Sex/gender differences in verbal fluency and verbal episodic memory – a meta-analysis

Journal:	*Perspectives on Psychological Science*
Manuscript ID	PPS-19-077.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	10-Jan-2022
Complete List of Authors:	Hirnstein, Marco; University of Bergen, Department of biological and medical psychology; University of Bergen Stuebs, Josephine; Maastricht University, Department of Neuropsychology and Psychopharmacology; University of Bergen, Department of Biological and Medical Psychology ; University of Oslo, Institute of Clinical Medicine Moè, Angelica; University of Padova, Department of General Psychology Hausmann, Markus; Durham University, Psychology
Keywords:	Action / Performance, Individual Differences, Sex / Gender, Methodology: Behavioral, Language / Communication, Cognition
User Defined Keywords:	cognitive gender differences, age, verbal ability, author effects

Abstract

Women are thought to fare better in verbal abilities, especially in verbal fluency and verbal memory tasks. However, the last meta-analysis on sex/gender differences in *verbal fluency* dates from 1988. While *verbal memory* has only recently been investigated meta-analytically, a comprehensive meta-analysis is lacking that focuses on verbal memory as it is typically assessed, for example, in neuropsychological settings. Based on 496 effect sizes and 355,173 participants, the current meta-analysis found that women/girls outperformed men/boys in phonemic fluency ($d = 0.12$-0.13) but not in semantic fluency ($d = 0.01$-0.02), where the sex/gender difference appeared to be category-depended. Women/girls also outperformed men/boys in recall ($d = 0.28$) and recognition ($d = 0.12$-0.17). Although effect sizes are small, the female advantage was relatively stable over the past 50 years and across lifetime. Published articles reported stronger female advantages than unpublished studies and first authors reported better performance for members of their own sex/gender. We conclude that a small female advantage in phonemic fluency, recall, and recognition exists and is partly subject to publication bias. Considerable variance suggests further contributing factors, such as participants' language and country/region.

Key words: verbal ability, gender, cognitive gender differences, verbal memory, age, author effects

Public significance statement: This meta-analysis showed that girls and women outperform boys and men in verbal episodic memory and phonemic fluency (e.g., generating words that start with a certain letter). The female advantage was smaller than previously thought, but still has important implications, for example, when verbal fluency and memory tasks are used in psychological assessment.

Sex/Gender Differences in Verbal Fluency and Verbal Episodic Memory. A Meta-analysis

After more than 100 years of psychological research, sex/gender[1] differences in cognitive abilities are still heavily debated (for review Halpern, 2012; Hyde, 2014). Spatial and mathematical abilities, in which men are commonly believed to excel, are very well researched. For instance, a male advantage in mental rotation, the ability to rotate complex figures in one's mind, has been reported in several meta-analyses with effect sizes around Cohen's d = 0.56-0.73 (Linn & Petersen, 1985; Voyer et al., 1995; Zell et al., 2015). By comparison, much less is known about verbal abilities, in which women/girls are commonly believed to excel. There is no unitary concept of verbal abilities but it relates to all aspects of open or inner language production and comprehension. Meta-analyses reported female advantages with medium effect sizes for writing ability, d = 0.53 to 0.61 (Hedges & Nowell, 1995), and reading comprehension, d = 0.23 to 0.68 (Reilly, 2012; Stoet & Geary, 2013). Verbal intelligence/reasoning (Feingold, 1988) and vocabulary (Hyde & Linn, 1988), on the other hand, did not reveal a female advantage with effect sizes smaller than d = 0.05 (Hyde, 2005, 2014).

The two verbal abilities, however, that textbooks and review articles typically refer to when claiming the existence of a female advantage are *verbal fluency* (sometimes also called "word fluency") and *verbal memory* (Andreano & Cahill, 2009; Halpern, 2012; Hamson et al., 2016; Hyde, 2014; Kimura, 2000; Miller & Halpern, 2014). Verbal fluency and verbal memory tests correlate with general cognitive abilities (Alexander & Smales, 1997; Kraan et al., 2013) and are frequently used in psychological assessments of developmental impairments in children

[1] Cognitive differences between men/boys and women/girls arise from a complex interplay of biological, psychological, *and* socio-cultural factors. These factors would be so intertwined that it would not be logical to distinguish between biology ('sex') and social environment ('gender'). In the current study, we therefore aimed for a neutral terminology, avoiding 'sex' or 'gender' as separate terms and instead using 'sex/gender' whenever possible. In certain contexts, however, it would be inappropriate to use "sex/gender" when addressing specific biological or social constructs, such as 'gender equality', 'gender stereotypes', 'sex hormones' or 'sex chromosomes'. When addressing first/last author effects, we refer to gender, because we identified authors as males or females simply based on their first name, not knowing their biological sex or gender identity.

(Gaillard et al., 2003; Pennington & Ozonoff, 1996), impairments and rehabilitation after stroke (Baldo et al., 2006; Barker-Collo & Feigin, 2006), and cognitive decline in dementia (Collie & Maruff, 2000; Zhao et al., 2013).

