al. 2009),
13 attention problems (Hjern et al. 2010), and auditory brainstem responses (Skoe et al. 2013). It is therefore
14 unsurprising that maternal education was associated with the degree of LiD in this study, given the known
15 relation between LiD and these other developmental factors (Moore et al. 2018; Hunter et al. 2023).
16 However, the finding is noteworthy because we controlled for cognitive and sensory functions using the
17 LiSN-S test and NIH Toolbox scores, which suggests that maternal education may have an independent
18 effect on LiD beyond these factors. Alternatively, maternal education may influence LiD through other
19 factors. Future studies could further explore and parse the influence of socioeconomic factors. In the
20 meantime, maternal education and socioeconomic factors may inform the risk of LiD and guide clinical
21 management, including the frequency of follow-up and intervention. Clinicians could also provide
22 support and guidance for parents with lower socioeconomic status who may face more challenges in
23 accessing resources or implementing home-based activities for their children. They could also refer the
24 children to other services, such as social workers or community programs, that can help them overcome
25 disadvantages.

27 **Persistent listening challenges**

28 This study provides evidence that children with LiD overall experience persistent listening
29 challenges across time, as indicated by significantly lower mean ECLiPS scores compared to TD children
30 across three waves of testing. Our analysis also accounts for the stability of scores for each participant
31 over time, as we used mixed-effect models to evaluate the longitudinal score changes of each participant.
32 The findings suggest that differences between TD and LiD groups are not due to a supra-normal
33 performance of TD children, as their ECLiPS scores were consistent with other independent samples of
34 children without LiD (Barry et al. 2015). The results are consistent with previous research (Del Zoppo et

1 al. 2015) and underscore the need for early diagnosis and active management to minimize the long-term
2 effects of LiD. Our findings also raise questions about the effectiveness of current management strategies
3 for LiD, which may not adequately address the unique needs of these children. In reviews of medical
4 records, we found that more than half the children with LiD had received specialist assessment and
5 interventions before and/or during the study period. To better evaluate the efficacy of these interventions,
6 future research should use quantitative outcome measures specific to each treatment. This will help
7 identify the most effective strategies for addressing the persistent listening challenges experienced by
8 children with LiD.

9

10 **Cognitive challenges**

11 Cognitive challenges of LiD participants were persistent throughout the study period, as shown
12 by our prior Wave 1 study (Petley et al. 2021; Stewart et al. 2022) and longitudinal data from this study.
13 LiD participants performed more poorly than controls on all the NIH Toolbox subscale and composite
14 scores. Modeling revealed that the NIH Total Composite score was the most predictive cognitive measure
15 for reported LiD, surpassing any combination of NIH Toolbox subscales. Moreover, the NIH Total
16 Composite score had a threefold higher beta weight than the next most predictive outcome, speech-in-
17 speech repetition (LiSN-S Spatial Advantage). This finding underscores the substantial role of cognitive
18 impairment in parent-reported LiD. It is worth noting that the LiD group in our study may have included a
19 variety of underlying etiologies, as we used the ECLiPS Total Scaled score as an inclusion criterion.
20 Consequently, our results are relevant to children with LiD as a whole, rather than to a specific underlying
21 condition of LiD. To address distinct needs, individual children with specific deficits should receive
22 evidence-based, targeted treatment (Dillon and Cameron 2021; Sharma et al. 2012; Musiek and Chermak
23 2007).

24

25 **Speech-in-speech listening**

26 One purpose for using LiSN-S was to characterize the specific pattern of impairment in LiD
27 participants using individual and derived scores of LiSN-S that represent different aspects of listening
28 abilities. In this study (Part 1a), the TD group outperformed the LiD group on all age-adjusted LiSN-S
29 scores. Our results thus highlight the significant challenges that LiD children face across multiple
30 domains of listening ability. Interestingly, in the Part 1b analysis, children with LiD scored lower only in
31 the Low Cue and Talker Advantage, similar to our previous findings in Wave 1 (Petley et al., 2021).
32 Taken together, the data suggest that Low Cue and Talker Advantage may be more sensitive measures for
33 differentiating TD and LiD than other LiSN-S scores for most children in this cohort.

