



Verbal and visual serial-order memory in deaf signers and hearing nonsigners: A systematic review and meta-analysis

Tyler C. McFayden¹ · Maria K. Gonzalez Aguiar² · Charlotte C. MacKenzie² · Anne McIntosh³ · Kristi S. Multhaup²

Accepted: 21 March 2023 / Published online: 3 April 2023
© The Psychonomic Society, Inc. 2023

Abstract

Previous research suggests Deaf signers may have different short-term and working memory processes compared with hearing nonsigners due to prolonged auditory deprivation. The direction and magnitude of these reported differences, however, are variable and dependent on memory modality (e.g., visual, verbal), stimulus type, and research design. These discrepancies have made consensus difficult to reach which, in turn, slows progress in areas such as education, medical decision-making, and cognitive sciences. The present systematic review and meta-analysis included 35 studies ($N = 1,701$ participants) that examined verbal ($n = 15$), visuospatial ($n = 10$), or both verbal and visuospatial ($n = 10$) serial-memory tasks comparing nonimplanted, Deaf signers to hearing nonsigners across the life span. Multivariate meta-analyses indicated a significant, negative effect of deafness on verbal short-term memory (forward recall), $g = -1.33$, $SE = 0.17$, $p < .001$, 95% CI $[-1.68, -0.98]$, and working memory (backward recall), $g = -0.66$, $SE = 0.11$, $p < .001$, 95% CI $[-0.89, -0.45]$, but no significant effect of deafness on visuospatial short-term memory, $g = -0.055$, $SE = 0.17$, $p = 0.75$, 95% CI $[-0.39, 0.28]$. Visuospatial working memory was not analyzed due to limited power. Population estimates for verbal and visuospatial short-term memory were moderated by age wherein studies with adults demonstrated a stronger hearing advantage than studies with children/adolescents. Quality estimates indicated most studies were of fair quality, with only 38% of studies involving Deaf authors. Findings are discussed in the context of both Deaf equity and models of serial memory.

Keywords Deaf signers · Verbal span · Visuospatial span · Short-term memory · Working memory

The ability to briefly maintain information in memory is related to higher cognitive performance in wide-ranging

areas including spatial navigation, math, reading, decision-making, and language (Baddeley, 1996). In cognitive theories, this ability has been termed *short-term memory* or *working memory*, although the definitions of these terms have varied widely (see Cowan, 2017, for a review). The present meta-analysis distinguishes between tasks that require the brief (15–30 s) maintenance of a limited amount of information, *short-term memory* (Atkinson & Shiffrin, 1968), and tasks that require the processing and manipulation of limited amounts of information, *working memory* (Baddeley & Hitch, 1974). More specifically, the present meta-analysis examines where group differences across nonimplanted, Deaf signers and hearing nonsigners do and do not occur.

The topic of deafness and memory has been studied since the mid-1900s (e.g., Blair, 1957). Despite over half a century of study, the literature has presented mixed results regarding whether, compared with hearing individuals, Deaf individuals demonstrate memory deficits (e.g., Hirshorn et al., 2012), strengths (e.g., Cardin et al., 2018),

Public significance statement

Short-term memory tests are part of standardized tests of intelligence. This meta-analytic systemic review reveals that hearing nonsigners have higher scores on verbal short-term memory tests than Deaf nonsigners, with moderate to large effect sizes. There was not a group difference on visuospatial short-term memory. These results suggest that visual span may be a more equitable method of assessing short-term memory.

✉ Tyler C. McFayden
tyler_mcfayden@med.unc.edu

¹ Carolina Institute for Developmental Disabilities, University of North Carolina Chapel Hill, 101 Renee Lynne Court, Carrboro, NC 27510, USA

² Davidson College, Department of Psychology, Davidson, NC, USA

³ Department of Arts and Humanities, University of Maryland, Global Campus, Adelphi, MD, USA

or no differences (e.g., Marshall et al., 2015). This lack of consensus can be attributed to numerous factors, including differences in stimuli (e.g., faces versus words; Bettger et al., 1997, and Geraci et al., 2008, respectively), method (e.g., recall the last seen item versus recall serial span; Hirshorn et al., 2012, and Boutla et al., 2004, respectively), and participant characteristics (e.g., hearing aids, cochlear implantation, age; cf. Conway et al., 2009, Kronenberger et al., 2018, and Arfé et al., 2015, respectively). This variability of findings limits the field's ability to assess models of memory and advance practical recommendations (e.g., best practices in education). Furthermore, the conflicting findings have contributed to and perpetuated misconceptions about Deaf signers, including the notion of Deaf signers having poor working memory (cf. Hamilton, 2011), or Deaf signers are “visual learners” (Marschark et al., 2017). Accordingly, the current meta-analysis begins to untangle inconsistent findings by focusing on a well-defined set of studies. The present research seeks to establish the direction and magnitude of possible memory differences between nonimplanted, Deaf signers, and hearing nonsigners for two distinct types of short-term and working memory: verbal-serial memory and visuospatial-serial memory.

Moreover, the current systematic review and meta-analysis focuses on prelingually Deaf users of a sign-based language (e.g., British Sign Language, American Sign Language) who do not use any auditory assistance (e.g., hearing aids, cochlear implants, bone-anchored hearing aids). A related literature examines short-term and working memory processes after receiving a cochlear implant to evaluate the role of *changes* to sound awareness on memory (e.g., Conway et al., 2009; Kronenberger et al., 2018). By contrast, the current systematic review and meta-analysis aims to evaluate how using sign language impacts short-term and working memory processes compared with using spoken language. Critically, evaluating individuals with residual hearing (e.g., use hearing aids for amplification), who were postlingually Deaf, or who used other forms of technology to access sound awareness (i.e., cochlear implants) would not address this question because such participants received at least some auditory input. By focusing only on nonimplanted, Deaf signers in contrast to hearing nonsigners, we aim to systematically evaluate the role of deafness on memory for items in serial order, both in verbal and visual tasks.

Verbal-serial-order memory

Verbal-serial-order tasks require participants to report back items exactly as presented; the items can be words, letters, sentences, or nameable stimuli such as digits. Digit span, for example, is used frequently in measures of intelligence

(e.g., Wechsler scales; Wechsler, 1997). The classic forward-digit-span task typically starts with a length-two series and after at least one correctly reported sequence (e.g., recall 4–9 as 4–9), the series length is increased by one. This process repeats until length nine is successfully completed or, alternatively, the participant can no longer report a correct sequence. The backward digit span task is identical, except participants recall digits in the reverse order (e.g., recall 4–9 as 9–4). In terms of the distinctions guiding this review, forward-digit span taps short-term memory, as it requires maintenance of information, whereas backward-digit span taps working memory, as it requires maintenance *and* manipulation of information (Baddeley & Hitch, 1974). This distinction between short-term and working memory is relative, as opposed to absolute, in the cognitive literature, because both tasks require similar levels of processing (Cowan, 2017). However, for the current review, these two domains are discussed separately to highlight the difference in task instructions and demands.

When considering verbal short-term memory (i.e., forward serial recall), nonimplanted Deaf signers of American Sign Language (ASL) often do not perform as well as hearing, nonsigning counterparts, including on tasks with printed letters (e.g., Bavelier et al., 2008), ASL letters (e.g., Boutla et al., 2004), ASL signs (Krakow & Hanson, 1985), digits (e.g., Koo et al., 2008), words (e.g., Geraci et al., 2008), and sentences (e.g., Streff et al., 1978). This pattern has been found both with Deaf children (e.g., Tomlinson-Keasey & Smith-Winberry, 1990) and adults (e.g., Bavelier et al., 2008). Effect sizes for group differences range from small (e.g., Andin et al., 2013) to moderate (e.g., Koo et al., 2008), suggesting unidentified sources of variance driving these differences in effect magnitudes. Theorists attribute these patterns to Deaf signers' lack of auditory experience, given verbal sequencing is a demand inherent to spoken language that is not placed in the language modality of ASL (Conway et al., 2009). In other words, Deaf signers have less experience with serial ordering than hearing nonsigners and the impact of this experience difference manifests in group differences on verbal short-term memory tasks.

Although previous studies seem to reach consensus on a Deaf deficit for forward-verbal-serial recall (e.g., Boutla et al., 2004), the results are not as clear when considering backward-verbal-serial recall. Some earlier studies suggest the backward-serial spans of Deaf signers are equivalent to those of hearing nonsigners (e.g., Wilson et al., 1997), predominantly due to hearing nonsigners demonstrating a decrease in backward span compared with forward span, whereas Deaf signers maintain their performance on forward and backward span tasks. These findings have emerged when considering simple digit spans (e.g., Wilson et al., 1997) and complex spans, such as operation span (Andin et al., 2013). Moreover, the pattern holds when evaluating backward

span in children (e.g., Alamargot et al., 2007) and adults (Boutla et al., 2004). By contrast, some studies suggest Deaf signers' backward spans may be higher than those of hearing nonsigners (Hamilton, 2011; Powell & Hiatt, 1996). Thus, while the findings for verbal-backward spans (working memory) are less consistent than those for verbal-forward spans (short-term memory), the existing data does not reveal any Deaf deficit in working memory—a contrast to the consistently reported Deaf deficit in short-term memory outlined above.

Taken together, most research suggests a hearing advantage for forward verbal serial order recall (e.g., Bavelier et al., 2008), but there is no clear consensus about the magnitude of this difference, or whether developmental cascades may impact this effect size. Furthermore, there is not yet consensus on whether group differences appear on backward verbal serial order recall, making verbal serial-working memory a particularly unclear area in studies of Deaf memory.