Verbal Fluency

Verbal Fluency refers to the ability to generate (orally or written) as many words as possible that fulfil a certain criterion, normally under time restrictions. The criterion is typically either *semantic*, also called "categorical fluency" (e.g., naming animals, fruits, etc.) or *phonemic* (e.g., naming words that begin with a specific letter), also called "lexical/letter fluency". Virtually all articles that claim women's/girls' superiority in verbal fluency refer to a landmark meta-analysis by Hyde and Linn (1988), who examined sex/gender differences in a few verbal abilities. The authors concluded that "speech production" or "verbal production" favored women by $d = 0.33$. However, the definition of "speech production" ("as occurs in essay writing or measures of spoken language", p. 55) is different from the verbal fluency definition above and, consequently, some studies in Hyde and Linn (1988) assessed different verbal abilities, such as quality of essays or written sentences (Harris & Seibel, 1976; Wormack, 1979) or how many words four-year-old children speak (Brownell & Smith, 1973). Moreover, the meta-analysis was based on only 14 studies, while the Web of Knowledge revealed that ~7500 references included the term "verbal fluency" since 1988.

Phonemic versus Semantic Fluency, Age, Cohort Effects, and Gender of First/Last Author

Heister (1982) found a female advantage when participants were asked to generate words beginning with the letters 'S' and 'M' (phonemic fluency), while no sex/gender differences emerged for naming things that are 'red' or 'round' (semantic fluency). Other studies reported a female advantage in semantic fluency (Acevedo et al., 2000) or did not find a sex/gender difference in either phonemic or semantic fluency (Kavé, 2005). Overall, it is unclear whether a female advantage exists in both semantic and phonemic fluency.

Furthermore, it is unclear at what age the putative female advantage arises and whether it changes across the lifespan. Some studies suggest a steeper decline in older men as compared to women (Maylor et al., 2007; Rodriguez-Aranda & Martinussen, 2006), while de Frias et al. (2006) found that the female advantage in semantic fluency was stable between 35 and 80 years. Based on semantic fluency data from more than 30,000 individuals (aged 50 to 84 years) in fourteen European countries, Weber et al. (2017; 2014) showed that women from younger cohorts performed better than women from older cohorts. Sex/gender differences also varied across European countries. Both findings were interpreted to show the impact of better access of women to resources and education (Weber et al., 2017; 2014). So far, it is unclear whether sex/gender differences in verbal fluency change with age or across cohorts.

Finally, Hyde and Linn (1988) found that female first authors reported a stronger female advantage (d = 0.15) than male first authors (d = 0.08). However, this finding was based on all verbal abilities and, though statistically significant, the difference was considered to be unsubstantial. The current study sought to replicate the findings by Hyde and Linn (1988) but more specifically with respect to verbal fluency. In addition, we also investigated the influence of gender of the last author, who is often the supervisor or more senior researcher overseeing the research effort.

Verbal Episodic Memory

As with verbal ability, there is no unitary definition of verbal memory. Nevertheless, there is a multitude of empirical data on what researchers considered verbal memory. Several studies found better performance in women (Catani et al., 2007; de Frias et al., 2006; Herlitz et al., 1997; Lowe et al., 2003) and a narrative review concluded that “females show an advantage at verbal memory” (Andreano & Cahill, 2009, p. 260). However, other studies found no sex/gender differences in verbal memory (Munnelly, 2016; Parsons et al., 2005). Meta-analyses on this issue were lacking until recently. Voyer et al. (2021) focused specifically on verbal *working* memory and found an overall significant female advantage that, however, was practically zero (*Hedge's* $g = 0.03$). Furthermore, sex/gender differences varied across different sample and task parameters: tasks with cued recall ($g = 0.08$) and free recall ($g = 0.15$) had a slightly elevated female advantage while there was a male advantage in complex span ($g = 0.04$) and no significant sex/gender difference in serial recall ($g < 0.01$) and simple span ($g < 0.01$).

Another meta-analysis (Asperholm et al., 2019) investigated sex/gender differences in long-term memory, specifically episodic memory. Long-term memory is typically divided into declarative (explicit) and non-declarative (implicit) memory, where declarative memory comprises episodic memory (i.e., the ability to remember specific events or situations at a particular place at a particular time) and semantic memory (i.e., the ability to remember concepts and facts). Asperholm et al. (2019) investigated sex/gender differences in episodic memory for different stimuli, including images, movies, faces, routes, locations but also verbal content such as words/sentences. Verbal content showed a small female advantage ($g = 0.28$). A wide range of studies/tasks were included in the verbal episodic category and the authors investigated whether the female advantage varied across, for example, neutral stimuli versus emotional stimuli, intentionally learnt versus incidentally learnt, or recall versus recognition. Subsequent analyses of moderator variables such as age, publication year, or geographical region took into account whether the stimulus material was verbal, images, movies, or faces, but did not distinguish between incidental/intentional, emotional/neutral, or recall/recognition, and only peer-reviewed articles were included.