1 Although the number of cases was small, the prevalence of SPD decreased at Wave 2, despite
2 ongoing listening, auditory, and cognitive challenges in the LiD group and without any participants
3 receiving specific treatment for SPD (Cameron et al. 2012). Furthermore, neither the Pattern score nor the
4 frequency of SPD differed significantly between the TD and LiD groups, suggesting that, at least in this
5 cohort, SPD does not represent the listening problems identified by the ECLiPS (Petley et al. 2021).
6 Future research clarifying the clinical significance and natural history of SPD would be beneficial.
7 Among LiSN-S scores, only the Spatial Advantage age-adjusted scores increased significantly from
8 Waves 1 to 2 in both groups. This could be due to developmental changes or inconsistent normalization
9 and is unlikely to be training effects after a 2-year separation between tests.

11 **Reliability**

12 We believe our findings are based on reliable longitudinal data, as the current study utilized a
13 large sample of children and had a 68% follow-up rate at Wave 2, much higher than a prior longitudinal
14 study of childhood LiD with only a 13% follow-up response rate (Del Zoppo et al. 2015). Additionally,
15 loss to follow-up at Wave 2 was random, suggesting a low likelihood of sampling bias. Findings from
16 Waves 1 and 2 thus likely represent the typical progression of LiD. Unfortunately, due to COVID, the
17 follow-up rate of Wave 3 dropped significantly, and sampling bias was introduced. However, analysis of
18 Wave 3 participants showed overall longer-term results that were consistent with the primary analysis of
19 Waves 1 and 2.

20 ECLiPS, LiSN-S, and NIH Toolbox scores all showed statistically significant correlations
21 between Waves 1 and 2. These correlations were similar for the TD and LiD groups. The high reliability
22 of the Toolbox scores observed in this study, even with long testing intervals of two years, seems to
23 alleviate recent concerns regarding these scores (Taylor et al. 2022). This high reliability may be
24 attributable to consistent testing procedures carried out by research staff at a single site, as opposed to
25 testing across multiple sites (Taylor et al. 2022). Although LiSN-S age-adjusted scores were significantly
26 correlated between Waves 1 and 2, the agreement between waves ranged from poor to good, and
27 correlation coefficients were generally lower than those reported in a previous study (Cameron et al.
28 2011). The higher retest reliability for that previous study may be explained by shorter testing intervals
29 (2-4 months) and older (above 12 years old) participants (Cameron et al. 2011).

31 **Limitation**

32 Due to the small sample size of Wave 3, we ran a separate analysis focusing on longitudinal
33 changes of Wave 3 participants to reduce sampling bias. Though this method reduced bias, the analysis
34 remained underpowered. Despite this limitation, our overall research findings are significant because we

1 showed long-lasting challenges of childhood LiD that may influence clinical management strategy, and
2 we identified a social factor, maternal education, that independently contributes to reported LiD.

3

4 **Conclusion**

5 Children with LiD and clinically normal audiograms have persistent auditory, listening, and
6 cognitive challenges through at least early adolescence. The degree of LiD can be independently
7 predicted by maternal education, cognitive processing, and spatial listening skills with a relatively larger
8 influence of cognitive processing over other factors.

