Childhood listening and associated cognitive difficulties persist into adolescence

Katsuaki Kojima[1,2,3], Li Lin[3], Lauren Petley[3,4], Nathan Clevenger[3], Audrey Perdew[3], Mark Bodik[3,5], Chelsea M. Blankenship[3, 6], Lina Motlagh Zadeh[3], 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
8. Departments of Otolaryngology and Neuroscience, University of Cincinnati College of Medicine, 260 Stetson St., Cincinnati, OH 45267
9. Manchester Centre for Audiology and Deafness, University of Manchester, Oxford Rd, Manchester M13 9PL, UK

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

Corresponding author: Katsuaki.kojima@cchmc.org

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

ACKNOWLEDGEMENTS

This research was supported by grant 5R01DC014078 from the National Institute of Deafness and other Communication Disorders (DRM), 2UL1TR001425-05A1 by the National Center for Advancing Translational Sciences of the National Institutes of Health (KK), and by the Cincinnati Children's Research Foundation. DRM received support from NIHR Manchester Biomedical Research Centre. The funding organization had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article. We would like to acknowledge that an earlier version of this manuscript was published as a preprint on medRxiv (https://doi.org/10.1101/2022.08.11.22278673).

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

Abstract (464 words /500 words)

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

Design: This longitudinal study had three data collection waves for children with LiD and TD aged 6-13 years at Wave 1, followed by assessments at 2-year (Wave 2) and 4-year (Wave 3) intervals. The primary analysis focused on data from Waves 1 and 2, while a secondary analysis encompassed all three waves despite high attrition at Wave 3. Caregivers completed the ECLiPS, while participants underwent the Listening in Spatialized Noise Sentences test (LiSN-S) and the NIH Toolbox Cognition Battery during each wave. The analysis consisted of: 1) examining longitudinal differences between TD and LiD groups in demographics, listening, auditory, and cognitive function; 2) identifying the functional domains contributing to LiD longitudinally; and 3) evaluating the test-retest reliability of the measures across waves. Mixed-effect models were employed to analyze longitudinal data.

Results: The study enrolled 169 participants, with 147, 100, and 31 children completing the required testing during Waves 1, 2, and 3 study visits, respectively. The mean ages at these waves were 9.5, 12.0, and 14.0 years old. Children with LiD consistently underperformed TD children in auditory and cognitive tasks across all waves. Maternal education, auditory and cognitive abilities independently predicted caregiver-reported listening skills. The measures exhibited significant correlations between Waves 1 and 2, displaying high reliability. Longitudinal analysis of Wave 3 participants corresponded with the primary analyses of Waves 1 and 2, reinforcing the enduring nature of listening difficulties.

Conclusion: Children with LiD and clinically normal audiograms experience persistent auditory, listening, and cognitive challenges through at least adolescence. The degree of LiD can be independently

predicted by maternal education, cognitive processing, and spatial listening skills. This study underscores the importance of early detection and intervention for childhood LiD and highlights the role of socioeconomic factors as contributors to these challenges.

INTRODUCTION

Listening is commonly defined as an active form of hearing. It facilitates an accurate understanding of speech, a key component of everyday communication. Successful communication allows healthy language development, as seen in improved language outcomes in children with hearing loss who received early interventions (Pimperton and Kennedy 2012). Problems with listening may similarly predispose children to challenges in language acquisition. It is difficult, however, to determine the direction of the causal pathway because children with primary language or other cognitive problems may have difficulties comprehending speech that prevents them from active hearing (listening). Listening difficulty (LiD) describes difficulty hearing and understanding speech and other suprathreshold sounds (Petley et al. 2021; Dillon and Cameron 2021). This is an umbrella term that describes symptoms, and the underlying causes of LiD can be problems in multiple domains, including hearing, language, cognition, or auditory processing (Dillon and Cameron 2021). Although some children presenting with LiD have hearing impairments, as indicated by elevated audiometric thresholds, many children with LiD have clinically normal audiometric sensitivity (Hind et al. 2011). To clarify the contribution of hearing and cognitive factors to LiD for individuals with clinically normal audiograms, we prospectively studied children with LiD, identified using a validated and standardized caregiver questionnaire, Evaluation of Children's Listening & Processing Skills (ECLiPS) (Barry and Moore 2021).

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

Whether LiD is a transient childhood problem that resolves with age or persists into adulthood is unclear. Answering this question is essential for developing management strategies and understanding the mechanisms and consequences of LiD. Persistent LiD may have a more significant impact on the individual, family, and society over a long time. More comprehensive assessment and more aggressive treatment strategies may therefore be warranted. Young adults with a history of LiD in childhood reported continuing communication difficulty almost 10 years after a clinical diagnosis (Del Zoppo et al. 2015).

Still, longitudinal quantitative data on listening, auditory, and cognitive abilities in children with earlier LiD is lacking.

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

Evaluating longitudinal cognitive performance during development can be challenging because of the different rates of development and potential practice effects (Sullivan et al. 2017). As part of the SICLiD study, we performed behavioral testing in the same cohort of children every two years to assess longitudinal changes. The NIH Cognition Toolbox is one test we used in the SICLiD study to assess cognition. Weintraub et al. showed that the Toolbox test results demonstrated high reliability in pediatric samples with short testing intervals between 7 and 21 days, with intraclass correlations (ICCs) between 0.76 and 0.99 (Weintraub et al. 2013b). However, the results of recent literature on the reliability for longer intervals (> 1 year) are mixed, with ICCs between 0.24 and 0.85 (Taylor et al. 2022) and 0.43 and 0.82 (Anokhin et al. 2022). The longitudinal data of the SICLiD study are suitable for evaluating the reliability of the NIH Cognition Toolbox over the years.

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

MATERIALS AND METHODS

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

Participants

The SICLID study is a longitudinal cohort study evaluating children with LiD and TD children with repeated assessment every two years. LiD and TD children were 6 to 13 years old at enrollment. The LiD group was recruited from children referred to CCH for LiD evaluation and from the community. All the participants in the TD group were recruited from the community. Recruitment from the community involved flyers posted in relevant CCH clinics (Audiology, Pediatrics, Speech-Language Pathology) and emails sent to all CCH employees and families interested in research. Interested caregivers completed (i) the background questionnaire and (ii) the ECLiPS (Barry and Moore 2021) to confirm eligibility for the LiD or TD group. Total standardized ECLiPS scores of ≥ 7 defined the TD group, and scores < 7 defined the LiD group. However, four children with LiD had a previous audiologic diagnosis of auditory processing disorder (APD) but scored 7 and 9 on the ECLiPS at enrollment. APD is one of the known underlying mechanisms of LiD (Dillon and Cameron 2021). Because of their diagnosis of APD, they were assigned to the LiD group. The inclusion criterion for both groups was English as the child's native language. Exclusion criteria for both groups were reported otologic (e.g., hearing loss), neurologic (e.g., epilepsy, stroke), psychiatric (e.g., attention deficits hyperactive disorder (ADHD) requiring medical treatment), developmental (e.g., autism spectrum disorder (ASD)), or language disorders (e.g., dyslexia), or severe intellectual insufficiency that would prevent or restrict their ability to complete testing procedures. Eligibility was determined based on caregiver responses to the background questionnaire. The TD group was age- and sex-matched by proportional sampling.