Visuospatial-serial-order memory

Previous theorists have postulated that deprivation in one sense (e.g., hearing) could lead to advanced skills in another (e.g., vision), as in the case of Deaf signers (for discussion, see Bavelier et al., 2006). Indeed, some studies support the notion that Deaf signers demonstrate stronger visuospatial-serial recall skills than hearing nonsigners in memory for designs (Blair, 1957) and forward recall from the Corsi Block Tapping Test (e.g., Cardin et al., 2018; Heled & Ohayon, 2021; Hirshorn et al., 2012). The Corsi Block Tapping Test (Corsi, 1972; hereafter abbreviated as *Corsi*) is traditionally a three-dimensional task where the experimenter and participant sit opposite each other with a board of blocks on the table between them. The test administrator taps on the blocks in a particular sequence, which the participant replicates. Corsi has inspired numerous adaptations for administration, including virtual tasks on a computer or iPad (e.g., Alamargot et al., 2007; Logan et al., 1996; McFayden et al., 2023). As a visuospatial analog of verbal-span tasks, forward Corsi has been used as a measure of visuospatial short-term memory and backward Corsi has been used as a measure of visuospatial working memory (Corsi, 1972). Of note, some theorists have challenged the use of forward and backward Corsi as analogues for forward and backward verbal recall, given the heavy working and processing load required in both conditions of the visuospatial task (Vandierendonck et al., 2004). For the purposes of the current review, forward and backward Corsi are considered separate to avoid intermixing different dependent measures of visual span.

Despite a research narrative of a sensory deprivation hypothesis (e.g., Hall & Bavelier, 2010), wherein nondeprived

senses strengthen due to a deprived sensory system, Marschark et al. (2016, 2017) urge researchers to question this generalization of a visuospatial advantage in Deaf signers. In addition to qualitatively evaluating whether Deaf signers may identify as visual learners (Marschark et al., 2013), Marschark et al. (2016, 2017) empirically demonstrate no significant group differences between Deaf signers and hearing nonsigners on forward-visual-serial tasks, which replicates prior reports (e.g., Flaherty & Moran, 2001; Marshall et al., 2015). Thus, whereas some research suggests a Deaf advantage in visual short-term memory (e.g., Hirshorn et al., 2012), recent work suggests Deaf signers' visuospatial short-term memory may be equivalent to that of hearing nonsigners (e.g., Marshall et al., 2015).

Fewer studies have considered backward-visuospatial-serial recall, although some studies conduct the Corsi Block Tapping test in forward and backward conditions (e.g., Heled & Ohayon, 2021). When evaluating backward-visuospatial-serial tasks, results again are mixed: Whereas some studies suggest group equivalency (e.g., Heled & Ohayon, 2021; Romero et al., 2014), others suggest audition to be an advantage (e.g., Marshall et al., 2015).

Ultimately, although a prominent narrative has emerged wherein Deaf signers have visual strengths (e.g., Hirshorn et al., 2012), when considering findings across multiple research teams (e.g., Flaherty & Moran, 2001; Marschark et al., 2016, 2017; Marshall et al., 2015) that narrative does not hold. The lack of consensus across studies warrants systematic evaluation.

Current study

Numerous studies have evaluated Deaf signers' memory compared with hearing nonsigners' memory and have reached variable conclusions, both about the direction (e.g., Hirshorn et al., 2012 vs. Boutla et al., 2004) and magnitude (e.g., Bavelier et al., 2006 vs. Boutla et al., 2004) of group differences. These discrepant findings not only have theoretical implications for understanding general frameworks related to memory but have equity implications when considering appropriate memory measurement for Deaf populations. Thus, the current study seeks to systematically evaluate verbal- and visual-serial-order performance of nonimplanted, Deaf signers compared with hearing nonsigners. In doing so, additional questions about Deaf memory that have not been addressed in individual studies will be investigated here, including differences as a function of task modality (verbal, visual) and whether age may moderate these effects. Thus, the current systematic review and meta-analysis has two primary aims, broken out into components: the first aim is to evaluate Deaf memory for *verbal-serial-recall*, which includes short-term memory (forward recall)

Table 1 Systematic review and meta-analysis inclusion and exclusion criteria

	Inclusion	Exclusion
Study Group (Deaf)	<ul style="list-style-type: none"> • Humans • Severe to profound hearing loss • Age of onset of severe to profound hearing loss prior to 3 years of age • Use of a sign-based language as form of communication 	<ul style="list-style-type: none"> • Animals • Diagnoses of comorbid neurocognitive or developmental delays • $NVIQ \leq 2SD$ from the normative mean • Use of cochlear implants or hearing aids • Mild to moderate hearing loss or age-related hearing loss
Comparison Group (Hearing)	<ul style="list-style-type: none"> • Humans • Normal hearing as reported by pure-tone average thresholds of 500 Hz, 1 kHz, 2 kHz and 4 kHz being >25 dB or self-reported normal hearing 	<ul style="list-style-type: none"> • Animals • Diagnoses of comorbid neurocognitive or developmental delays • $NVIQ \leq 2SD$ from the normative mean
Task	ANY of the following <ul style="list-style-type: none"> • Verbal short-term memory task • Verbal working-memory task • Visuospatial short-term memory task • Visuospatial working-memory task • Span Task must be variable and dependent on participant's performance 	None of the following <ul style="list-style-type: none"> • Verbal short-term memory task • Verbal working-memory task • Visuospatial short-term memory task • Visuospatial working-memory task • Span task was set or fixed, not dependent on participant performance
Types of Studies	ANY of the following: <ul style="list-style-type: none"> • Cross-sectional studies • Between-groups design • Longitudinal studies 	<ul style="list-style-type: none"> • Theoretical or review papers • Case studies • Within-group designs with hearing only or deaf only (no comparison group)

$NVIQ$ = Nonverbal Intelligence Quotient; SD = standard deviation, Hz = Hertz, kHz = kilohertz

and working memory (backward recall); the second aim is to evaluate deaf memory for *visuospatial-serial-recall*, which includes short-term memory (forward recall) and working memory (backward recall).

Method

Transparency and openness

We adhered to the PRISMA 2020 guidelines for systematic reviews (Page et al., 2021). All data, analysis code, and research materials (including our coding scheme) are available online (https://osf.io/vwns8/?view_only=26fd6130332d4cc0a3281a720639352e). Data were modeled using RStudio Version 2021.09.0 with the metafor Version 3.4-0 and metaSEM Version 1.2.5.1 (Cheung, 2019; Viechtbauer & Cheung, 2010). This review project was preregistered on PROSPERO (https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=167987).

Eligibility criteria

The inclusion and exclusion criteria were specified in advance and documented in the preregistration on PROSPERO. The current meta-analysis compares verbal and visuospatial spans of nonimplanted Deaf signers and hearing nonsigners across the life span. Thus, study exclusion

criteria included any form of hearing assistance (e.g., cochlear implantation or hearing aids), any task where memory span was not directly measured using a serial-recall task, or any study where only Deaf participants were assessed without a hearing group. Detailed study inclusion and exclusion criteria are in Table 1.

Search strategy

The search strategy was developed by the first and final authors and was verified by a research librarian to ensure accuracy and inclusivity of retrieved articles. Electronic searches for publications in English were conducted in PsycINFO and PubMed. The query string used was:

(Deaf* OR "hard of hearing" OR "hearing loss" OR "hearing disorder*") and ((corsi OR "visual short term memory" OR "visual short-term memory" OR "visuospatial short term memory" OR "visuospatial short-term memory" OR "visuospatial memory" OR "visual memory") OR ("serial recall" OR "order recall" OR "forward recall" OR "backward recall" OR "verbal working memory" OR "verbal short term memory" OR "verbal short-term memory" OR "span" OR Recall (Learning))).

To reduce bias, two independent searches were conducted to identify all relevant studies (authors T.C.M. and M.K.G.A.). Specifiers on the search included selection of

‘Peer Reviewed’ and ‘English’ language. Search results were exported into Covidence¹; duplicates were automatically removed based on DOI, author, and year of publication. Title and abstracts were hand-screened by authors T.C.M. and M.K.G.A. based on inclusion and exclusion criteria. Any conflicts were resolved by consensus and a novel reviewer (author K.S.M.). Based on the title/abstract review, full-text screens were conducted by authors T.C.M. and M.K.G.A., again with consensus and conflict resolution conducted by author K.S.M. This searching and screening process was conducted twice—once in June 2021 and again in September 2022; the second phase was conducted searching for articles published only 2021–2022 to capture any articles that may have been missed during the manuscript-writing phase. Relevant articles to be included in the study were hand-searched by author T.C.M. to identify any potentially eligible publications. In addition, the first author invited corresponding authors of primary publications to share additional research, published or unpublished, that met the inclusion criteria. Lastly, the first author disseminated email requests for grey or existing literature related to the current study via three cognitive LISTSERVs.

Coding

The first author developed a coding manual and data extraction process in consultation with the second and final author (available in OSF: https://osf.io/vwns8/?view_only=26fd6130332d4cc0a3281a720639352e). The coding procedure was piloted by the first and second authors using eight articles, after which the coding manual and data extraction forms were revised accordingly. Data extracted from articles included study characteristics, participant demographics, methods, and outcome measures. The first, second, and third authors coded all articles using the refined coding manual and data extraction form. If selected articles had missing data, or required clarification on study details, T.C.M. contacted corresponding authors up to three times. Data presented in visual formats only (e.g., figures) were extracted using Web Plot Digitizer (Rohatgi, 2021). Data coding documents were compared using 4TOPS software² and CloudyExcel³; interrater reliability for continuous dependent variables was assessed using intraclass correlation. Inconsistencies between the three coders were discussed by all five authors and resolved by consensus.

¹ <https://support.covidence.org/help/how-can-i-cite-covidence>

² 4TOPS Compare Spreadsheets: http://www.4tops.com/compare_excel_files.htm

³ CloudyExcel Compare: <https://www.cloudyexcel.com/compare-excel/>

Assessment of study quality

Three domains of study quality were coded: methodological quality, comprehensiveness of reporting, and hearing bias. Comprehensiveness of reporting and hearing bias measures were developed with consultation from author A.M., a member of the Deaf community.

Methodological quality was assessed using an adapted version of National Institute of Health’s (NIH) Quality Assessment Tool for Case-Control Studies (NIH, 2021). Of the 12 criteria proposed by the NIH, five were changed or adjusted to reflect the use of a Deaf sample; one was omitted because it was encapsulated by the inclusion criteria for the review (i.e., cases and controls differentiated; full criteria in Table 2). Studies were assigned values of 0, 1, or “CD” for cannot determine. In accordance with Tawfit et al. (2019), a score 0–3 was considered poor quality, 4–7 as fair, and 8–11 as good.