9

10

11

1 REFERENCES

- 2 Anokhin, Andrey P., Monica Luciana, Marie Banich, Deanna Barch, James M. Bjork, Marybel R.
3 Gonzalez, Raul Gonzalez, Frank Haist, Joanna Jacobus, Krista Lisdahl, Erin McGlade,
4 Bruce McCandliss, Bonnie Nagel, Sara Jo Nixon, Susan Tapert, James T. Kennedy, and
5 Wesley Thompson. 2022. 'Age-related changes and longitudinal stability of individual
6 differences in ABCD Neurocognition measures', *Developmental Cognitive Neuroscience*,
7 54.
- 8 Barry, J. G., and D. R. Moore. 2021. 'ECLIPS: Evaluation of children's listening and processing
9 skills (2nd ed.)', *Cincinnati Children's Hospital*.
- 10 Barry, J. G., D. Tomlin, D. R. Moore, and H. Dillon. 2015. 'Use of Questionnaire-Based
11 Measures in the Assessment of Listening Difficulties in School-Aged Children', *Ear*
12 *Hear*, 36: e300-13.
- 13 Bench, J., A. Kowal, and J. Bamford. 1979. 'The BKB (Bamford-Kowal-Bench) sentence lists
14 for partially-hearing children', *Br J Audiol*, 13: 108-12.
- 15 Benjamini, Yoav, and Yosef Hochberg. 1995. 'Controlling the false discovery rate: a practical and
16 powerful approach to multiple testing', *Journal of the Royal statistical society: series B*
17 *(Methodological)*, 57: 289-300.
- 18 Bishop, Dorothy VM. 2006. *CCC-2: Children's Communication Checklist-2* (Pearson).
- 19 Brown, D. K., S. Cameron, J. S. Martin, C. Watson, and H. Dillon. 2010. 'The North American
20 Listening in Spatialized Noise-Sentences test (NA LiSN-S): normative data and test-retest
21 reliability studies for adolescents and young adults', *J Am Acad Audiol*, 21: 629-41.
- 22 Cameron, S., and H. Dillon. 2007. 'Development of the Listening in Spatialized Noise-Sentences
23 Test (LiSN-S)', *Ear Hear*, 28: 196-211.
- 24 Cameron, S., and H. Dillon. 2011. 'Development and evaluation of the LiSN & learn auditory
25 training software for deficit-specific remediation of binaural processing deficits in
26 children: preliminary findings', *J Am Acad Audiol*, 22: 678-96.
- 27 Cameron, S., H. Dillon, H. Glyde, S. Kanthan, and A. Kania. 2014. 'Prevalence and remediation
28 of spatial processing disorder (SPD) in Indigenous children in regional Australia', *Int J*
29 *Audiol*, 53: 326-35.
- 30 Cameron, S., H. Glyde, and H. Dillon. 2011. 'Listening in Spatialized Noise-Sentences Test
31 (LiSN-S): normative and retest reliability data for adolescents and adults up to 60 years
32 of age', *J Am Acad Audiol*, 22: 697-709.
- 33 Cameron, S., H. Glyde, and H. Dillon. 2012. 'Efficacy of the LiSN & Learn auditory training
34 software: randomized blinded controlled study', *Audiol Res*, 2: e15.
- 35 Christensen, D. L., L. A. Schieve, O. Devine, and C. Drews-Botsch. 2014. 'Socioeconomic
36 status, child enrichment factors, and cognitive performance among preschool-age
37 children: results from the Follow-Up of Growth and Development Experiences study',
38 *Res Dev Disabil*, 35: 1789-801.
- 39 Del Zoppo, C., L. Sanchez, and C. Lind. 2015. 'A long-term follow-up of children and
40 adolescents referred for assessment of auditory processing disorder', *Int J Audiol*, 54:
41 368-75.
- 42 Dettman, S. J., D. Pinder, R. J. Briggs, R. C. Dowell, and J. R. Leigh. 2007. 'Communication
43 development in children who receive the cochlear implant younger than 12 months: risks
44 versus benefits', *Ear Hear*, 28: 11S-18S.
- 45 Dillon, H., and S. Cameron. 2021. 'Separating the Causes of Listening Difficulties in Children',
46 *Ear Hear*, 42: 1097-108.