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

Procedure

All caregivers of eligible and interested participants reviewed the informed consent form with a study staff member and discussed the purpose, procedures, risks and benefits, duration, and expectations involved in the study. According to institutional policy, children aged 11 and above also assented using a child-friendly version of the consent document. All participants received financial compensation for their

participation. Each participant completed audiometric and behavioral testing during the study visit, as listed below. Several additional tests, including brainstem and cortical evoked responses and structural and functional MRI, were also performed during the visit as part of the SICLiD study. The results of these additional tests are beyond the scope and not included in the current manuscript. Some of the results of these additional tests were reported elsewhere (Hunter et al. 2021; Stewart et al. 2022).

Background Questionnaire

The study staff developed a background questionnaire to collect information about the study participants, including demographic information, medical and surgical history, and socioeconomic status. This questionnaire lacks reliability or validity data because of its basic nature. The specific information collected by this questionnaire was age, sex, race, ethnicity, maternal education (a proxy for socioeconomic position (Sherar et al. 2016)), the primary language used at home, and diagnosis and treatment received for otologic, neurologic, behavioral, developmental, and psychiatric problems. The questionnaire includes specific medical conditions as examples, such as ADHD, ASD, and dyslexia. Caregivers completed the same background questionnaire at each Wave. Demographic variables at each time point are summarized in **Table 1**.

Caregiver Evaluation of Children's Listening

The ECLiPS (Barry and Moore 2021) was administered at each wave to assess participants' listening and communication abilities. This questionnaire consists of 38 straightforward statements (items) describing behaviors commonly observed in children. Caregivers rated their child on a five-point scale, ranging from strongly disagree to strongly agree, based on their agreement with each statement. These ratings were averaged to derive age-scaled scores for five subscales: speech and auditory processing (SAP), environmental and auditory sensitivity (EAS), language/literacy/laterality (L/L/L), memory and attention (M&A), and pragmatic and social skills (PSS), each containing 6 to 9 distinct items. The subscales were further grouped into Language, Listening, Social, and Total aggregate (composite) scores, all of which were standardized with a population mean of 10 (SD = 3) based on British data (Barry et al. 2015). All ECLiPS subscales exhibit high test-retest reliability, with intraclass correlations (ICCs) above 0.8 (Barry and Moore 2021). Construct validity has been demonstrated through convergence with other established tests measuring similar skills, including the Children's Auditory Processing Performance Scale (CHAPS; (Smoski et al. 1998)), the CCC-2 (Bishop 2006), and the Social Communication Questionnaire (Rutter et al. 2003).

As mentioned in the introduction, LiD is an umbrella term describing a symptom potentially associated with abnormalities in auditory, auditory processing, cognitive, and/or language domains.

Although the ECLiPS provides scores for five latent factors related to listening (SAP, EAS, L/L/L, M&A, and PSS), these scores are not validated for diagnosing specific impairments in each area. Moreover, our study's objective is to understand the LiD irrespective of the underlying causes. Consequently, eligibility for inclusion in the LiD category was determined based on the ECLiPS Total composite score (below 7, which defines a cutoff of < 1 SD from the population mean). The Total composite score represents the overall degree of LiD (Barry and Moore 2021) and is appropriate for this purpose.

Audiometry

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

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

The Listening in Spatialized Noise – Sentences (LiSN-S) test (Cameron and Dillon 2007; Brown et al. 2010) measures the ability to listen and repeat simple, spoken sentences in the presence of distracting sentences, developed using the same child-friendly criteria as the BKB sentences (Bench et al. 1979). We selected LiSN-S as the auditory assessment tool for the SICLiD study because of its multifaceted approach to evaluating spatial and speech-hearing abilities, as well as its specific design for use in pediatric populations. The LiSN-S (US Edition (Brown et al. 2010)) was administered at each wave using a commercial CD played on a laptop, a task-specific soundcard, and Sennheiser HD 215 headphones. Participants were asked to repeat a series of target sentences, presented using virtual space generic head-related transfer functions (Humanski and Butler 1988), directly in front (0°: diotic) while ignoring two distracting speakers. There are four listening conditions in which the distractors change voice (different or same as target) and/or position (0° and 90°) (Humanski and Butler 1988). The LiSN-S is adaptive; the level of the target speaker decreases or increases in SNR relative to the distracting speech if the listener responds correctly or incorrectly during up to 30 sentences in each condition. The speech reception threshold (SRT) represents 50% correct SNR for the condition. The Low Cue condition is where distractors are in the same voice as the target and 0° relative to the listener. Distractors are in different voices and ±90° relative to the listener in the High Cue condition. The derived scores of the LiSN-S are the Talker Advantage and Spatial Advantage, so-called because each is the difference

between SRTs from two conditions. This subtraction process should, to varying extent, separate auditory from cognitive influences (Cameron and Dillon 2007; Petley et al. 2021). Spatial Advantage thus represents the participant's ability to separate the spatial source of the speech using purely auditory cues, while Talker Advantage demonstrates the participant's ability to distinguish individual talkers, which involves both acoustic and linguistic information (Perrachione et al. 2011; Quinto et al. 2020). Test-retest comparisons for the four listening conditions of the LiSN-S showed significant improvement from the first test to the second but no differences in the advantage scores (Cameron et al. 2011).

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

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

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

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

Cognition (National Institutes of Health, NIH Toolbox)

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

All participants in this study completed the Picture Vocabulary Test (PVT), Flanker Inhibitory Control and Attention Test (Flanker), Dimensional Change Card Sort test (DCCS), and Picture Sequence Memory test (PSMT). Each test produced a US age-corrected standardized score. The scores from all four tests were combined to calculate a single Early Childhood Composite (ECC), which provides a measure of overall cognitive skill for children aged six years and older. The PVT is an adaptive test that evaluates the participants' vocabulary skills. Participants listen to an audio recording of a word and select one of four pictures that closely matches the word's meaning. The Flanker assesses the participant's ability to inhibit responses and focus their attention. Participants report the direction of a central visual stimulus in a series of five similar, flanking stimuli, where the flanking stimuli may either be congruent or incongruent. The DCCS tests cognitive flexibility (attention switching). Target and test card stimuli vary

along two dimensions, shape, and color. Participants are asked to match test cards to the target card according to a specified dimension that varies for each trial. The PVT and DCCS scores both measure accuracy and reaction time. The PSMT assesses episodic memory by presenting illustrated objects and activities along with audio-recorded descriptive phrases. Participants are scored on the number of adjacent picture pairs correctly remembered over two learning trials. The length of the picture sequences varied from 6 to 18 pictures based on the participant's age. The ECC score derived from these tests allows for a comprehensive evaluation of overall cognitive performance in early childhood.