For comprehensiveness of reporting, studies received up to one point for reporting each of five demographic characteristics of included samples: age (assigned a half point if only one statistic was reported; full credit for reporting two statistics such as mean and standard deviation, which was only relevant for age), sex, deafness onset, degree of hearing loss, and years of education. All domains were assessed independently by the first three authors with discrepancies resolved by consensus with all five authors. After one full round of coding assessment quality, an additional coding training was conducted which included revisions of the operational definition of several codes. The first three authors then re-coded quality and comprehensiveness codes independently and all five authors met to reach consensus on any remaining discrepant codes.

Lastly, the coders evaluated whether a member of the authorship team identified as Deaf or Hard of Hearing (scored as 0, 1) to evaluate hearing bias in publication. Deaf membership was determined by web searches and online profile data.

Data analysis

Power

Power calculations for a random effects model were used to determine the minimum number of studies needed to provide sufficient power. Prior research indicated that deafness had a small-to-moderate, negative effect on verbal memory, and a small, positive effect on visual memory. Power calculations indicated with an average group size of 23 participants, 12 studies would be needed for a moderate effect size (~.50) to reach sufficient power (~.80 in a high heterogeneity model), and 20 studies would be needed for a small effect size (~.20) to reach sufficient power (Valentine et al., 2010).

Table 2 Adapted quality assessment criteria from NIH's Quality Assessment Tool for Case-Controls

NIH Quality Assessment of Case-Control Studies	Adapted Quality Assessment Criteria
1. Was the research question or objective in this paper clearly stated and appropriate?	1. Was the research question or objective in this paper clearly stated and appropriate?
2. Was the study population clearly specified and defined?	2. Was the study population clearly specified and defined?
3. Did the authors include a sample size justification?	3. Did the authors include a sample size justification?
4. Were controls selected or recruited from the same or similar population that gave rise to the cases (including the same timeframe)?	4. Were controls selected or recruited from the same or similar population that gave rise to the cases (including the same timeframe)?
**5. Were the definitions, inclusion and exclusion criteria, algorithms or processes used to identify or select cases and controls valid, reliable, and implemented consistently across all study participants?	**5. Were inclusion and exclusion criteria explicitly stated for the deaf group?
**6. Were the cases clearly defined and differentiated from controls?	**Question 6 removed; study inclusion criteria required groups be differentiated based on hearing status
**7. If less than 100 percent of eligible cases and/or controls were selected for the study, were the cases and/or controls randomly selected from those eligible?	**7. Were inclusion and exclusion criteria explicitly stated for the hearing group?
8. Was there use of concurrent controls?	8. Was there use of concurrent controls?
**9. Were the investigators able to confirm that the exposure/risk occurred prior to the development of the condition or event that defined a participant as a case?	**9. Were participants tested in their dominant language?
**10. Were the dependent measures (span measures) derived from previously-established, standardized measures?	**10. Were the reliability and validity of the dependent measures reported from prior work and/or assessed in the study?
**11. Were the assessors of exposure/risk blinded to the case or control status of participants?	**11. Were the data collectors blind to the study hypotheses?
12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?	12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?

** = items that were adapted for use in deaf/hearing research studies. Question 6 from the NIH Quality Assessment criteria was removed in the revised version, as hearing and deaf groups were differentiated based on inclusion/exclusion criteria for the systematic review

Effect sizes

Effect sizes were calculated to represent the difference in serial-span scores between nonimplanted, Deaf signers and hearing, nonsigning participants using the standardized mean difference (Hedges's g), which is less susceptible to upward bias compared with Cohen's d (Hedges, 1981). Effect sizes were coded so that positive values reflect superior memory performance for Deaf signers compared with hearing nonsigners; negative values reflect a Deaf disadvantage.

Meta-analyses

Given that multiple effect sizes were generated from interdependent data (e.g., both a forward and backward effect size reported in one study), a multivariate meta-analysis was conducted for each topic (e.g., visual short-term memory) that had sufficient studies in the literature. The use of a multivariate meta-analysis in the case of non-independent effect sizes results in more precise estimates (smaller confidence intervals) than univariate meta-analysis and is the most appropriate model to use when sampling covariances are unknown (Cheung, 2019), as is the case with

the current data. In the case of small sample size for one outcome (i.e., visuospatial working memory) a univariate meta-analysis with only one outcome was conducted (Harrer et al., 2021).

Given the included studies had diverse populations and heterogeneity was anticipated, a random effects estimation model was used to account for heterogeneity by assuming additional variance beyond the studies in the analysis and adjusting study weights according to the extent of variation, which facilitates generalizability of findings (Borenstein et al., 2009). Heterogeneity was evaluated using the Q statistic and I^2 statistic. A significant Q rejects the null hypothesis of homogeneity and indicates that the variability among effect sizes is greater than what is likely to have resulted from study-level variability alone (Lipsey & Wilson, 2001). The I^2 statistic describes percentage of total variation across studies that is due to heterogeneity rather than chance (Higgins & Green, 2009). With significant heterogeneity, exploratory moderator analyses were conducted to determine whether other study characteristics were systematically associated with primary outcomes. Exploratory moderator variables were chosen a priori and included participant age, stimuli (e.g., letters, digits, words), and publication year.

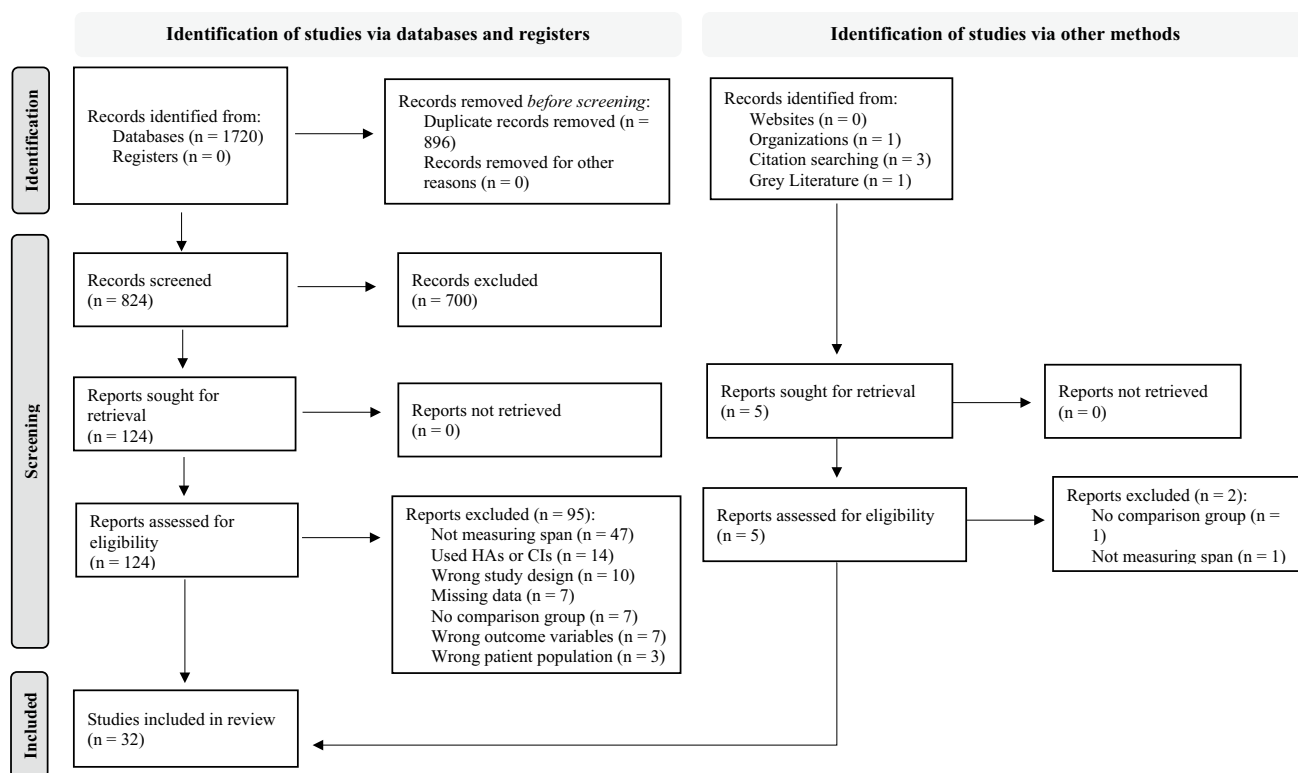


Fig. 1 PRISMA flowchart for included and excluded studies

Bias

Mean effects were assessed for degree of publication bias using a funnel plot of the effect sizes by their standard error and the trim-and-fill procedure (Duval & Tweedie, 2000) and by conducting Egger's linear regression method (Egger et al., 1997). To assess whether one or more studies had a substantial statistical impact on the summary effect, a Baujat plot was used to graphically assess studies that contributed considerably to the overall heterogeneity (Baujat et al., 2002) and studies were further evaluated using influence statistics (Viechtbauer & Cheung, 2010), which evaluate the presence of an outlier using eight statistical models. When outliers were detected, the analyses were run with and without the outlier to assess its effect on the overall findings.

Results

Study selection

Figure 1 shows the study selection results using the PRISMA flowchart (Page et al., 2021). Searches conducted using databases retrieved 1,720 records; of those records, 896 were removed as duplicates. With the 824 records remaining, 700 were excluded based on title and abstract review, which

left 124 records that entered the full-text review process. Authors were able to retrieve all 124 articles for full-text review; of these, 95 were further excluded due to a variety of reasons aligned with the inclusion and exclusion criteria (see Fig. 1 for further detail). After the full-text review process, 29 articles derived from database searches were extracted. From these 29 articles, two additional articles were hand-selected from the references to be included. One additional, unpublished article was identified to be included via grey literature searches, which resulted in a total of 32 articles to be coded and extracted.

The interrater agreement at each level of the decision-making process was acceptable (Norcini, 1999): 90% (title and abstract screening) and 75% (full text review). For data extraction, the coder interrater reliability was excellent, $r = .981$.