- 1 Fiscella, K., and D. R. Williams. 2004. 'Health disparities based on socioeconomic inequities:
2 implications for urban health care', *Acad Med*, 79: 1139-47.
- 3 Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis*
4 (John Wiley & Sons).
- 5 Harding, Jessica F, Pamela A Morris, and Diane Hughes. 2015. 'The relationship between
6 maternal education and children's academic outcomes: A theoretical framework', *Journal*
7 *of Marriage and Family*, 77: 60-76.
- 8 Hechtman, L., J. M. Swanson, M. H. Sibley, A. Stehli, E. B. Owens, J. T. Mitchell, L. E. Arnold,
9 B. S. Molina, S. P. Hinshaw, P. S. Jensen, H. B. Abikoff, G. Perez Algorta, A. L. Howard,
10 B. Hoza, J. Etcovitch, S. Houssais, K. D. Lakes, J. Q. Nichols, and M. T. A. Cooperative
11 Group. 2016. 'Functional Adult Outcomes 16 Years After Childhood Diagnosis of
12 Attention-Deficit/Hyperactivity Disorder: MTA Results', *J Am Acad Child Adolesc*
13 *Psychiatry*, 55: 945-52 e2.
- 14 Hind, S. E., R. Haines-Bazrafshan, C. L. Benton, W. Brassington, B. Towle, and D. R. Moore.
15 2011. 'Prevalence of clinical referrals having hearing thresholds within normal limits', *Int*
16 *J Audiol*, 50: 708-16.
- 17 Hjern, A., G. R. Weitoft, and F. Lindblad. 2010. 'Social adversity predicts ADHD-medication in
18 school children--a national cohort study', *Acta Paediatr*, 99: 920-4.
- 19 Humanski, R. A., and R. A. Butler. 1988. 'The contribution of the near and far ear toward
20 localization of sound in the sagittal plane', *J Acoust Soc Am*, 83: 2300-10.
- 21 Hunter, L. L., C. M. Blankenship, L. Lin, N. T. Sloat, A. Perdew, H. Stewart, and D. R. Moore.
22 2021. 'Peripheral Auditory Involvement in Childhood Listening Difficulty', *Ear Hear*, 42:
23 29-41.
- 24 Hunter, L. L., C. M. Blankenship, B. Shinn-Cunningham, L. Hood, L. M. Zadeh, and D. R.
25 Moore. 2023. 'Brainstem auditory physiology in children with listening difficulties()',
26 *Hear Res*, 429: 108705.
- 27 Keith, Robert W. 2009. *SCAN-3 for children: Tests for auditory processing disorders* (Pearson).
- 28 Little, Roderick JA. 1988. 'A test of missing completely at random for multivariate data with
29 missing values', *Journal of the American statistical Association*, 83: 1198-202.
- 30 Lohvansuu, K., M. Torppa, T. Ahonen, K. Eklund, J. A. Hamalainen, P. H. T. Leppanen, and H.
31 Lyytinen. 2021. 'Unveiling the Mysteries of Dyslexia-Lessons Learned from the
32 Prospective Jyvaskyla Longitudinal Study of Dyslexia', *Brain Sci*, 11.
- 33 Magnuson, Katherine A, Holly R Sexton, Pamela E Davis-Kean, and Aletha C Huston. 2009.
34 'Increases in maternal education and young children's language skills', *Merrill-Palmer*
35 *Quarterly (1982-):* 319-50.
- 36 Moeller, M. P. 2000. 'Early intervention and language development in children who are deaf and
37 hard of hearing', *Pediatrics*, 106: E43.
- 38 Moore, D. R., S. L. Sieswerda, M. M. Grainger, A. Bowling, N. Smith, A. Perdew, S. Eichert, S.
39 Alston, L. W. Hilbert, L. Summers, L. Lin, and L. L. Hunter. 2018. 'Referral and
40 Diagnosis of Developmental Auditory Processing Disorder in a Large, United States
41 Hospital-Based Audiology Service', *J Am Acad Audiol*, 29: 364-77.
- 42 Musiek, Frank E, and Gail D Chermak. 2007. 'Handbook of central auditory processing disorder',
43 (No Title).
- 44 Perrachione, T. K., S. N. Del Tufo, and J. D. Gabrieli. 2011. 'Human voice recognition depends
45 on language ability', *Science*, 333: 595.