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

Analysis

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

In Part 1 of the analysis, we investigated differences between two groups, TD and LiD on their demographics, auditory, and cognitive function. Demographic information included age, sex, race, ethnicity, and maternal education. Test scores of ECLiPS, LiSN-S, and NIH Cognition Toolbox were compared. We used mixed-effect models (Verbeke 1997; Fitzmaurice et al. 2012) to analyze the longitudinal data. These models can handle repeated measures and are flexible on missing data. We assumed the data were missing at random (MAR) and used Restricted Maximum Likelihood (REML) estimation. We tested the MAR assumption with Little's test (Little 1988). A mixed-effect model has fixed effects and random effects. Fixed effects are variables that have a consistent effect on the outcome variable, while random effects are variables that have a varying effect on the outcome variable depending

on the level of another variable. Mixed-effect models are useful for analyzing data that have repeated measurements or hierarchical structures. Mixed-effect models can use both dependent and independent variables of multiple time points by modeling the change in the outcome variable over time as a function of fixed and random effects. This allows accounting for within-subject correlation and between-subject variations in the data.

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

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

In Part 2 of our analysis, our objective was to identify the functional domains that contributed to LiD in the longitudinal data of all participants. In our prior cross-sectional study, we developed a model that predicted 42% of the variance in ECLiPS Total Scaled scores at Wave 1 (Petley et al. 2021). This model included four predictors: the SCAN-3:C Composite Score, the LiSN-S Talker Advantage score, and the NIH PV and DCCS scores. Here, we tested if this model predicted ECLiPS Total Scaled scores at Wave 2. Next, we developed a new mixed-effect model using the longitudinal data to predict ECLiPS Total Scaled scores. Candidate predictors were maternal education, five LiSN-S scores (Cue, Advantage, and Pattern scores), and measures of cognitive function (either the seven NIH Cognition Toolbox subtests

or the four NIH Toolbox composite scores). Separate models were tested for the NIH Cognition Toolbox subtests and composite scores to avoid including duplicate variables in the model. Owing to its smaller sample size, Wave 3 was omitted from this longitudinal model.

In the mixed-effect model for Part 2 of the analysis, all candidate predictors were initially examined against the ECLiPS Total Scaled score through univariate analyses using a mixed-effect model, with participants used as a random effect and predictor variables and the interaction term (predictor x wave) as fixed effects. Only variables with $p < 0.1$ in the univariate analysis were included in the backward selection. Pairwise comparisons among the levels of significant variables were adjusted for using Tukey-Kramer multiple adjustment. Beta weights for each predictor variable were calculated to indicate the strength of their effects on the ECLiPS Total Scaled score.

The aim of Part 3 of the analysis was to evaluate the test-retest reliability of the measures. To achieve this, Pearson correlation coefficients and intra-class correlation coefficients (ICC) were computed to investigate the consistency of the test outcomes between Waves 1 and 2 for all participants. Wave 3 data were excluded from the analysis due to the limited sample size. Furthermore, to determine whether there were any variations in the correlation coefficients between the TD and LiD groups, the coefficients were calculated separately for each group, and z-tests were performed. To avoid type I errors, the p-values were adjusted for multiple comparisons using the B-H adjustment method (Benjamini and Hochberg 1995) with an FDR of less than 5%. B-H adjustment was applied to four LiSN-S scores, seven NIH Cognition Toolbox sub-scores, four NIH Cognition Toolbox composite scores, and eight ECLiPS sub-scores. All analyses used SAS version 9.4 (SAS Institute, Cary, N.C.), setting the two-sided significance level at 0.05.

RESULTS

LiD group associated with lower maternal education

Table 1 shows the number and demographics of participants in each group and wave. As Part 1 of our analysis, we compared the TD and LiD groups on their demographics, auditory, and cognitive function. Among the demographic variables, only maternal education differed significantly between the two groups ($p < .0001$), with a significant Group x Wave interaction observed (Wave 1 vs. Wave 3, $p < .0001$; see Table 1). Specifically, the TD group consistently had higher maternal education than the LiD group across all waves, and this difference increased from Wave 1 to Wave 3. To address the potential confounding effect of maternal education, we included it as a covariate in the subsequent analyses.

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

As expected, the LiD group had lower ECLiPS scores than the TD group at Wave 1 (see Figure 1A and Table 2). This finding is consistent with our prior cross-sectional study (Petley et al. 2021). The difference between the groups persisted at Wave 2 (group effect, $p < .0001$). There was no significant change in ECLiPS scores across the two waves for either group after B-H adjustment (wave effect, p values for each sub-score ranging from 0.01 to 0.67, higher than the B-H critical p-value 0.0063). Furthermore, there was no significant Group x Wave interaction effect, after B-H adjustment ($p = 0.02 – 0.91$), indicating that the listening skills of both groups remained stable over time.

LiD group had consistently lower cognitive scores

The LiD group scored significantly lower than the TD group on all seven sub-tests and four composite scores of the NIH Cognition Toolbox across Waves 1 and 2 (group effect, $p < .0001$). As shown in Figure 1B and Table 3, the scores did not change significantly over time for either group following B-H adjustment (wave effect, p values ranging from 0.008 to 1.00 for each sub-score; Group x Wave interaction, p values 0.01 - 0.84; B-H critical p-value 0.0071).

LiD group had consistently lower skills in segregating sentences

The LiD group had lower LiSN-S z-scores than the TD group across waves (see Figure 1C and Table 2; group effect, $p < .015$; Group x Wave interaction, $p = 0.12 – 1.00$). Spatial Advantage scores increased significantly from Wave 1 to Wave 2 for both groups ($p = .007$). Other LiSN-S z-scores did not show any significant change over time. Spatial Pattern Scores were similar between groups and waves (group effect, $p = .41$; wave effect, $p = .11$; Group x Wave interaction, $p = .86$). Notably, fewer children met the diagnostic criteria for SPD based on Pattern Scores at Wave 2 than at Wave 1: only three children in the LiD group compared to thirteen children in total at Wave 1 (7 LiD, 6 TD; $p = 0.57$ for Wave 1 and $p = 0.24$ for Wave 2).

Differences between the TD and LiD groups persisted through adolescence

We examined longitudinal changes in listening skills, speech-in-speech skills, and cognitive function across three waves based on 30 participants who completed all three waves (Figure 2). The TD group had higher ECLiPS sub-scores and Total Scaled scores than the LiD group (group effect, $p < .0001$), as in the main analysis of Waves 1 and 2 data. However, both groups improved their listening skills over time. ECLiPS speech and auditory processing sub-score (wave effect $p = .004$; Group x Wave interaction, $p = 0.18$) and composite Listening score (wave effect $p = .003$; Group x Wave interaction, $p = 0.93$) were significantly higher in Wave 2 than Wave 1 across both groups. Moreover, ECLiPS Total Scaled score in Wave 3 was significantly higher than Wave 1 (wave effect, $p = 0.028$; Group x Wave interaction, p =

0.71). In contrast to the listening skills, there was no change in cognitive function or speech-in-speech skills over time for either group. The TD group had significantly higher NIH Cognition Toolbox scores and two LiSN-S z scores (Low Cue and Talker Advantage) than the LiD group (group effect, $p < .003$). There was no significant main wave effect or Group x Wave interaction for either NIH Cognition Toolbox or LiSN-S scores. Participants who did not participate in Wave 2 were random (Little's $p = 0.74$). However, those who did not participate in Wave 3 were non-random (Little's $p = 0.01$). This suggests systemic reasons for participant attrition.