Study characteristics

Thirty-two articles, representing 37 studies were eligible for inclusion (see Table 3). Upon further inspection, two studies did not report any measure of variance (e.g., standard deviation, standard error) and thus could not be included in effect size generation, which rendered the final sample size to 35 studies, comprising 1,701 participants ($n = 816$ deaf). Studies were published in nine countries and spanned

Table 3 Characteristics of included studies

Article #	Study #	Authors and date	Study location	Deaf		Hearing		Verbal Span?	Visuospatial Span?
				<i>N</i>	<i>M</i> age	<i>N</i>	<i>M</i> age		
1	1	Alamargot et al. (2007)	France	15	14.02	15	13.55		Y
2	2	Andin et al. (2013)	Sweden	18	27.83	18	28.17	Y	
2	3	Andin et al. (2013)	Sweden	24	38.58	30	34.17	Y	
3	4	Arfé et al. (2015)	Italy	29	10.9	29	10.3	Y	
4	5	Bavelier et al. (2008)	USA	12	22	20	20	Y	
4	6	Bavelier et al. (2008)	USA	20	21	20	18.9	Y	
5	7	Blair (1957)	USA	53	NR	53	NR	Y	Y
6	8	Boutla et al. (2004)	USA	12	25	12	23	Y	
6	9	Boutla et al. (2004)^	USA	20	21	20	22	Y	
6	10	Boutla et al. (2004)	USA	36	18	18	20	Y	
7	11	Cardin et al. (2018)	UK	12	25.7	16	28.3		Y
8	12	Chincotta & Chincotta (1996)	China	15	NR	15	NR	Y	
9	13	Edwards et al. (2021)	USA	42	NR	38	NR	Y	
10	14	Emmorey et al. (2017)	USA	35	33.1	35	22.5	Y	Y
11	15	Farjardo et al. (2008)	Italy	30	16.8	31	16.9	Y	Y
12	16	Flaherty & Moran (2001)	Ireland	24	21	23	25	Y	
13	17	Flaherty & Moran (2004)	Ireland	20	23	20	24	Y	
13	18	Flaherty & Moran (2004)	Ireland	20	23	20	24	Y	
14	19	Geraci et al. (2008)	Italy	16	44.13	16	44.88	Y	Y
15	20	Gozzi et al. (2011)	Italy	12	44	12	46	Y	
16	21	Hall et al. (2018)	USA	45	8.17	45	8.33		Y
17	22	Harris & Moreno (2004)	UK	29	7.95	30	8.03		Y
18	23	Heled & Ohayon (2021)	Israel	20	32.55	20	32.7		Y
19	24	Hermelin & O'Connor (1975)	UK	10	12.17	10	10.33	Y	
20	25	Koo et al. (2008)	USA	13	23.1	9	30.4	Y	Y
21	26	Marschark et al. (2013)	USA	39	NR	32	NR		Y
22	27	Marschark et al. (2015)	USA	42	NR	45	NR		Y
23	28	Marshall et al. (2015)	UK	27	9.17	28	9		Y
24	29	McDaniel (1980)	USA	11	10	10	10		Y
25	30	McFayden et al. (2023)	USA	33	19.6	32	19.5	Y	Y
26	31	Olsson & Furth (1966)	USA	15	NR	15	NR	Y	Y
27	32	Romero Lauro et al. (2014)	Italy	18	39.33	18	36.5		Y
28	33	Tomlinson-Keasey & Smith-Winberry (1990)	USA	22	13.73	62	12.65	Y	Y
29	34	Tractenberg (2002)	USA	19	23.3	27	20.1	Y	Y
30	35	Tzeng (2002)	Taiwan	30	NR	30	NR	Y	
31	36	Wilson et al. (1997)	USA	16	9.17	31	9.3	Y	Y
32	37	Wilson & Emmorey (2006)^	USA	12	NR	16	NR	Y	

“NR” = Not Reported, “*M* age” = Mean age in years, *N* = sample size, “Y” indicates yes, this dependent measure was present in select study; ^ indicates studies excluded from effect size calculations due to no reported measures of variance

64 years (range: 1957–2021, *M* = 2004, *SD* = 16.07 yrs). Studies were predominantly conducted with adults (*n* = 22) and in English (*n* = 28), followed by Italian (*n* = 5). Ten studies contained both verbal and visuospatial span tasks.

When considering verbal and visual memory separately, 21 articles comprising 25 studies contained a verbal serial span task (all contained forward trials, *n* = 12 contained

backward trials) comprising 1,202 participants (*n* = 576 Deaf). Deaf participants were an average age of 23.47 years (*SD* = 9.95, range: 9.17–44.13) and hearing participants were an average of 23.01 years old (*SD* = 10.01, range: 9.30–46.00). The most common stimulus for verbal span was digits (*n* = 16), followed by letters (*n* = 5) and words (*n* = 4).

Comparatively, 20 articles comprising 20 studies contained a visuospatial span task (all contained forward trials, $n = 4$ contained backward trials), comprising 1,080 participants ($n = 510$ Deaf). Deaf participants were an average age of 20.61 years ($SD = 11.65$, range: 7.95–44.13) and hearing participants were an average of 20.17 years old ($SD = 11.42$, range: 8.03–44.80). The most common stimuli for the visuospatial span tasks were Corsi blocks ($n = 15$), others included pictures of objects ($n = 2$), lights from the game Simon, Knox cubes, or nonsense forms.

Quality of evidence

Interrater agreement for the quality, comprehensiveness, and deaf authorship codes was high (93%). Quality of evidence data is available in Table 4. Studies were overall of fair quality ($M = 4.5$, $SD = 1.58$, range: 0–11) with 25 studies being of fair quality as evidenced by scores on the adapted NIH criteria; 10 of the 37 studies were in the “poor” range and only two studies were in the “good” range. Zero studies included sample size justifications or information to suggest data collectors were blind to study hypotheses; only one study indicated the use of concurrent controls, suggesting these practices are not common in research with Deaf participants. Excluding these three codes (i.e., sample size justification, blind to hypotheses, concurrent controls) did not significantly change the average quality score ($M = 4.47$, $SD = 1.54$), but did alter the range distribution (zero studies “poor” [range 0–2], 30 studies “fair” [range 3–5], and seven studies “good” [range 6–8]).

For comprehensiveness of reporting, the average score was 3.3/5 metrics (range: 0–5) with the modal study reporting on 4/5 domains. The domain with the least reporting was educational level (46% of studies reported). Lastly, 43% of studies had a Deaf author included in their authorship team; however, as some articles contained multiple studies, this number decreased to 38% when considering article authorship teams (12/32).

Meta-analyses

Verbal serial order

Funnel plots for forward and backward verbal serial order are available in Supplementary Figures 1 and 2. Trim-and-fill analyses for publication bias (Duval & Tweedie, 2000) imputed zero studies to the right of the mean (positive Hedges’s g effect sizes) for both forward and backward recall. However, Egger’s test (Egger et al., 1997) revealed the funnel plot asymmetry for forward recall was significant, $z = -5.75$, $p < .001$, suggesting evidence of publication bias favoring larger, negative effect sizes (suggesting a hearing bias). Funnel plot asymmetry for backward recall was not

significant, $z = 1.21$, $p = .22$, suggesting a lack of publication bias. Next, inspecting the forest plot identified one statistical outlier in the forward recall condition (Gozzi et al., 2011; Supplementary Figure 3), which was confirmed with the Baujat plot (Baujat et al., 2002) and indicated as an outlier on 100% (8 out of 8) of the influence plots (Viechtbauer & Cheung, 2010). Supplemental Figure 3 shows the full forest plot with the outlier depicted in red; Fig. 2 shows the plot without the outlier. Due to strong evidence of a statistical outlier, the meta-analysis was conducted with and without Gozzi et al. (2011). Comparing the two meta-analyses suggested statistically different estimates, $\chi^2 = 20.16$, $p < .001$; to be conservative, the multivariate meta-analysis was conducted without Gozzi et al. (2011). No studies in the backward recall conditions were indicated as outliers (see Fig. 3).

Results of the multivariate meta-analysis suggested a significant effect of deafness on verbal-serial recall, both forward, $g = -1.33$, $SE = 0.17$, $p < .001$, 95% CI [-1.68, -0.98] (see Fig. 2), and backward, $g = -0.66$, $SE = 0.11$, $p < .001$, 95% CI [-0.89, -0.45] (see Fig. 3). Effect sizes for verbal short-term (forward) and working memory (backward) were significantly related, $r = 0.89$; however, effect sizes and estimates of variance were significantly different between forward and backward recall, $\chi^2(df = 1) = 17.89$, $p = .001$, suggesting the effects of deafness are not the same on forward and backward verbal recall. Heterogeneity indicated low within-study heterogeneity, I^2 forward = 0.86, I^2 backward = 0.49; and high across-study variation, $Q = 166.476$, $p < .001$, suggesting other factors may also be accounting for significant variability in verbal serial recall performance.

Due to the significant heterogeneity indicators, separate meta-regressions were conducted to evaluate the role of participant age, study stimuli (letters versus digits), and publication year on the population estimate for verbal short-term and working memory. Results for the meta-regression with age indicated a significant regression coefficient for age in verbal short-term memory, $B = 0.05$, $p = .004$, 95% CI [0.017, 0.09], but not working memory, $B = 0.007$, $p = .65$, 95% CI [-0.02, 0.03]. The statistically significant age moderation for verbal short-term memory indicated the effect of deafness was stronger in studies with higher age participants (adults compared with children/adolescents); no relation was detected for working memory. This pattern may be due to the fact that the R^2 value when only group was considered was higher for working memory than for short-term memory (the bump in R^2 by age did not have as much room to increase for backward span as it did for forward span). Incorporating age into the final model accounted for 34% of the variability in verbal short-term memory and 64% of variability in verbal working memory, $R^2 = 0.339$ and 0.636, respectively. Results for the other two meta-regressions indicated no

Table 4 Study scores for quality assessment, comprehensiveness of reporting, and Deaf authorship