- 1 Petley, L., L. L. Hunter, L. Motlagh Zadeh, H. J. Stewart, N. T. Sloat, A. Perdew, L. Lin, and D.
2 R. Moore. 2021. 'Listening Difficulties in Children With Normal Audiograms: Relation to
3 Hearing and Cognition', *Ear Hear*.
- 4 Pimperton, H., and C. R. Kennedy. 2012. 'The impact of early identification of permanent
5 childhood hearing impairment on speech and language outcomes', *Arch Dis Child*, 97:
6 648-53.
- 7 Pungello, E. P., I. U. Iruka, A. M. Dotterer, R. Mills-Koonce, and J. S. Reznick. 2009. 'The
8 effects of socioeconomic status, race, and parenting on language development in early
9 childhood', *Dev Psychol*, 45: 544-57.
- 10 Quinto, A., S. Abu El Adas, and S. V. Levi. 2020. 'Re-Examining the Effect of Top-Down
11 Linguistic Information on Speaker-Voice Discrimination', *Cogn Sci*, 44: e12902.
- 12 Reiss, F. 2013. 'Socioeconomic inequalities and mental health problems in children and
13 adolescents: a systematic review', *Soc Sci Med*, 90: 24-31.
- 14 Roy, A., L. Hechtman, L. E. Arnold, J. M. Swanson, B. S. G. Molina, M. H. Sibley, A. L.
15 Howard, and M. T. A. Cooperative Group. 2017. 'Childhood Predictors of Adult
16 Functional Outcomes in the Multimodal Treatment Study of Attention-
17 Deficit/Hyperactivity Disorder (MTA)', *J Am Acad Child Adolesc Psychiatry*, 56: 687-95
18 e7.
- 19 Rutter, Michael, A Bailey, and Catherine Lord. 2003. 'SCQ', *The Social Communication
20 Questionnaire. Torrance, CA: Western Psychological Services*.
- 21 Sharma, M., I. Dhamani, J. Leung, and S. Carlile. 2014. 'Attention, memory, and auditory
22 processing in 10- to 15-year-old children with listening difficulties', *J Speech Lang Hear
23 Res*, 57: 2308-21.
- 24 Sharma, M., S. C. Purdy, and A. S. Kelly. 2009. 'Comorbidity of auditory processing, language,
25 and reading disorders', *J Speech Lang Hear Res*, 52: 706-22.
- 26 Sharma, M., S. C. Purdy, and A. S. Kelly. 2012. 'A randomized control trial of interventions in
27 school-aged children with auditory processing disorders', *Int J Audiol*, 51: 506-18.
- 28 Shaywitz, Sally E., Jack M. Fletcher, John M. Holahan, Abigail E. Shneider, Karen E.
29 Marchione, Karla K. Stuebing, David J. Francis, Kenneth R. Pugh, and Bennett A.
30 Shaywitz. 1999. 'Persistence of Dyslexia: The Connecticut Longitudinal Study at
31 Adolescence', *Pediatrics*, 104: 1351-59.
- 32 Sherar, L. B., T. P. Griffin, U. Ekelund, A. R. Cooper, D. W. Esliger, E. M. van Sluijs, L. Bo
33 Andersen, G. Cardon, R. Davey, K. Froberg, P. C. Hallal, K. F. Janz, K. Kordas, S.
34 Kriemler, R. R. Pate, J. J. Puder, L. B. Sardinha, A. F. Timperio, and A. S. Page. 2016.
35 'Association between maternal education and objectively measured physical activity and
36 sedentary time in adolescents', *J Epidemiol Community Health*, 70: 541-8.
- 37 Skoe, E., J. Krizman, and N. Kraus. 2013. 'The impoverished brain: disparities in maternal
38 education affect the neural response to sound', *The Journal of neuroscience : the official
39 journal of the Society for Neuroscience*, 33: 17221-31.
- 40 Smoski, W, Michael A Brunt, and J Curtis Tannahill. 1998. 'Children's auditory performance
41 scale (CHAPS)', *Tampa, FL: Educational Audiology Association*.
- 42 St Clair, M. C., A. Pickles, K. Durkin, and G. Conti-Ramsden. 2011. 'A longitudinal study of
43 behavioral, emotional and social difficulties in individuals with a history of specific
44 language impairment (SLI)', *J Commun Disord*, 44: 186-99.