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

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

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

As Part 3 of the analysis, we evaluated the test-retest reliability of the measures (Figure 3; Table 4). The ECLiPS Total Scaled scores showed good reliability across Waves 1 and 2, with no differences between TD and LiD groups ($r = 0.52$ for both groups, $p = 1.000$). The NIH Toolbox sub-tests and composite scores also had significant and consistent correlations across waves (ICCs for each sub-test ranging from 0.48 to 0.82) and groups (p values from z-tests ranging from 0.07 to 0.96). Lastly, the LiSN-S scores revealed a moderate but significant agreement between waves, with ICCs ranging from 0.28 to 0.42.

DISCUSSION

Impact of study

We found that ECLiPS, LiSN-S, and NIH Toolbox scores were stable over time and that children with LiD, identified by the ECLiPS caregiver questionnaire, performed worse than TD children on auditory and cognitive tasks over multiple years. Moreover, we found that maternal education was a

significant predictor of caregiver-reported listening skills, independent of auditory and cognitive skills. This suggests that socioeconomic factors may play a crucial role in this complex condition. Our findings of the enduring nature of LiD have clinical implications for the early detection and treatment of childhood LiD, which may have long-term effects on affected individuals.

Socioeconomic factors in listening difficulties

The current study is the first to show how socioeconomic factors affect listening skills in children with LiD after partially controlling for auditory and cognitive skills measured by the LiSN-S test and NIH Toolbox scores. Previous studies have linked maternal education to various aspects of child development, such as children's academic outcomes (Harding et al. 2015), language skills (Magnuson et al. 2009), attention problems (Hjern et al. 2010), and auditory brainstem responses (Skoe et al. 2013). Therefore, it is not surprising that maternal education was associated with the degree of LiD in our study, potentially mediated by these factors. However, our finding is noteworthy because we controlled for cognitive and sensory functions using the LiSN-S test and NIH Toolbox scores, which suggests that maternal education may have an independent effect on LiD beyond these factors. Alternatively, maternal education may influence LiD through other factors that were not captured by our measures. Future studies should explore the underlying mechanism of LiD and include socioeconomic factors in their evaluation. Maternal education and socioeconomic factors should also be considered in the clinical assessment of children with LiD, as they may inform the risk of LiD and guide the clinical management, including the frequency of follow-up and intervention. Clinicians can also provide more support and guidance for parents with lower socioeconomic status who may face more challenges in accessing resources or implementing home-based activities for their children. They can also refer children with lower socioeconomic status to other services such as social workers or community programs that can help them overcome some of their disadvantages.

Persistent listening challenges

Our study provides evidence that children with LiD continue to experience persistent listening challenges over time, as indicated by significantly lower ECLiPS scores compared to TD children across three waves of testing. Our finding accounts for the stability or variability of scores in each participant over time as we used mixed-effect models to evaluate the longitudinal score changes of each participant. Importantly, our findings also suggest that the difference between TD and LiD groups is not due to a supra-normal performance of TD children, as their ECLiPS scores were consistent with other independent samples of children without LiD (Barry et al. 2015). These results are consistent with previous research (Del Zoppo et al. 2015) and underscore the need for early diagnosis and aggressive management to minimize the long-term effects of LiD. Our findings also raise questions about the effectiveness of current

management strategies for LiD, which may not adequately address the unique needs of these children. In a separate study (Hunter et al., in preparation), we found that more than half of the children in the LiD group had received specialist assessment and interventions before and/or during the study period. To better evaluate the efficacy of these interventions, future research should use quantitative outcome measures specific to each treatment. This will help identify the most effective strategies for addressing the persistent listening challenges experienced by children with LiD.

Cognitive challenges

Cognitive challenges of LiD participants were persistent throughout the study period, as shown by our prior Wave 1 study (Petley et al. 2021) and the longitudinal data from the current study. The current study also demonstrated that LiD participants performed worse than controls on all the NIH Toolbox subscale and composite scores. Moreover, our prediction model revealed that the NIH Total Composite score was the most predictive cognitive measure for LiD surpassing any combination of NIH Toolbox subscales. These results confirmed the broad and complex cognitive challenges associated with LiD. It is worth noting that the LiD group in our study included various underlying etiologies, as we used a low ECLiPS Total Scaled score as an inclusion criterion. Consequently, our results are relevant to children with LiD as a whole, rather than to a specific underlying condition of LiD. To address their distinct needs, individual children with specific deficits should receive targeted treatment (Dillon and Cameron 2021).

Spatial hearing

In the main analysis of Wave 1 and 2 data, the TD group outperformed the LiD group on all age-adjusted LiSN-S scores. Our purpose of using LiSN-S was to characterize the specific pattern of impairment in LiD participants using derived scores of LiSN-S that represent different aspects of listening abilities. In our previous study of Wave 1 data, we observed that TD children scored higher than LiD children only on two scores of LiSN-S, namely Low Cue and Talker Advantage (Petley et al., 2021). These scores reflect the ability to use cognitive strategies rather than auditory information to comprehend speech in a noisy environment. Our main analysis of Wave 1 and 2 data confirmed this finding and also discovered that TD children outperformed LiD children on Spatial Advantage, which measures sensory function by assessing the ability to use spatial cues to separate speech. Our results highlight the significant challenges that LiD children face across multiple domains of listening ability. Interestingly, in the secondary analysis of Wave 3 participants in the current study, children with LiD scored lower in Low Cue and Talker Advantage only, similar to the prior study (Petley et al., 2021). Taken together, the data

suggest that Low Cue and Talker Advantage may be more sensitive measures for differentiating TD and LiD than other LiSN-S scores.

Although the number of cases was small, the prevalence of SPD decreased at Wave 2, even with ongoing listening, auditory, and cognitive challenges in the LiD group and despite none of the participants receiving specific treatment for SPD (Cameron et al. 2012). Furthermore, neither the Pattern score nor the frequency of SPD differed significantly between the TD and LiD groups, suggesting that, at least in our cohort, SPD does not represent the listening problems identified by the ECLiPS (Petley et al. 2021). Future research clarifying the clinical significance and natural history of SPD would be beneficial. Among LiSN-S scores, only the Spatial Advantage age-adjusted scores increased significantly from Waves 1 to 2 in both groups. This could be due to developmental changes or inconsistent normalization and is unlikely to be training effects after a 2-year separation between tests.

Reliability

A prior longitudinal study of childhood LiD was based on a self-reported questionnaire with a follow-up response rate of approximately 13% (Del Zoppo et al. 2015). The current study used a large and heterogeneous sample of children and had a follow-up rate at Wave 2 of 68%. Additionally, loss to follow-up at Wave 2 was random, suggesting a low likelihood of sampling bias. Therefore, findings from Waves 1 and 2 likely represent the typical progression of LiD. Unfortunately, due to COVID, the follow-up rate of Wave 3 dropped significantly, and sampling bias was introduced. A separate longitudinal analysis of Wave 3 participants showed overall longer-term results that were consistent with the primary analysis of Waves 1 and 2.