Study	Quality Score	Range	Comprehensiveness Score	Deaf Author
Alamargot et al. (2007)	7.5	Good	3/5	0
Andin et al. (2013): Exp 1	3	Poor	4/5	1
Andin et al. (2013): Exp 2	3	Poor	3/5	1
Arfè et al. (2015)	4	Fair	4/5	0
Bavelier et al. (2008): Exp 1	5	Fair	0.5/5	1
Bavelier et al. (2008): Exp 3	4	Fair	1/2	1
Blair (1957)	3.5	Fair	1.5/5	0
Boutla et al. (2004): Exp 1	4.5	Fair	1/2	1
Boutla et al. (2004): Exp 2	5	Fair	1/2	1
Boutla et al. (2004): Exp 3	4	Fair	1/2	1
Cardin et al. (2018)	7	Fair	4/5	1
Chincotta & Chincotta (1996)	5	Fair	4.5/5	0
Edwards et al. (2021)	7	Fair	0.5/5	1
Emmorey et al. (2017)	2.5	Poor	4/5	0
Farjardo et al. (2008)	7	Fair	5/5	0
Flaherty & Moran (2001)	2	Poor	5/5	0
Flaherty & Moran (2004): Exp 1	3	Poor	4/5	0
Flaherty & Moran (2004): Exp 2	3	Poor	4/5	0
Geraci et al. (2008)	5.5	Fair	4/5	0
Gozzi et al. (2011)	5	Fair	4.5/5	0
Hall et al. (2018)	6	Fair	4/5	1
Harris & Moreno (2004)	4	Fair	2/5	1
Heled & Ohayon (2021)	5	Fair	4/5	0
Hermelin & O'Connor (1975)	2	Poor	3/5	0
Koo et al. (2008)	8	Good	5/5	1
Marschark et al. (2013)	4	Fair	2/5	1
Marschark et al. (2015)	5	Fair	3/5	1
Marshall et al. (2015)	5	Fair	4/5	1
McDaniel (1980)	3	Poor	0.5/5	0
McFayden et al. (2023)	4	Fair	5/5	1
Olsson & Furth (1966)	5	Fair	1.5/5	0
Romero Lauro et al. (2014)	5	Fair	4/5	0
Tomlinson-Keasey & Smith-Winberry (1990)	6.5	Fair	4.5/5	0
Tractenberg (2002)	4.5	Fair	3/5	0
Tzeng (2002)	2	Poor	3/5	0
Wilson et al. (1997)	5	Fair	2/5	0
Wilson & Emmorey (2006)	2	Poor	0/5	0
Summary Statistic (<i>M</i> , Mode, Proportion)	4.5	Fair	3.3/5	43%

M = Mean; Scores from 0–3 were considered Poor quality, 4–7 as Fair, and 8–11 as Good (Tawfit et al., 2019). Deaf authorship assigned a score of 1 if one of the manuscript authors identifies as Deaf

significant effect of stimuli (verbal short-term memory: $B = 0.31$, $p = .36$, 95% CI $[-0.36, 0.98]$; verbal working memory: $B = -0.008$, $p = .97$, 95% CI $[-0.42, 0.41]$) or publication year (verbal short-term memory: $B = 0.01$, $p = .30$, 95% CI $[-0.01, 0.03]$; verbal working memory: $B = -0.007$, $p = .72$, 95% CI $[-0.01, 0.007]$).

Visuospatial serial order

Due to the small sample size of studies reporting on visuospatial working memory ($n = 4$), there was insufficient power to include backward recall as an outcome in the analyses (est. power $[1 - \beta] = 0.14$ for moderate

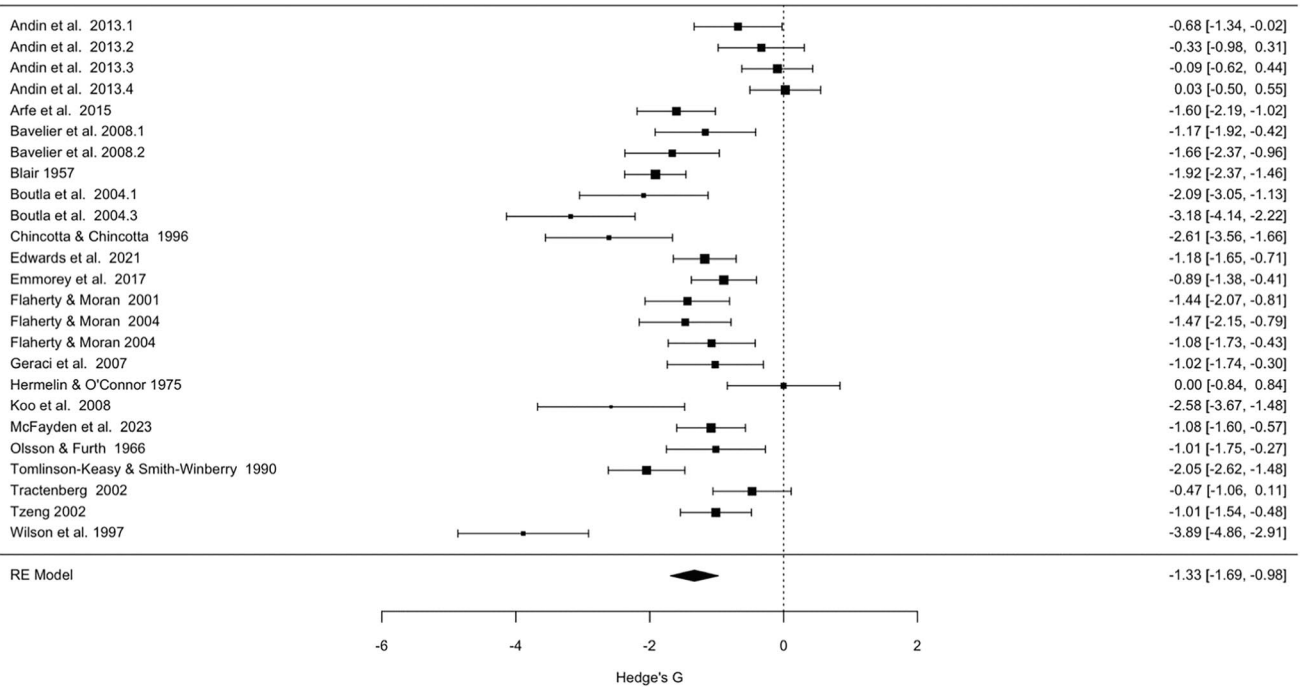


Fig. 2 Forest plots of population estimates and variance for forward verbal recall

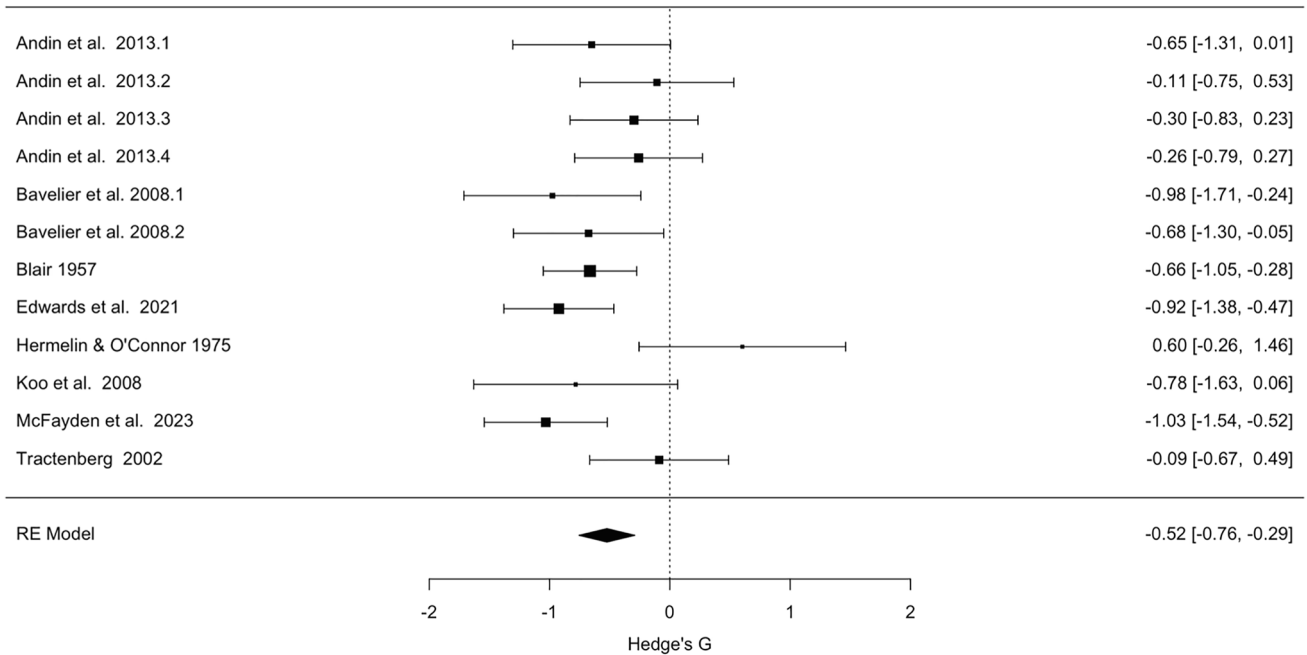


Fig. 3 Forest plots of population estimates and variance for backward verbal recall

heterogeneity given average sample size of 18.75 with proposed effect size of 0.2). Thus, a univariate meta-analysis was conducted with forward visual recall only. The

same assumptions and random effect models were used for the univariate analysis, as indicated in the Method section.