- 1 Stewart, H. J., E. K. Cash, L. L. Hunter, T. Maloney, J. Vannest, and D. R. Moore. 2022. 'Speech
2 cortical activation and connectivity in typically developing children and those with
3 listening difficulties', *Neuroimage Clin*, 36: 103172.
- 4 Sullivan, E. V., T. Brumback, S. F. Tapert, D. Prouty, R. Fama, W. K. Thompson, S. A. Brown, K.
5 Cummins, I. M. Colrain, F. C. Baker, D. B. Clark, T. Chung, M. D. De Bellis, S. R.
6 Hooper, B. J. Nagel, B. N. Nichols, W. Chu, D. Kwon, K. M. Pohl, and A. Pfefferbaum.
7 2017. 'Effects of prior testing lasting a full year in NCANDA adolescents: Contributions
8 from age, sex, socioeconomic status, ethnicity, site, family history of alcohol or drug
9 abuse, and baseline performance', *Dev Cogn Neurosci*, 24: 72-83.
- 10 Sweetow, Robert W., and Jennifer Henderson-Sabes. 2004. 'The case for LACE', *The Hearing*
11 *Journal*, 57: 32-35.
- 12 Taylor, B. K., M. R. Frenzel, J. A. Eastman, A. I. Wiesman, Y. P. Wang, V. D. Calhoun, J. M.
13 Stephen, and T. W. Wilson. 2022. 'Reliability of the NIH toolbox cognitive battery in
14 children and adolescents: a 3-year longitudinal examination', *Psychol Med*, 52: 1718-27.
- 15 Tomlin, D., H. Dillon, M. Sharma, and G. Rance. 2015. 'The Impact of Auditory Processing and
16 Cognitive Abilities in Children', *Ear Hear*, 36: 527-42.
- 17 Verbeke, Geert. 1997. 'Linear mixed models for longitudinal data.' in, *Linear mixed models in*
18 *practice* (Springer).
- 19 Weintraub, S., P. J. Bauer, P. D. Zelazo, K. Wallner-Allen, S. S. Dikmen, R. K. Heaton, D. S.
20 Tulskey, J. Slotkin, D. L. Blitz, N. E. Carlozzi, R. J. Havlik, J. L. Beaumont, D. Mungas, J.
21 J. Manly, B. G. Borosh, C. J. Nowinski, and R. C. Gershon. 2013a. 'I. NIH Toolbox
22 Cognition Battery (CB): introduction and pediatric data', *Monogr Soc Res Child Dev*, 78:
23 1-15.
- 24 Weintraub, S., S. S. Dikmen, R. K. Heaton, D. S. Tulskey, P. D. Zelazo, P. J. Bauer, N. E.
25 Carlozzi, J. Slotkin, D. Blitz, K. Wallner-Allen, N. A. Fox, J. L. Beaumont, D. Mungas,
26 C. J. Nowinski, J. Richler, J. A. Deocampo, J. E. Anderson, J. J. Manly, B. Borosh, R.
27 Havlik, K. Conway, E. Edwards, L. Freund, J. W. King, C. Moy, E. Witt, and R. C.
28 Gershon. 2013b. 'Cognition assessment using the NIH Toolbox', *Neurology*, 80: S54-64.
- 29 Yoshinaga-Itano, C., A. L. Sedey, D. K. Coulter, and A. L. Mehl. 1998. 'Language of early- and
30 later-identified children with hearing loss', *Pediatrics*, 102: 1161-71.
- 31
32

1 **Figure legends**

2

3 **Figure 1. All test scores were consistently lower in the LiD group across Waves 1 and 2. A.** Violin
4 plots of ECLiPS Total Scaled scores showing the probability density of the data. Violin plots are overlaid
5 with boxplots indicating each group's median and interquartile range in Waves 1 and 2. The horizontal
6 dashed line reflects the expected standard score (here, 10). Wave 3 data are not included in the figure due
7 to the significantly smaller sample compared with Waves 1 and 2 (see Tables 1-3). **B.** Early Childhood
8 Composite scores from the NIH Cognition Toolbox. **C.** LiSN-S Cue and Advantage scaled Z-Scores. LiD
9 indicates listening difficulty; TD, typically developing.

10

11 **Figure 2. Individual data of Wave 3 participants showed improved listening skills over 3 waves. A.**
12 Line plots of ECLiPS total scaled scores across 4 years. **B.** Same as A for NIH-Toolbox Early Childhood
13 Composite scaled scores. **C.** Same as A for LiSN-S Cue and Advantage scores. LiD indicates listening
14 difficulty; TD, typically developing.

15

16 **Figure 3.** Listening skills and cognitive abilities showed good to excellent correlations between Waves 1
17 and 2. Auditory skills showed poor agreement between waves. **A.** Scatter plots of ECLiPS total scaled
18 scores at Wave 1 and 2. Dashed lines show regression lines for each group. The black line indicates the
19 regression line for the entire data combining groups. **B.** Same as A for Early Childhood Composite. **C.**
20 Same as A for LiSN-S Cue and Advantage scores. LiD indicates listening difficulty; TD, typically
21 developing.