ECLiPS, LiSN-S, and NIH Toolbox scores all showed statistically significant correlations between Waves 1 and 2. These correlations did not vary between the TD and LiD groups. The high reliability of the Toolbox scores observed in this study, even with long testing intervals of two years, seems to alleviate recent concerns regarding these scores (Taylor et al. 2022). This may be attributable to our consistent testing procedure carried out by research staff at a single site, as opposed to testing across multiple sites (Taylor et al. 2022). Moreover, LiSN-S age-adjusted scores were significantly correlated between Waves 1 and 2. Nonetheless, the agreement between waves ranged from poor to good, and correlation coefficients were generally lower than those reported in a previous study (Cameron et al. 2011). The higher retest reliability for that previous study may be explained by its shorter testing intervals (2-4 months) and older (above 12 years old) participants (Cameron et al. 2011).

Limitation

Due to the small sample size of Wave 3, we ran a separate analysis focusing on longitudinal changes of Wave 3 participants to reduce sampling bias. Though this method reduced bias, the analysis remained underpowered. Despite this limitation, our overall research findings are significant because we showed long-lasting challenges of childhood LiD that may influence clinical management strategy, and we identified a social factor, maternal education, that independently contributes to LiD.

Conclusion

Children with LiD and clinically normal audiograms have persistent auditory, listening, and cognitive challenges through at least early adolescence. The degree of LiD can be independently predicted by maternal education, cognitive processing, and spatial listening skills.

REFERENCES

Anokhin, Andrey P., Monica Luciana, Marie Banich, Deanna Barch, James M. Bjork, Marybel R. Gonzalez, Raul Gonzalez, Frank Haist, Joanna Jacobus, Krista Lisdahl, Erin McGlade, Bruce McCandliss, Bonnie Nagel, Sara Jo Nixon, Susan Tapert, James T. Kennedy, and Wesley Thompson. 2022. 'Age-related changes and longitudinal stability of individual differences in ABCD Neurocognition measures', *Developmental Cognitive Neuroscience*, 54.

Barry, J. G., and D. R. Moore. 2021. 'ECLIPS: Evaluation of children's listening and processing skills (2nd ed.)', *Cincinnati Children's Hospital*.

Barry, J. G., D. Tomlin, D. R. Moore, and H. Dillon. 2015. 'Use of Questionnaire-Based Measures in the Assessment of Listening Difficulties in School-Aged Children', *Ear Hear*, 36: e300-13.

Bench, J., A. Kowal, and J. Bamford. 1979. 'The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children', *Br J Audiol*, 13: 108-12.

Benjamini, Yoav, and Yosef Hochberg. 1995. 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal statistical society: series B (Methodological)*, 57: 289-300.

Bishop, Dorothy VM. 2006. *CCC-2: Children's Communication Checklist-2* (Pearson).

Brown, D. K., S. Cameron, J. S. Martin, C. Watson, and H. Dillon. 2010. 'The North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): normative data and test-retest reliability studies for adolescents and young adults', *J Am Acad Audiol*, 21: 629-41.

Cameron, S., and H. Dillon. 2007. 'Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)', *Ear Hear*, 28: 196-211.

Cameron, S., and H. Dillon. 2011. 'Development and evaluation of the LiSN & learn auditory training software for deficit-specific remediation of binaural processing deficits in children: preliminary findings', *J Am Acad Audiol*, 22: 678-96.

Cameron, S., H. Dillon, H. Glyde, S. Kanthan, and A. Kania. 2014. 'Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia', *Int J Audiol*, 53: 326-35.

Cameron, S., H. Glyde, and H. Dillon. 2011. 'Listening in Spatialized Noise-Sentences Test (LiSN-S): normative and retest reliability data for adolescents and adults up to 60 years of age', *J Am Acad Audiol*, 22: 697-709.

Cameron, S., H. Glyde, and H. Dillon. 2012. 'Efficacy of the LiSN & Learn auditory training software: randomized blinded controlled study', *Audiol Res*, 2: e15.

Christensen, D. L., L. A. Schieve, O. Devine, and C. Drews-Botsch. 2014. 'Socioeconomic status, child enrichment factors, and cognitive performance among preschool-age children: results from the Follow-Up of Growth and Development Experiences study', *Res Dev Disabil*, 35: 1789-801.

Del Zoppo, C., L. Sanchez, and C. Lind. 2015. 'A long-term follow-up of children and adolescents referred for assessment of auditory processing disorder', *Int J Audiol*, 54: 368-75.

Dillon, H., and S. Cameron. 2021. 'Separating the Causes of Listening Difficulties in Children', *Ear Hear*, 42: 1097-108.

Fiscella, K., and D. R. Williams. 2004. 'Health disparities based on socioeconomic inequities: implications for urban health care', *Acad Med*, 79: 1139-47.

Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis* (John Wiley & Sons).

Harding, Jessica F, Pamela A Morris, and Diane Hughes. 2015. 'The relationship between maternal education and children's academic outcomes: A theoretical framework', *Journal of Marriage and Family*, 77: 60-76.

Hechtman, L., J. M. Swanson, M. H. Sibley, A. Stehli, E. B. Owens, J. T. Mitchell, L. E. Arnold, B. S. Molina, S. P. Hinshaw, P. S. Jensen, H. B. Abikoff, G. Perez Algorta, A. L. Howard, B. Hoza, J. Etcovitch, S. Houssais, K. D. Lakes, J. Q. Nichols, and M. T. A. Cooperative Group. 2016. 'Functional Adult Outcomes 16 Years After Childhood Diagnosis of Attention-Deficit/Hyperactivity Disorder: MTA Results', *J Am Acad Child Adolesc Psychiatry*, 55: 945-52 e2.

Hind, S. E., R. Haines-Bazrafshan, C. L. Benton, W. Brassington, B. Towle, and D. R. Moore. 2011. 'Prevalence of clinical referrals having hearing thresholds within normal limits', *Int J Audiol*, 50: 708-16.

Hjern, A., G. R. Weitoft, and F. Lindblad. 2010. 'Social adversity predicts ADHD-medication in school children--a national cohort study', *Acta Paediatr*, 99: 920-4.

Humanski, R. A., and R. A. Butler. 1988. 'The contribution of the near and far ear toward localization of sound in the sagittal plane', *J Acoust Soc Am*, 83: 2300-10.

Hunter, L. L., C. M. Blankenship, L. Lin, N. T. Sloat, A. Perdew, H. Stewart, and D. R. Moore. 2021. 'Peripheral Auditory Involvement in Childhood Listening Difficulty', *Ear Hear*, 42: 29-41.

Keith, Robert W. 2009. *SCAN-3 for children: Tests for auditory processing disorders* (Pearson).

Little, Roderick JA. 1988. 'A test of missing completely at random for multivariate data with missing values', *Journal of the American statistical Association*, 83: 1198-202.

Lohvansuu, K., M. Torppa, T. Ahonen, K. Eklund, J. A. Hamalainen, P. H. T. Leppanen, and H. Lyytinen. 2021. 'Unveiling the Mysteries of Dyslexia-Lessons Learned from the Prospective Jyvaskyla Longitudinal Study of Dyslexia', *Brain Sci*, 11.