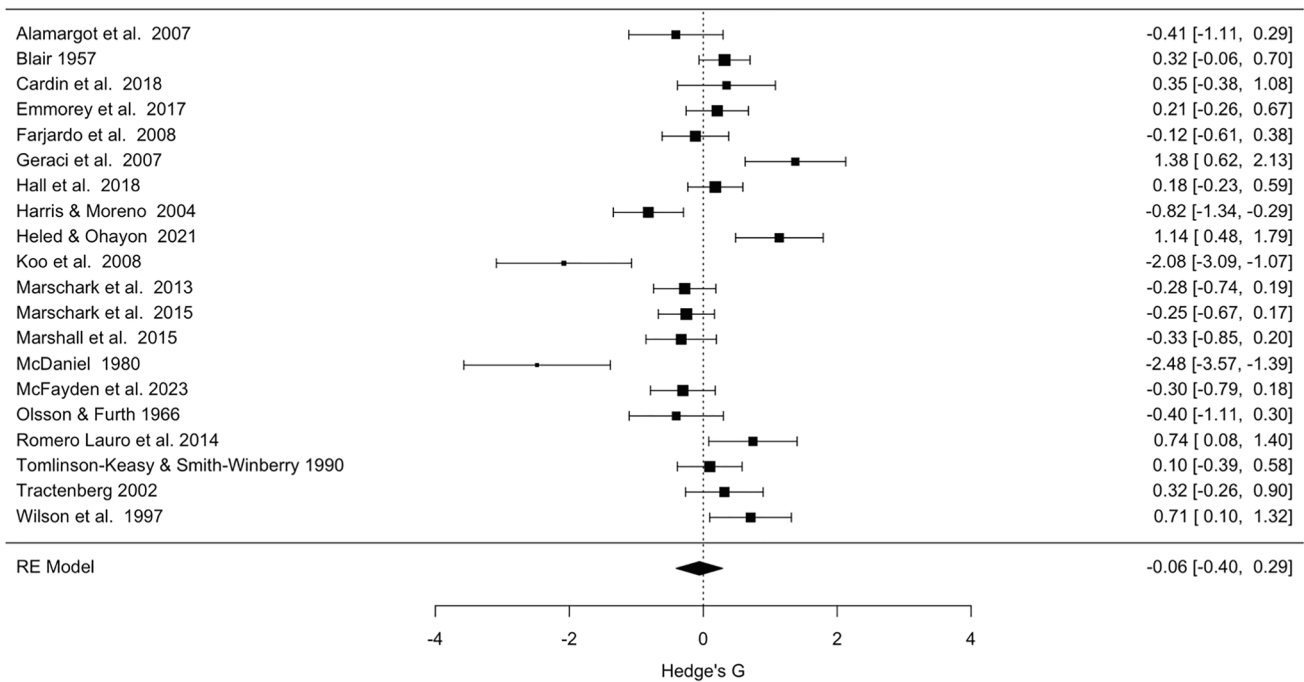


Fig. 4 Forest plots of population estimates and variance for forward visual recall

The trim-and-fill analysis for publication bias (Duval & Tweedie, 2000) imputed zero studies to the right of the mean; however, Egger's test (Egger et al., 1997) revealed the funnel plot asymmetry was significant, $z = -2.12$, $p = .03$, suggesting evidence of publication bias. Visual inspection of the funnel plot (Supplementary Figure 4) indicates a bias favoring larger, negative effect sizes (negative effect sizes indicate hearing participants scoring higher than Deaf participants). Next, visually inspecting the forest plot identified two potential statistical outliers (Koo et al., 2008; McDaniel, 1980), which were confirmed with the Baujat plot (Baujat et al., 2002). However, neither study reached statistical significance on any of the eight measures of statistical influence (Viechtbauer & Cheung, 2010). Thus, there did not appear to be a strong reason for exclusion, so the model was run including both studies.

Results of the univariate meta-analysis indicated no significant effect of deafness on visuospatial short-term memory, $g = -0.055$, $SE = 0.17$, $p = 0.75$, 95% CI [-0.39, 0.28] (see Fig. 4). Heterogeneity indicated low within-study heterogeneity, $I^2 = 0.86$, and moderate across study variation, $Q = 94.87$, $p < .001$, suggesting other factors may also be accounting for significant variability in visuospatial short-term memory.

Due to the significant heterogeneity Q value, a meta-regression was conducted to evaluate the role of participant age on the population estimate for visuospatial short-term memory. Results indicated a significant regression coefficient for age, $B = 0.041$, $p = .015$, 95% CI [-0.019, 0.86].

The statistically significant age moderation indicated the effect of deafness on forward visual recall was stronger in studies with participants of higher ages (adults compared with children/adolescents). Incorporating age into the final model accounted for approximately 13% of variability in visuospatial short-term memory, $R^2 = 0.126$.

A meta-regression was also conducted with publication year as a moderator. Results indicated no significant effect of publication year, $B = 0.0097$, 95% CI [-0.010, 0.019], $p = .394$. There was not sufficient variability in visuospatial stimuli to assess for moderation.

Discussion

The current systematic review and meta-analysis evaluated the impact of Deafness on serial-order memory, specifically visual- and verbal-serial-order short-term (forward recall) and working (backward recall) memory. Importantly, the current meta-analysis was only able to generate population estimates for verbal short-term, verbal working, and visual short-term memory. Due to currently limited sample size and insufficient power, visual working memory represents an area for future study.

Results of the multivariate meta-analyses for verbal memory indicated significant effects of Deafness on both short-term and working memory, wherein hearing nonsigners performed significantly better than Deaf signers. This finding held across stimuli and publication year. Despite the

correlation between forward and backward verbal-serial-order effect sizes indicating a strong relationship, forward and backward recall effect sizes were significantly different, suggesting that Deafness has a greater impact on short-term memory (forward recall) compared with working memory (backward recall) of verbal items. Additionally, verbal short-term memory was significantly moderated by age, insofar as studies with adult participants reported a larger hearing advantage than studies with child/adolescent participants. Interestingly, this age moderation was not significant for verbal working memory.

Results of the univariate meta-analysis for visual short-term memory indicated no significant effect of Deafness on visual forward recall, suggesting Deaf signers and hearing nonsigners have equivalent skills in forward visual serial recall tasks. Despite the lack of significant group differences, a significant moderation did emerge wherein studies with adult participants demonstrated a larger gap between Deaf signers and hearing nonsigners than studies with child/adolescent participants, which was also noted in verbal short-term memory. Although variance estimates suggest other factors may also moderate the relationship between Deafness and visual short-term memory, publication year was not significant, and there was not sufficient variability in stimulus type to probe for moderation as almost all studies (15/20) used a Corsi task.

The results of the visual meta-analysis provide a key takeaway message: Although previous literature has suggested that Deaf signers demonstrate weaker short-term and working memory for serial recall items (e.g., Bavelier et al., 2008; Conway et al., 2009; Kronenberger et al., 2018), the current visual results suggest Deaf signers are similarly skilled as hearing nonsigners at remembering visual items in a serial order when presented in a forward direction. Thus, the previous claim about Deaf signers having poorer serial order memory globally is *not* supported. When paired with the verbal results, discussed in greater detail below, we glean a better understanding of the relative strengths and weaknesses of Deaf signers studied in the current systematic review. The present results suggest that when considering the routine assessment of working memory in educational and cognitive contexts, the hearing bias of using verbal items, such as digits, may be disproportionately impacting Deaf signers and underestimating their abilities.

Indeed, the results of the verbal multivariate analyses supported the direction of the group findings reported in individual studies, wherein hearing nonsigners demonstrate a significant advantage in forward recall of serially presented items compared with Deaf signers. Previous explanations have attributed this finding to the nature of the language modalities of both groups, wherein spoken English relies more heavily on serial order compared with ASL (e.g., Cardin et al., 2018; Conway et al., 2009). Interestingly, the

hearing advantage replicates with backward recall of serially presented verbal items although the overall observed effect size was significantly smaller than for forward recall. These results suggest backward serial recall may not be as impacted in people who use sign-based languages compared with forward-recall. One area of future study could investigate whether a hearing advantage, including a larger one for forward than for backward recall, holds when using ASL words for the Deaf group and spoken words for the hearing group. Moreover, it will be important to assess whether the patterns replicate on sentence span tasks that use sentence structure observed in both native languages because these tasks may have higher external validity for daily exercises in working and short-term memory. Although the results of the multivariate meta-analysis suggested stimuli type did not moderate this relationship, only a few studies of the 25 analyzed studies included non-digits ($n = 5$ letters, $n = 4$ words), and thus perhaps was not powered enough to detect moderation.

The nonsignificant stimuli moderation for verbal forward and backward serial recall was unexpected. Many previous works (e.g., Boutla et al., 2004; Flaherty & Moran, 2001, 2004; Krakow & Hanson, 1985; Logan et al., 1996; Wilson et al., 1997) have used within-participants designs to compare spans of different verbal stimuli, such as digits, letters, and words. Although findings have been mixed, some studies (e.g., Boutla et al., 2004) demonstrated that there is no difference between Deaf signers and hearing nonsigners on forward verbal span when native language stimuli are used (i.e., read aloud digits for hearing nonsigners, signed numbers for Deaf signers). The lack of significant moderation of stimuli type in the current meta-analysis does not support these previous claims. However, perhaps multiple interactions are taking place that were not accounted for in the current review. Not only could stimuli type vary, but the presentation modality and recall modality are also domains that could vary between studies, ranging from printed stimuli/recall, verbal stimuli/recall, or a combination. Although there are too many factors to have considered for the current meta-analysis, it stands to reason that the multiple modalities of presentation, recall, instructions, and task could have cascading impacts on the results above and beyond the stimuli selected for each group.

Although the current results suggested no significant stimuli moderation, there was a significant effect of age on both forward verbal and forward visual serial recall. These results suggested that the effect size increased in magnitude along with participant age, wherein the gap between hearing and Deaf participants grew larger with increasing age. In other words, for children and adolescents, Deaf signers are at a small disadvantage in forward serial-order recall skills compared with hearing nonsigners, but by adulthood, differences of larger magnitude emerge. These findings also

support what we term the language-use hypothesis, wherein with more experience using a language modality (e.g., ASL or spoken English), our memory systems narrow towards that specific language modality. Thus, as we age, we may become less flexible with memory in unfamiliar language modalities. This process mirrors what we know about language experience in early development, wherein perception in infancy changes from language-general to language-specific, otherwise called perceptual narrowing, which is observed in both speech and sign (e.g., Maurer & Werker, 2014). Perhaps there is a similar process of perceptual narrowing related to one's native language modality that continues with age beyond the infancy period. Indeed, as one's brain adaptively and developmentally prunes irrelevant connections over time to increase efficiency, these age-related findings are consistent with previous literature suggesting perceptual narrowing which makes adults less flexible in their memory compared with younger children (Maurer & Werker, 2014). Future studies may also consider how cognitive aging may impact memory for serial recall in Deaf signers compared with hearing nonsigners, as none of the current studies included older adults.

Whereas the size of the Deaf–hearing difference on both verbal and visual forward recall was modified by age, there was no age effect for verbal backward recall. In other words, the size of the Deafness effect increased with age for short-term memory but not for working memory. This finding is counter to the pattern that would be expected based on the adult cognitive aging literature where age-related differences on short-term memory tasks can be explained by age-related slowing (Multhaup et al., 1996) and these are small as compared with larger age-related differences in working memory (e.g., Wingfield et al., 1988). The contrasting aging effects found in the present meta-analysis with those in the adult aging literature highlight the need for life span research in Deaf as well as hearing samples.