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

Running head: Longitudinal study of childhood listening difficulties

1 **Table 1. Participant numbers and demographic information**

	Wave 1		Wave 2		Wave 3	
	TD	LiD	TD	LiD	TD	LiD
N	79	68	48	52	15	16
Age, year (SD)	9.3 (2.0)	9.7 (1.9)	12.1 (2.2)	11.9 (2.0)	13.4 (1.9)	14.5 (2.2)
Female, n (%)	33 (42)	24 (35)	24 (50)	16 (31)	6 (40)	6 (38)
Maternal education, n (%)						
High school	1 (1)	12 (18)	1 (2)	9 (18.4)	0 (0)	3 (19)
College	8 (10)	17 (25)	3 (6)	15 (30.6)	0 (0)	6 (38)
Bachelor	50 (63)	25 (37)	31 (65)	16 (32.7)	11 (73)	3 (19)
Post-grad	20 (25)	14 (21)	13 (27)	9 (18.4)	4 (27)	4 (25)
Race, non-white (%)	13 (16)	17 (25)	5 (10)	16 (31)	1 (7)	9 (56)
Ethnicity, Hispanic or Latino (%)	3 (4)	4 (6)	2 (4)	3 (6)	1 (7)	2 (12)

2

Running head: Longitudinal study of childhood listening difficulties

1 **Table 2. Auditory and listening test scores**

Testing measures*	Wave 1		Wave 2		Wave 3	
	TD	LiD	TD	LiD	TD	LiD
	n = 79	n = 67	n = 47	n = 49	n = 13	n = 15
ECLiPS scores						
Speech and auditory processing	10.4 (1.7)	3.0 (1.7)	10.5 (2.5)	4.2 (2.6)	11.4 (2.2)	6.0 (3.3)
Environmental and auditory sensitivity	10.8 (2.5)	4.8 (2.6)	9.8 (2.7)	4.9 (2.7)	10.6 (3.1)	7.9 (3.2)
Language/literacy/laterality	10.5 (2.8)	3.9 (2.4)	10.2 (2.6)	3.9 (2.4)	10.9 (2.2)	4.9 (2.8)
Memory and attention	10.7 (2.5)	4.2 (2.0)	10.8 (3.0)	4.8 (2.7)	12.2 (3.5)	6.5 (3.4)
Pragmatic and social skills	10.9 (2.4)	5.1 (1.8)	10.9 (2.4)	5.7 (2.2)	12.1 (3.1)	7.9 (3.4)
Language	10.6 (2.4)	3.6 (2.1)	10.6 (2.9)	3.8 (2.4)	11.8 (3.2)	5.3 (2.9)
Listening	10.7 (2.4)	3.4 (1.7)	10.8 (2.6)	4.2 (2.3)	12.1 (3.1)	6.3 (3.4)
Social	11.0 (2.4)	3.6 (2.1)	10.6 (2.9)	3.8 (2.4)	11.6 (3.4)	7.7 (3.6)
Total	10.8 (2.4)	3.0 (1.8)	10.5 (2.7)	3.5 (2.3)	11.8 (3.1)	5.9 (3.2)
LiSN-S scores						
	n = 76	n = 65	n = 43	n = 47	n = 15	n = 15
Low Cue, z score	-0.13 (1.00)	-0.76 (1.18)	0.16 (1.16)	-0.96 (1.10)	0.33 (1.32)	-0.87 (1.08)
High Cue, z score	0.28 (1.04)	-0.39 (1.32)	0.43 (0.81)	-0.31 (1.26)	0.68 (0.91)	0.25 (1.10)
Talker Advantage, z score	-0.05 (0.81)	-0.54 (0.99)	0.20 (0.95)	-0.34 (0.96)	0.55 (1.05)	-0.55 (1.01)
Spatial Advantage, z score	-0.23 (1.36)	-0.66 (1.69)	0.24 (1.20)	-0.29 (1.52)	0.60 (1.07)	-0.17 (0.88)
Pattern Score	7.43 (1.83)	7.25 (2.14)	7.79 (1.13)	7.64 (2.09)	8.28 (1.46)	8.76 (1.23)

2 *All values are mean (SD) unless otherwise noted.