Magnuson, Katherine A, Holly R Sexton, Pamela E Davis-Kean, and Aletha C Huston. 2009. 'Increases in maternal education and young children's language skills', *Merrill-Palmer Quarterly (1982-)*: 319-50.

Perrachione, T. K., S. N. Del Tufo, and J. D. Gabrieli. 2011. 'Human voice recognition depends on language ability', *Science*, 333: 595.

Petley, L., L. L. Hunter, L. Motlagh Zadeh, H. J. Stewart, N. T. Sloat, A. Perdew, L. Lin, and D. R. Moore. 2021. 'Listening Difficulties in Children With Normal Audiograms: Relation to Hearing and Cognition', *Ear Hear*.

Pimperton, H., and C. R. Kennedy. 2012. 'The impact of early identification of permanent childhood hearing impairment on speech and language outcomes', *Arch Dis Child*, 97: 648-53.

Pungello, E. P., I. U. Iruka, A. M. Dotterer, R. Mills-Koonce, and J. S. Reznick. 2009. 'The effects of socioeconomic status, race, and parenting on language development in early childhood', *Dev Psychol*, 45: 544-57.

Quinto, A., S. Abu El Adas, and S. V. Levi. 2020. 'Re-Examining the Effect of Top-Down Linguistic Information on Speaker-Voice Discrimination', *Cogn Sci*, 44: e12902.

Reiss, F. 2013. 'Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review', *Soc Sci Med*, 90: 24-31.

Roy, A., L. Hechtman, L. E. Arnold, J. M. Swanson, B. S. G. Molina, M. H. Sibley, A. L. Howard, and M. T. A. Cooperative Group. 2017. 'Childhood Predictors of Adult Functional Outcomes in the Multimodal Treatment Study of Attention-Deficit/Hyperactivity Disorder (MTA)', *J Am Acad Child Adolesc Psychiatry*, 56: 687-95 e7.

Rutter, Michael, A Bailey, and Catherine Lord. 2003. 'SCQ', *The Social Communication Questionnaire. Torrance, CA: Western Psychological Services*.

Shaywitz, Sally E., Jack M. Fletcher, John M. Holahan, Abigail E. Shneider, Karen E. Marchione, Karla K. Stuebing, David J. Francis, Kenneth R. Pugh, and Bennett A. Shaywitz. 1999. 'Persistence of Dyslexia: The Connecticut Longitudinal Study at Adolescence', *Pediatrics*, 104: 1351-59.

Sherar, L. B., T. P. Griffin, U. Ekelund, A. R. Cooper, D. W. Esliger, E. M. van Sluijs, L. Bo Andersen, G. Cardon, R. Davey, K. Froberg, P. C. Hallal, K. F. Janz, K. Kordas, S. Kriemler, R. R. Pate, J. J. Puder, L. B. Sardinha, A. F. Timperio, and A. S. Page. 2016. 'Association between maternal education and objectively measured physical activity and sedentary time in adolescents', *J Epidemiol Community Health*, 70: 541-8.

Skoe, E., J. Krizman, and N. Kraus. 2013. 'The impoverished brain: disparities in maternal education affect the neural response to sound', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33: 17221-31.

Smoski, W, Michael A Brunt, and J Curtis Tannahill. 1998. 'Children's auditory performance scale (CHAPS)', *Tampa, FL: Educational Audiology Association*.

St Clair, M. C., A. Pickles, K. Durkin, and G. Conti-Ramsden. 2011. 'A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI)', *J Commun Disord*, 44: 186-99.

Stewart, H. J., E. K. Cash, L. L. Hunter, T. Maloney, J. Vannest, and D. R. Moore. 2022. 'Speech cortical activation and connectivity in typically developing children and those with listening difficulties', *Neuroimage Clin*, 36: 103172.

Sullivan, E. V., T. Brumback, S. F. Tapert, D. Prouty, R. Fama, W. K. Thompson, S. A. Brown, K. Cummins, I. M. Colrain, F. C. Baker, D. B. Clark, T. Chung, M. D. De Bellis, S. R. Hooper, B. J. Nagel, B. N. Nichols, W. Chu, D. Kwon, K. M. Pohl, and A. Pfefferbaum. 2017. 'Effects of prior testing lasting a full year in NCANDA adolescents: Contributions from age, sex, socioeconomic status, ethnicity, site, family history of alcohol or drug abuse, and baseline performance', *Dev Cogn Neurosci*, 24: 72-83.

Taylor, B. K., M. R. Frenzel, J. A. Eastman, A. I. Wiesman, Y. P. Wang, V. D. Calhoun, J. M. Stephen, and T. W. Wilson. 2022. 'Reliability of the NIH toolbox cognitive battery in children and adolescents: a 3-year longitudinal examination', *Psychol Med*, 52: 1718-27.

Verbeke, Geert. 1997. 'Linear mixed models for longitudinal data.' in, *Linear mixed models in practice* (Springer).

Weintraub, S., P. J. Bauer, P. D. Zelazo, K. Wallner-Allen, S. S. Dikmen, R. K. Heaton, D. S. Tulsky, J. Slotkin, D. L. Blitz, N. E. Carlozzi, R. J. Havlik, J. L. Beaumont, D. Mungas, J. J. Manly, B. G. Borosh, C. J. Nowinski, and R. C. Gershon. 2013a. 'I. NIH Toolbox Cognition Battery (CB): introduction and pediatric data', *Monogr Soc Res Child Dev*, 78: 1-15.

Weintraub, S., S. S. Dikmen, R. K. Heaton, D. S. Tulsky, P. D. Zelazo, P. J. Bauer, N. E. Carlozzi, J. Slotkin, D. Blitz, K. Wallner-Allen, N. A. Fox, J. L. Beaumont, D. Mungas, C. J. Nowinski, J. Richler, J. A. Deocampo, J. E. Anderson, J. J. Manly, B. Borosh, R.

Havlik, K. Conway, E. Edwards, L. Freund, J. W. King, C. Moy, E. Witt, and R. C. Gershon. 2013b. 'Cognition assessment using the NIH Toolbox', *Neurology*, 80: S54-64.