The present study is the first to systematically evaluate short-term and working memory of nonimplanted Deaf signers compared with hearing signers. A related systematic review and meta-analysis evaluated these domains with Deaf cochlear implant (CI) users to evaluate the impact of restored auditory access on memory processes. Although their meta-analyses only contained two to four studies per dependent variable, Akçakaya et al. (2022) reported significant effect sizes for digit span ($g = -1.194$), comparable to our results ($g = -1.33$). Their findings of digit span backwards were not significant ($g = -0.26$), which contrasts with our significant verbal working memory population estimate ($g = -0.52$). The backward span pattern across meta-analyses is consistent with the idea that hearing experience may affect performance (e.g., Conway et al., 2009), although the consistent forward span difference across studies is not. Importantly, although our results do closely mirror those

reported by Akçakaya et al., our novel results underscore the importance of evaluating group differences in short-term and working memory prior to cochlear implantation in Deaf signers. Indeed, although longitudinal work would be required to evaluate this claim, comparing results from both meta-analyses may indicate that cochlear implantation does not improve short-term memory of verbal stimuli but may improve working memory of verbal stimuli.

Theoretical and practical implications

The present findings bring clarity to a literature that includes multiple discrepant findings. The first two meta-analyses revealed significant, negative effects of deafness on verbal forward and verbal backward recall, or verbal short-term memory and working memory, respectively. By contrast, the third meta-analysis failed to detect an effect of deafness on visuospatial forward recall, or visuospatial short-term memory. One theoretical approach to understanding the effect of deafness on cognition, the auditory scaffolding hypothesis (Conway et al., 2009), suggests that the Deaf signers' relative lack of experience with sequentially ordered language should result in a general deficit in serial-order task performance. The third meta-analysis, however, challenges this view because it reveals similar performance of Deaf signers and hearing nonsigners on visuospatial serial-order recall. The effect of deafness on forward and backward serial-order recall of verbal information remains, however. With these group differences established clearly by this first two meta-analyses of span scores in Deaf signers and hearing nonsigners, next steps include doing more detailed analyses of the error types that the groups make across conditions. The M3 error model (Oberauer & Lewandowsky, 2019) may be a particularly helpful guide for these next steps (see McFayden et al., 2023, for detailed discussion). In addition to establishing group differences in verbal span, the first two meta-analyses also showed that the size of the Deaf signer disadvantage increased as participant age increased from children/adolescents to adults. Another important next step is to broaden data collection to older adulthood. Indeed, the cognitive aging literature has discussed the effect of disrupted sensory input on cognitive aging (e.g., Phillips et al., 2022). To our knowledge, there has not been studies comparing older adults who are Deaf signers with older adults who are hearing nonsigners. Such data would contribute to current discussions in the cognitive aging literature as well as clarify whether the age-related increase in effect sizes continues throughout adulthood, asymptotes, or is cubic.

The meta-analyses' clarification of the data patterns also has practical implications. For example, there are narratives of Deaf signers being less intelligent compared with hearing nonsigners, which is not empirically supported by the

current findings (cf. the “deaf and dumb” literature, still being published in the 21st century, as recent as 2021; Eling & Finger, 2021). The present findings counter this narrative, in part because there is no group difference in visual short-term memory. The data are also relevant in applied settings. For example, the systematic disadvantage for Deaf signers, compared with hearing nonsigners, on verbal span tasks coupled with no group difference on visuospatial span tasks suggests that the best means of obtaining accurate, nonaudist measures of memory is to assess short-term memory with visuospatial tools. Further discussion about implications for audist biases and equity is below.

Quality

Importantly, the quantitative results presented here may be hedged by quality estimates in the “fair” range, with several articles landing in the “poor” range as indicated by adapted NIH criteria for case-control studies. However, as zero studies included data on three out of the 12 domains assessed by the NIH (e.g., sample size justifications, blinded hypotheses, concurrent controls), our quality results suggest that these metrics are not being routinely considered in research with Deaf participants, perhaps due to unique factors of the sample (e.g., Deaf participants may be harder to recruit, thus reliance on a sample size calculation is not common). Removing these three quality domains significantly increased study quality, which suggests research with Deaf communities is not as systematically rigorous as other experimental work. As the NIH Quality Assessment checklist was created in 2013 and the current study spans over 50 years, only 15 studies were published after the creation of the quality domain criteria. Pearson correlations indicated no significant relationship between publication date and quality score, $r = .24$, $p = .16$ (Supplementary Figure 5), suggesting these quality domains may stand the test of time, and that significant efforts are needed to improve quality of studies with Deaf participants.

Future quality estimates for Deaf scholarship should consider (a) creating discipline-specific quality guidelines in accordance with research ethics with Deaf participants (see Singleton et al., 2014, for a review), or (b) evaluating comprehensiveness of reporting guidelines instead, as evidenced in the current review. Data from the 35 studies suggested comprehensive reporting of age, sex, hearing status, and age of deafness. The category with the least reporting was educational level (46% of studies reported), which is imperative for authors to assess and report given current and historical exclusion of deaf students and scholars from educational pursuits (Schick et al., 2006). Additionally, by recommendation from Deaf scholar, author AM, two other comprehensiveness metrics were evaluated, including vision

screening and access to closed captioning at home, which were reported on at low incidences (closed captioning: 0%, vision: 8%). Future studies may wish to consider what relevant demographic or health considerations are important to consider when working with Deaf communities, as they may relate to task design (e.g., vision may be important to address if conducting a visuospatial task), and/or may occur at higher prevalence rates than in hearing populations (Chia et al., 2006).

One important take-away from the quality assessment process was the lack of Deaf authorship in the majority of articles and studies on Deaf memory. Previous research (e.g., Singleton et al., 2014) has underscored the importance of involving Deaf and signing communities in research to address many ethical research barriers including distrust towards researchers, auditory biases in measure selection and auditory bias in reporting of results. Although rates of Deaf authorship ranged from 38% (articles) to 43% (studies), the authorship metric used in the current meta-analysis was dichotomous and does not account for other areas of bias, such as Deaf tokenism (e.g., feeling that Deaf scholars are sometimes treated as tokens in research teams; Singleton et al., 2014). Funding for research projects should target Deaf scholars conducting research in this domain with an emphasis on Deaf perspectives driving future scholarship.

Conclusion

Nonimplanted, Deaf signers, previously conceptualized to have weaker memory for items presented serially (Wilson et al., 1997), demonstrate similar visual spans and smaller verbal spans compared with hearing nonsigners. The effect sizes observed for significant group differences were in the moderate (backward verbal recall) to large (forward verbal recall) range, suggesting a hearing advantage. These effects are stronger with age, increasing from childhood through adulthood. By contrast, when Deaf signers are assessed in a modality commensurate with their primary language, specifically a visual modality, they demonstrate no significant differences in forward span compared with hearing nonsigners. These results challenge guiding theories that a sign-based language disrupts serial-order processing, as evidenced by no group differences in serial visual recall. As results from the quality assessments of Deaf bias further indicate, our current systems show an audism bias (Eckert & Rowley, 2013) towards hearing nonsigners in the design of studies with Deaf participants and the discussion of results from an audist-centered perspective. Future research should evaluate the types of errors made during short-term and working-memory tasks to further elucidate memory mechanisms, as well as expand developmental research to older adult populations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-023-02282-6>.

Author note Special thanks to the research and synthesis librarians at Virginia Tech for their assistance developing search terms, troubleshooting, and providing training on the systematic review and meta-analysis process: Cozette Comer, Frances McCarty, Kiri Goldbeck DeBose, Ginny Pannabecker, and Rachel Miles. Thank you to Dr. Mike Cheung for his assistance with adaptations to the ‘metafor’ R script package. This research was supported by the Vail Professorship (Davidson College; author K.S.M.) and a training fellowship from NICHD to UNC-Chapel Hill (T32 HD040127; author T.C.M.). The authors have no conflicts of interest to disclose.

Preregistration is available on Prospero: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=167987.

Data, code, and materials are available on Open Science Framework (anonymized link): https://osf.io/vwns8/?view_only=26fd6130332d4cc0a3281a720639352e

References

*References marked with an asterisk indicate studies included in the meta-analysis.