3
4
5
6
7
8

Running head: Longitudinal study of childhood listening difficulties

1 **Table 3. Cognitive test scores**

Testing measures*	Wave 1		Wave 2		Wave 3	
	TD	LiD	TD	LiD	TD	LiD
NIH Cognition Toolbox Composite Scores	n = 56	n = 49	n = 42	n = 45	n = 15	n = 15
Picture vocabulary	114 (14)	98 (13)	112 (14)	93 (11)	113 (15)	91(16)
Flanker test	101(15)	90 (14)	98 (12)	86 (14)	107 (15)	80 (12)
List sorting working memory [†]	110 (15)	94 (14)	110 (14)	92 (13)	117 (15)	89 (12)
Dimensional change card sort test	104 (15)	89 (13)	108 (19)	90 (15)	113 (21)	89 (17)
Pattern comparison processing speed test [†]	103 (22)	87 (20)	115 (22)	86 (22)	123 (19)	90 (22)
Picture sequence memory test	112 (19)	95 (20)	113 (18)	99 (13)	128 (17)	95 (14)
Oral reading recognition test [†]	107 (13)	90 (12)	110 (12)	92 (11)	106 (14)	85 (11)
Fluid cognition composite [†]	109 (18)	85 (17)	114 (17)	84 (17)	126 (14)	81 (18)
Crystallized cognition composite [†]	112 (14)	93 (12)	113 (13)	91 (10)	111 (15)	86 (14)
Total composite [†]	112 (15)	87 (14)	116 (15)	85 (13)	123 (15)	80 (15)
Early Childhood composite	112 (16)	89 (14)	112 (16)	87 (13)	122 (16)	82 (18)

2 *All values are mean (SD) unless otherwise noted.

3 [†]Data missing for 9 children in Wave 1 (TD: n = 51, LiD: n = 45).

	Pearson correlation coefficient	ICC (95% CI), Unadjusted	ICC (95% CI), Adjusted by Wave	Wave 2 – Wave 1			
				N	Mean	SD	Mean/SD
ECLIPS total scaled score	0.85	0.86 (0.80, 0.90)	0.86 (0.80, 0.90)	97	0.10	2.33	0.04
NIH-toolbox age-corrected standard score							
Picture vocabulary	0.66	0.65 (0.51-0.77)	0.65 (0.52-0.77)	81	-2.26	12.48	-0.18
Flanker test	0.51	0.42 (0.26-0.60)	0.48 (0.32-0.64)	79	-5.59	15.17	-0.37
List sorting working memory	0.66	0.68 (0.53-0.79)	0.68 (0.55-0.80)	75	-1.99	13.48	-0.15
Dimensional change card sort test	0.48	0.53 (0.35-0.69)	0.52 (0.35-0.69)	79	0.12	17.97	0.01
Pattern comparison processing speed test	0.58	0.56 (0.40-0.71)	0.56 (0.40-0.70)	71	4.58	27.02	0.17
Picture sequence memory test	0.44	0.45 (0.27-0.64)	0.45 (0.27-0.64)	81	2.46	20.73	0.12
Oral reading recognition test	0.70	0.68 (0.54-0.79)	0.70 (0.57-0.80)	74	2.66	11.23	0.24
Fluid cognition composite	0.74	0.75 (0.64-0.84)	0.76 (0.64-0.84)	68	-0.03	16.79	0.00
Crystallized cognition composite	0.80	0.79 (0.69-0.86)	0.79 (0.69-0.86)	72	0.84	9.66	0.09
Total composite	0.82	0.82 (0.73-0.88)	0.82 (0.74-0.88)	66	0.51	12.58	0.04
Early childhood composite	0.73	0.73 (0.61-0.82)	0.74 (0.63-0.83)	74	-2.10	13.89	-0.15
LiSN-S, z score							
Low Cue A	0.29	0.27 (0.13-0.49)	0.28 (0.13-0.49)	89	0.41	2.13	0.19
High Cue	0.35	0.32 (0.17-0.52)	0.33 (0.18-0.53)	89	0.23	1.40	0.16
Talker Advantage	0.32	0.29 (0.14-0.51)	0.33 (0.17-0.54)	89	0.18	1.26	0.14
Spatial Advantage	0.45	0.39 (0.24-0.56)	0.42 (0.27-0.58)	89	0.33	1.69	0.20

1 **Table 4. Pearson correlation coefficients, estimated reliability coefficients, and test-retest differences between waves 1 and 2**

2

3