Figure legends

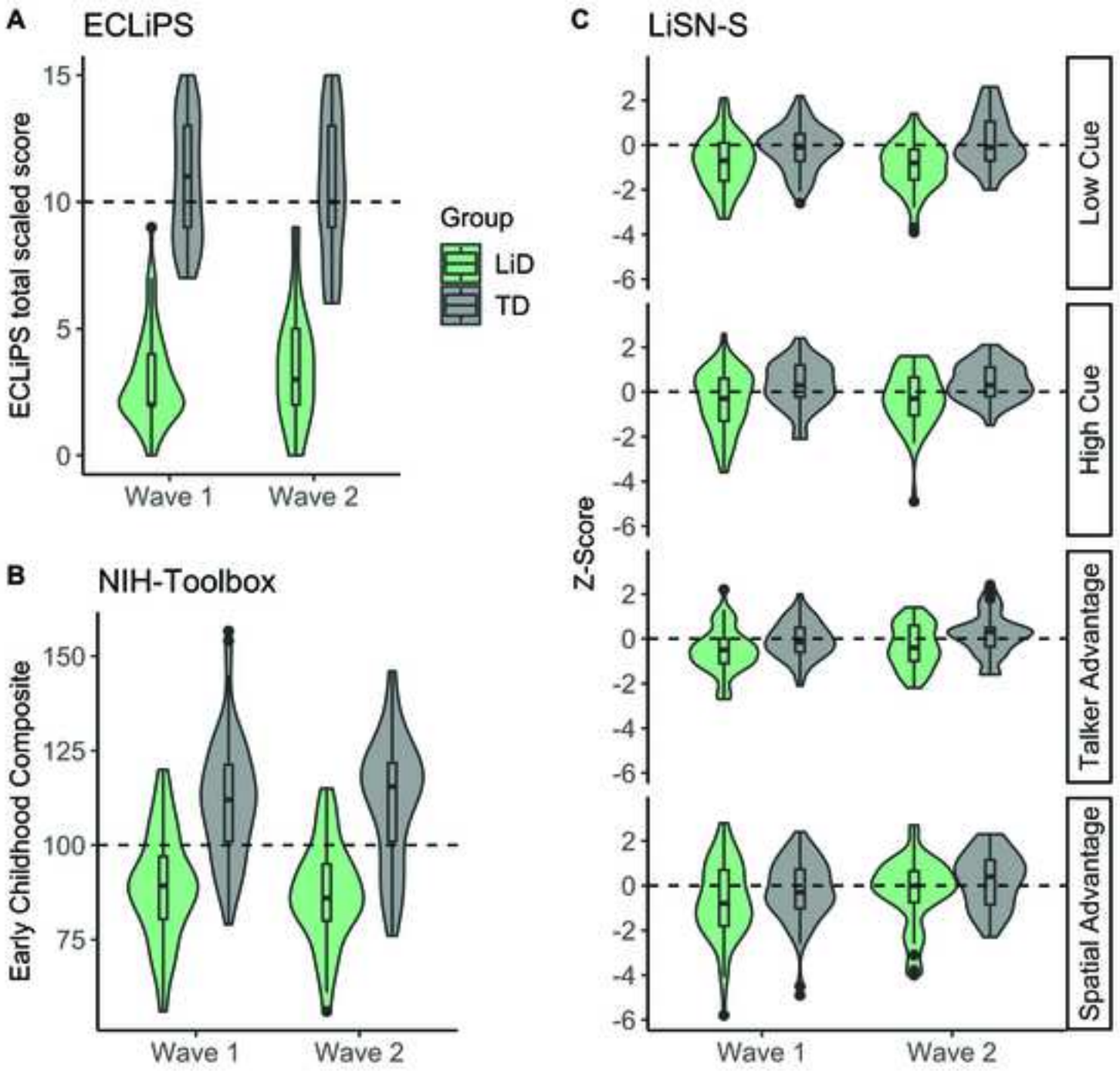
Figure 1. All test scores were consistently lower in the LiD group across Waves 1 and 2. A. Violin plots of ECLiPS Total Scaled scores showing the probability density of the data. Violin plots are overlaid with boxplots indicating each group's median and interquartile range in Waves 1 and 2. The horizontal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

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

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

Figure1

Click here to access/download;Figure;figure1._CMYKtif.tif



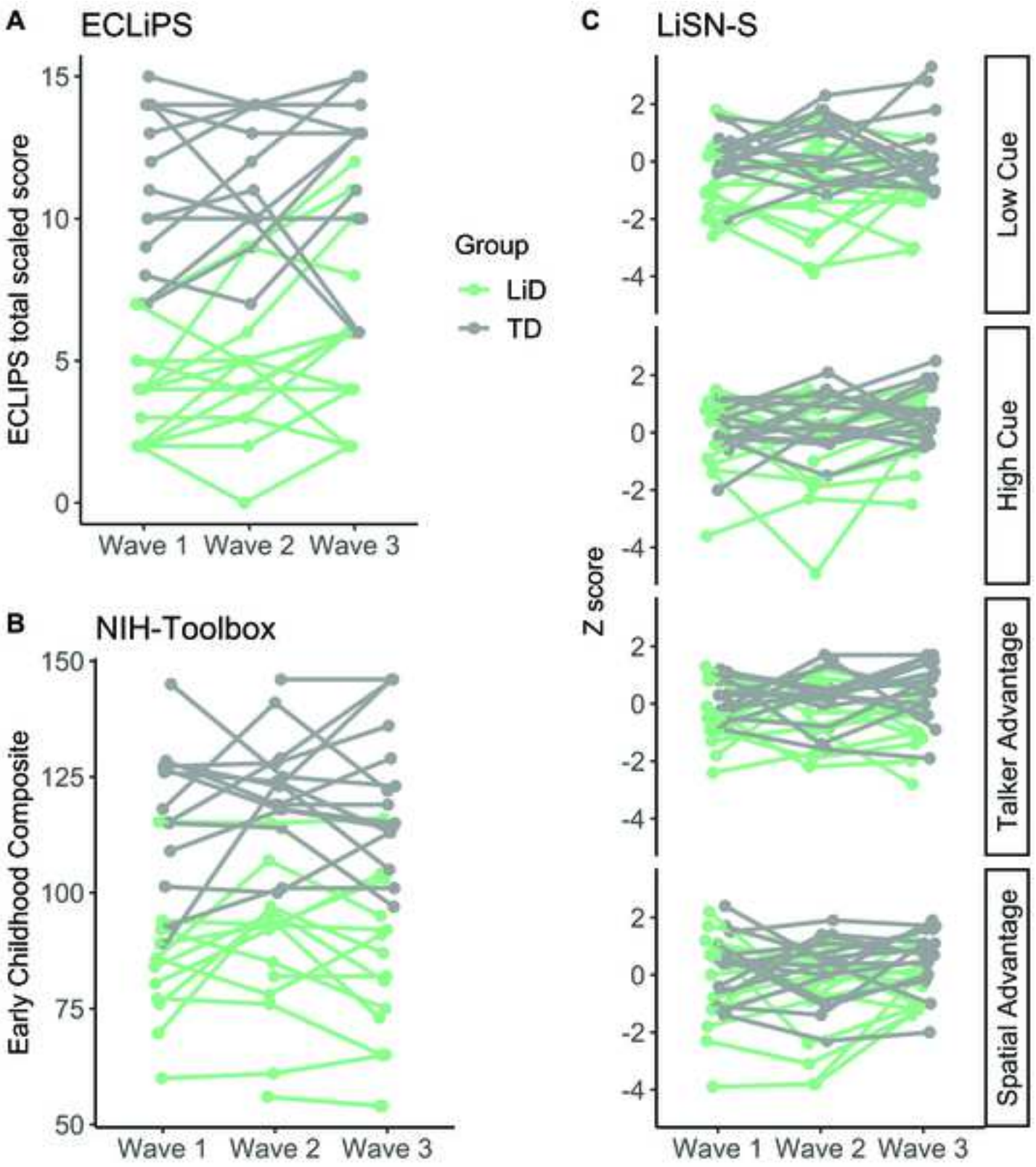

A
ECLiPS
ECLIPS total scaled score
15
10
5
0
Wave 1
Wave 2
Wave 3
B
NIH-Toolbox
Early Childhood Composite
150
125
100
75
50
Wave 1
Wave 2
Wave 3
Group
LiD
TD
C
LiSN-S
Z score
2
0
-2
-4
Low Cue
High Cue
Talker Advantage
Spatial Advantage
Wave 1
Wave 2
Wave 3