- Akçakaya, H., Jayakody, D. M. P., & Dogan, M. (2022). *Systematic review and meta-analysis of STM and WM in long-term CI users*. Advance online publication. <https://doi.org/10.1007/s40688-022-00408-6>
- Alamargot, D., Lambert, E., Thebault, C., & Dansac, C. (2007). Text composition by deaf and hearing middle-school students: The role of working memory. *Reading and Writing, 20*, 333–360. <https://doi.org/10.1007/s11145-006-9033-y>
- Andin, J., Orfanidou, E., Cardin, V., Holmer, E., Capek, C. M., Woll, B., Rönnerberg, J., & Rudner, M. (2013). Similar digit-based working memory in deaf signers and hearing non-signers despite digit span differences. *Frontiers in Psychology, 4*. <https://doi.org/10.3389/fpsyg.2013.00942>
- *Arfé, B., Rossi, C., & Sicoli, S. (2015). The contribution of verbal working memory to deaf children’s oral and written production. *Journal of Deaf Studies and Deaf Education, 20*(3), 203–214.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2*, 89–195.
- Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America, 93*(24), 13468–13472.
- Baddeley, A., & Hitch, G. J. (1974). Working Memory. *Psychology of Learning and Motivation, 8*, 47–89.
- Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine, 21*(18), 2641–2652.
- Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences, 10*(11), 512–518.
- Bavelier, D., Newport, E. L., Hall, M., Supalla, T., & Boutla, M. (2008). Ordered short-term memory differs in signers and speakers: Implications for models of short-term memory. *Cognition, 107*(2), 433–459.
- Bettger, J., Emmorey, L., McCullough, S., & Bellugi, U. (1997). Enhanced facial discrimination: Effects of experience with American Sign Language. *The Journal of Deaf Studies and Deaf Education, 2*(4), 223–233.
- Blair, F. X. (1957). A study of the visual memory of deaf and hearing children. *American Annals of the Deaf, 102*(2), 254–263 <http://www.jstor.org/stable/44399367>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>
- Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7*, 997–1002.
- Cardin, V., Rudner, M., De Oliveira, R. F., Andin, J., Su, M. T., Beese, L., Woll, B., & Rönnerberg, J. (2018). The organization of working memory networks is shaped by early sensory experience. *Cerebral Cortex, 28*(10), 3540–3554.
- Cheung, M. W.-L. (2019). A guide to conducting meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29*, 387–396.
- Chia, E., Mitchell, P., Rochtchina, E., Foran, S., Golding, M., & Wang, J. J. (2006). Association between vision and hearing impairments and their combined effects on quality of life. *Arch Ophthalmol, 124*(10), 1465–1470.
- *Chincotta, M., & Chincotta, D. (1996). Digit span, articulatory suppression, and the deaf: A study of the Hong Kong Chinese. *American Annals of the Deaf, 141*(3), 252–257.
- Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science, 18*(5), 275–279.
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Unpublished doctoral dissertation. McGill University.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*, 1158–1170.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
- Eckert, R. C., & Rowley, A. J. (2013). Audism: A theory and practice of audiocentric privilege. *Humanity & Society, 37*(2), 101–130.
- *Edwards, L., Marschark, M., Kronenberger, W. G., Crowe, K., & Walton, D. (2021). Inferencing abilities of deaf college students: Foundations and implications for metaphor comprehension and theory of mind. *Journal of Developmental and Physical Disabilities, 33*, 233–258.
- Egger, M., Smith, D. G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634.
- Eling, P., & Finger, S. (2021). Franz Joseph Gall on the “deaf and dumb” and the complexities of the mind. *Journal of the History of the Neurosciences, 30*(2), 128–140.
- *Emmorey, K., Giezen, M. R., Petrich, J. A. F., Spurgeon, E., & Farnady, L. O. (2017). The relation between working memory and language comprehension in signers and speakers. *Acta Psychologica, 177*, 69–77.
- *Farjardo, I., Benedetti, P., Arfé, B., & Altoé, G. (2008). Hyperlink format, categorization abilities and memory span as contributors to deaf users hypertext access. *The Journal of Deaf Studies and Deaf Education, 13*(2), 241–256.
- Flaherty, M., & Moran, A. (2001). Memory span for Arabic numerals and digit words in Japanese kanji in deaf signers. *Japanese Psychological Research, 43*(2), 63–71.
- *Flaherty, M., & Moran, A. (2004). Deaf signers who know Japanese remember words and numbers more effectively than deaf signers who know English. *American Annals of the Deaf, 149*(1), 39–45.
- Geraci, C., Gozzi, M., Papagno, C., & Cecchetto, C. (2008). How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition, 106*(2), 780–804.
- *Gozzi, M., Geraci, C., Cecchetto, C., Perugini, M., & Papagno, C. (2011). Looking for an explanation for the low sign span. Is order

- involved? *The Journal of Deaf Studies and Deaf Education*, 16(1), 101–107.
- Hall, M. L., & Bavelier, D. (2010). Working memory, deafness, and sign language. In M. Marschark & E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (Vol. 2, pp. 458–472). Oxford University Press.
- *Hall, M. L., Eigsti, I.-M., Bortfeld, H., & Lillo-Martin, D. (2018). Executive function in deaf children: Auditory access and language access. *Journal of Speech, Language, and Hearing Research*, 61(8), 1970–1988.
- Hamilton, H. (2011). Memory skills of deaf learners: Implications and applications. *American Annals of the Deaf*, 156(4), 402–423.
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman & Hall/CRC Press. <https://doi.org/10.1201/9781003107347>
- *Harris, M., & Moreno, C. (2004). Deaf children's use of phonological coding: Evidence from reading, spelling, and working memory. *The Journal of Deaf Studies and Deaf Education*, 9(3), 253–268.
- Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107–128.
- Heled, E., & Ohayon, M. (2021). Visuospatial and tactile working memory in individuals with congenital deafness. *The Journal of Deaf Studies and Deaf Education*, 26(3), 314–321.
- *Hermelin, B., & O'Connor, N. (1975). The recall of digits by normal, deaf and autistic children. *British Journal of Psychology*, 66(2), 203–209.
- Higgins, J. P. T., & Green, S. (2009). *Cochrane handbook for systematic reviews of interventions (Version 5.0.2)*. John Wiley & Sons.
- Hirshorn, E., Fernandez, N., & Bavelier, D. (2012). Routes to short-term memory indexing: Lessons from deaf native users of American Sign Language. *Cognitive Neuropsychology*, 29(1/2), 85–103.
- Koo, D., Crain, K., LaSasso, C., & Eden, G. F. (2008). Phonological awareness and short-term memory in hearing and deaf individuals of different communication backgrounds. *Annals of the New York Academy of Sciences*, 1145(1), 83–99.
- Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. *Memory & Cognition*, 13, 265–272.
- Kronenberger, W. G., Henning, S. C., Ditmars, A. M., Roman, A. S., & Pisoni, D. B. (2018). Verbal learning and memory in prelingually deaf children with cochlear implants. *International Journal of Audiology*, 57(10), 746–754.
- Lipsey, M., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). SAGE Publications.
- Logan, K., Mayberry, M., & Fletcher, J. (1996). The short-term memory of profoundly deaf people for words, signs, and abstract spatial stimuli. *Applied Cognitive Psychology*, 10(2), 105–119.
- Marschark, M., Morrison, C., Lukowski, J., Borgna, G., & Convertino, C. (2013). Are deaf students visual learners? *Learning and Individual Differences*, 25, 156–162.
- Marschark, M., Paivio, A., Spencer, L. J., Durkin, A., Borgna, G., Convertino, C., & Machmer, E. (2017). Don't assume deaf students are visual learners. *Journal of Developmental and Physical Disabilities*, 29, 153–171.
- Marschark, M., Sarchet, T., & Trani, A. (2016). Effects of hearing status and sign language use on working memory. *The Journal of Deaf Studies and Deaf Education*, 21(2), 148–155.
- Marschark, M., Spencer, L. J., Durkin, A., Borgna, G., Convertino, C., Machmer, E., Kronenberger, W. G., & Trani, A. (2015). Understanding language, hearing status, and visual-spatial skills. *The Journal of Deaf Studies and Deaf Education*, 20(4), 310–330.
- Marshall, C., Jones, A., Denmark, T., Mason, K., Atkinson, J., Botting, N., & Morgan, G. (2015). Deaf children's non-verbal working memory is impacted by their language experience. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00527>
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56(2), 154–178.
- *McDaniel, E. D. (1980). Visual memory in the deaf. *American Annals of the Deaf*, 125(1), 17–20.
- McFayden, T. C., Faust, M. E., McIntosh, A., Multhaup, K., (2023). Effect of visual presentation format and recall direction on letter span and error patterns in Deaf signing and hearing adults. *Attention, Perception, & Psychophysics*. Advance online publication. <https://doi.org/10.3758/s13414-023-02664-x>
- Multhaup, K. S., Balota, D. A., & Cowan, N. (1996). Implications of aging, lexicality, & item length for the mechanisms underlying memory span. *Psychonomic Bulletin & Review*, 3, 112–120.
- National Institutes of Health. (2021). Quality assessment tool of case-control studies. Available online at: <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools> (Accessed 17, January 2021)
- Norcini, J. J., Jr. (1999). Standards and reliability in evaluation: when rules of thumb don't apply. *Academic Medicine: Journal of the Association of American Medical Colleges*, 74(10), 1088–1090.
- Oberauer, K., & Lewandowsky, S. (2019). Simple measurement models for complex working-memory tasks. *Psychological Review*, 126(6), 880–932.
- *Olsson, J. E., & Furth, H. G. (1966). Visual memory span in the deaf. *The American Journal of Psychology*, 79(3), 480–484.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Bourtron, I., Hoffman, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Phillips, N. A., Isler, L., Kabir, R., Hämäläinen, A., Wittich, W., Pichora-Fuller, M. K., & Mick, P. (2022). Hearing and visual acuity predict cognitive function in adults aged 45–85 years: Findings from the baseline wave of the Canadian Longitudinal Study on Aging (CLSA). *Psychology and Aging*, 37(8), 891–912.
- Powell, D. H., & Hiatt, M. D. (1996). Auditory and visual recall of forward and backward digit spans. *Perceptual and Motor Skills*, 82, 1099–1103.
- Rohatgi, A. (2021). *WebPlotDigitizer* (Version 4.5) [Web application]. <https://automeris.io/WebPlotDigitizer>
- Romero Lauro, L. J., Crespi, M., Papagno, C., & Cecchetto, C. (2014). Making sense of an unexpected detrimental effect of sign language use in a visual task. *The Journal of Deaf Studies and Deaf Education*, 19(3), 358–365.
- Schick, B., Williams, K., & Kupermintz, H. (2006). Look who's being left behind: Educational interpreters and access to education for deaf and hard-of-hearing students. *The Journal of Deaf Studies and Deaf Education*, 11(1), 3–20.
- Singleton, J. L., Jones, G., & Hanumantha, S. (2014). Toward ethical research practice with deaf participants. *Journal of Empirical Research on Human Research Ethics*, 9(3), 59–66.
- Streff, M., Barefoot, S., Walter, G., & Crandall, K. (1978). A comparative study of hearing-impaired and normal-hearing young adults- verbal and nonverbal abilities. *Journal of Communication Disorders*, 11(6), 489–498.
- Tawfit, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, 47. <https://doi.org/10.1186/s41182-019-0165-6>
- Tomlinson-Keasey, C., & Smith-Winberry, C. (1990). Cognitive consequences of congenital deafness. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 151(1), 103–115.
- *Tractenberg, R. E. (2002). Exploring hypotheses about phonological awareness, memory, and reading achievement. *Journal of Learning Disabilities*, 35(5), 407–424.

- *Tzeng, S. (2002). Working memory, language production rate, and reading comprehension of Chinese deaf readers. *Bulletin of Special Education, 22*, 155–169.
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*(2), 215–247.
- Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*, 57–79.
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125.
- Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale* (3rd ed). The Psychological Corporation.
- Wilson, M., Bettger, J. G., Niculae, I., & Klima, E. S. (1997). Modality of language shapes working memory: Evidence from digit span and spatial span in ASL signers. *The Journal of Deaf Studies and Deaf Education, 2*(3), 150–160.
- *Wilson, M., & Emmorey, K. (2006). Comparing sign language and speech reveals a universal limit on short-term memory capacity. *Psychological Science, 17*(8), 682–683.
- Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research, 14*(2), 103–107.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.