Figure3

Click here to access/download;Figure;figure3_CMYK.tif

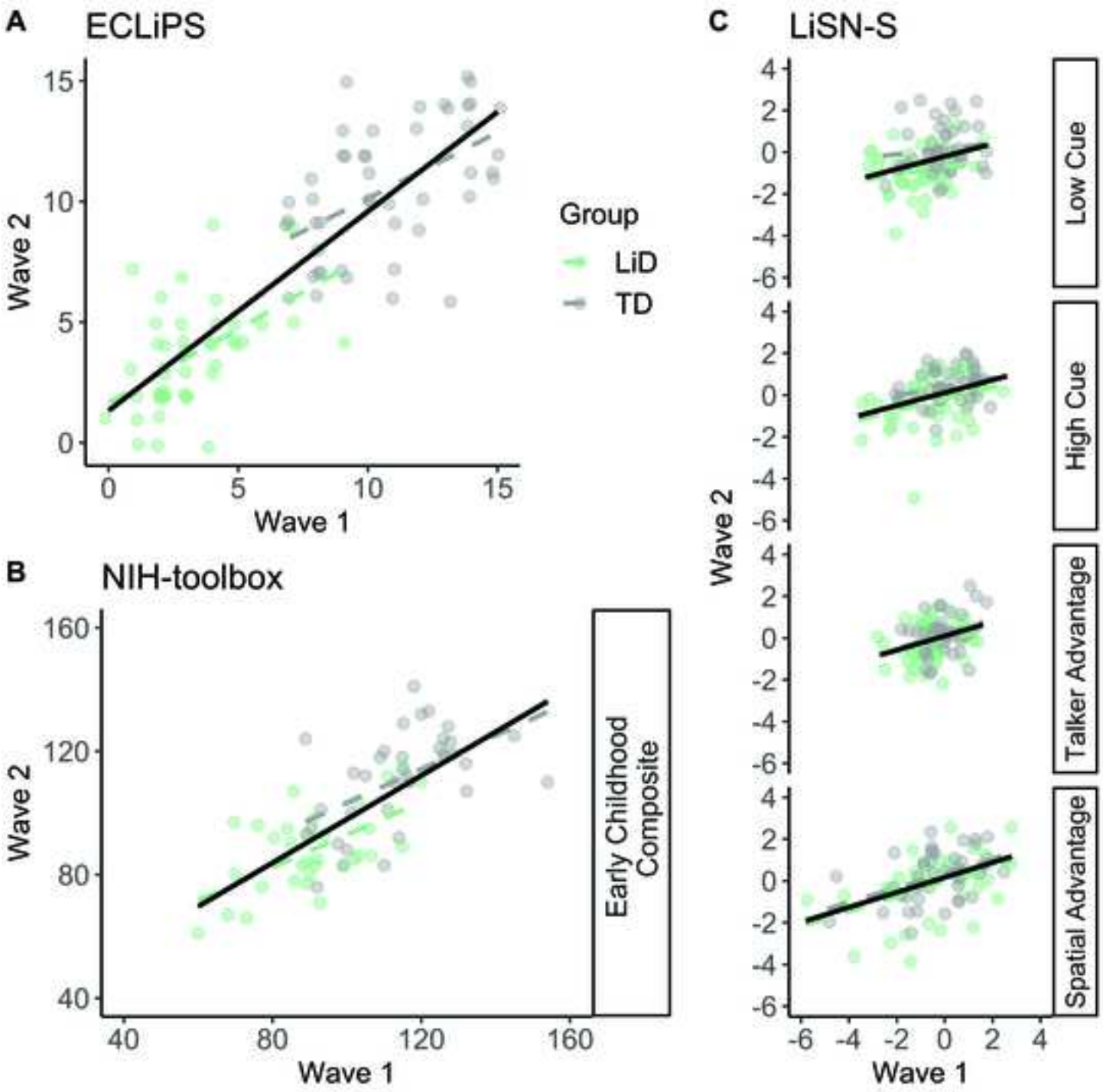


Table 1. Participant numbers and demographic information

	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)

Table 2. Auditory and listening test scores

	Wave 1		Wave 2		Wave 3	
Testing measures*	TD	LiD	TD	LiD	TD	LiD
ECLiPS scores	n = 79	n = 67	n = 47	n = 49	n = 13	n = 15
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, dB	7.43 (1.83)	7.25 (2.14)	7.79 (1.13)	7.64 (2.09)	8.28 (1.46)	8.76 (1.23)

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

Table 3. Cognitive test scores

	Wave 1		Wave 2		Wave 3	
Testing measures*	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)
Crystalized 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)

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

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

Table 4. Pearson correlation coefficients and estimated reliability coefficients between waves 1 and 2

	Pearson correlation coefficient	ICC (95%CI), Unadjusted	ICC (95%CI), Adjusted by Wave
ECLIPS total scaled score	0.85	0.86 (0.80, 0.90)	0.86 (0.80, 0.90)
LiSN-S, z score			
Low Cue	0.29	0.27 (0.13-0.49)	0.28 (0.13-0.49)
High Cue	0.35	0.32 (0.17-0.52)	0.33 (0.18-0.53)
Talker Advantage	0.32	0.29 (0.14-0.51)	0.33 (0.17-0.54)
Spatial Advantage	0.45	0.39 (0.24-0.56)	0.42 (0.27-0.58)
NIH-toolbox age-corrected standard score			
Picture vocabulary	0.66	0.65 (0.51-0.77)	0.65 (0.52-0.77)
Flanker test	0.51	0.42 (0.26-0.60)	0.48 (0.32-0.64)
List sorting working memory	0.66	0.68 (0.53-0.79)	0.68 (0.55-0.80)
Dimensional change card sort test	0.48	0.53 (0.35-0.69)	0.52 (0.35-0.69)
Pattern comparison processing speed test	0.58	0.56 (0.40-0.71)	0.56 (0.40-0.70)
Picture sequence memory test	0.44	0.45 (0.27-0.64)	0.45 (0.27-0.64)
Oral reading recognition test	0.70	0.68 (0.54-0.79)	0.70 (0.57-0.80)
Fluid cognition composite	0.74	0.75 (0.64-0.84)	0.76 (0.64-0.84)
Crystallized cognition composite	0.80	0.79 (0.69-0.86)	0.79 (0.69-0.86)
Total composite	0.82	0.82 (0.73-0.88)	0.82 (0.74-0.88)
Early childhood composite	0.73	0.73 (0.61-0.82)	0.74 (0.63-0.83)