Like Asperholm et al. (2019), the present study was interested in episodic long-term memory and thus discarded studies/tasks that primarily assess working memory. In contrast to Asperholm et al. (2019), the current study had a narrower focus on verbal episodic memory, which we investigated with a broader literature search. That is, we examined exclusively *verbal* episodic memory (not memory for routes and locations) and only included studies with neutral stimuli (as opposed to emotional stimuli), in which participants learnt material intentionally (as opposed to incidentally). The intentional learning of neutral stimuli is a key feature of frequently used neuropsychological tests on verbal long-term memory, such as the California Verbal Learning Test (CVLT; Delis et al., 2000), the Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996), or the Wechsler Memory Scale (WMS; (Wechsler, 2009). Further in contrast to Asperholm et al. (2019), the literature search of the current study also comprised 'gray' literature such as PhD/Master's theses to investigate whether sex/gender differences are subject to publication effects. Moreover, the current study examined, for the first time, possible effects of first/last authors' gender on sex/gender differences in verbal episodic memory. Finally, we performed these analyses separately for *recognition* (i.e., when cues are provided for the material that had to be memorized) and *recall* (i.e., absence or lack of cues), because the female advantage appeared to be consistently larger for recall than for recognition (Asperholm et al., 2019; Voyer et al., 2021). The fact that only 14 and 18 of our 168 included studies overlapped with Voyer et al. (2021) and Asperholm et al. (2019), respectively, demonstrates that different aspects of verbal memory were investigated in the current study. Henceforth, we thus used the term "verbal episodic memory" when referring to the data that were analyzed in the present study, and "verbal memory" when referring to verbal memory in general.

Aims and Hypotheses

A female advantage is frequently assumed in verbal fluency and verbal memory. For *verbal fluency,* this assumption is based on an early meta-analysis by Hyde and Linn (1988) that required an update. For *verbal memory*, a meta-analysis was missing that focuses specifically on *verbal* episodic memory – complementary to two recent meta-analyses about verbal working memory (Voyer et al., 2021) and episodic memory in general (Asperholm et al., 2019). The present study thus aimed to reveal the magnitude of the putative female advantage in verbal fluency and verbal episodic memory. For both, we additionally examined the impact of potentially modulating factors such as publication year, type of publication (articles versus PhD/Master theses), participants' age, semantic fluency versus phonemic fluency, recall versus recognition, and gender of first/last author. We hypothesized a female advantage (i) in both verbal fluency and verbal episodic memory of intentionally learnt neutral stimuli (Andreano & Cahill, 2009; Halpern, 2012; Miller & Halpern, 2014), (ii) that has increased over the past 50 to 60 years due to better access to education for women (Weber et al., 2017; Weber et al., 2014), (iii) which emerges across all age groups but becomes larger in older adults (Maylor et al., 2007; Rodriguez-Aranda & Martinussen, 2006), and (iv) is affected by the gender of the first (Hyde & Linn, 1988) and last author.

Methods

The meta-analysis including literature search, study selection, data analysis, and presentation of results was performed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009) and the recommendations for meta-analyses described by Borenstein et al. (2009). Data analysis was carried out with Comprehensive Meta-Analysis (CMA) version 3.3.070 (Borenstein et al., 2014).

Literature Search and Study Selection

Search Terms and Databases

Between 22nd to 29th October 2016, the databases PsychInfo, ISI Web of Knowledge, and PubMed were searched for relevant literature. Between 13th and 19th September 2019, we additionally searched the "ProQuest Dissertation & Theses" database to identify unpublished PhD and Master's theses. The search terms and number of identified references are listed in Supplementary Table S1. An additional 16 studies were identified through other sources such as comprehensive literature reviews and references used in previously identified publications. After removing 38,322 duplicates, the remaining 28,305 hits were screened for suitability. Screening comprised reading both title and full abstract. In isolated cases, references were excluded solely based on title, for example, in case the title indicated that the reference was a review or meta-analysis without original data, or the topic of the reference was outside the scope of the present meta-analysis (e.g., "Persephone in the underworld: the motherless hero in novels by Burney, Radcliffe, Austen, Bronte, Eliot, and Woolf"). Some older PhD and Master's theses often did not have abstracts, in which case the whole thesis was screened. Details about the exclusion criteria and procedure during screening is provided in the Supplementary material.

Study Selection: Final Inclusion Criteria

Of the 2,984 references that were included after screening of abstract/title, 72 full-texts could not be obtained. The remaining 2,912 references then underwent a full-text search for eligibility. Inclusion criteria were:

(1) Use of phonemic/semantic fluency and/or verbal episodic memory (recognition/recall) tests that comply with the aforementioned definitions of verbal fluency and verbal episodic memory. Examples for verbal fluency are the Controlled Oral Word Association Test (COWAT; Benton, 1967) or the F-A-S Test (Spreen & Benton, 1977), the Thurstone Word Fluency Test (Thurstone & Thurstone, 1962), or any test in which participants had to generate as many words as possible starting/ending with, or containing certain letters, as well as providing as many examples as possible for a specific category. Not included were data from tests such as finding synonyms or essay writing (which were considered too peripheral for verbal fluency). Anagram tasks were excluded on the grounds that they draw on numerical and spatial abilities (Wilson et al., 1954).

For verbal episodic memory, we excluded tasks that measured exclusively or predominantly working memory such as Digit span forwards or backwards from the Wechsler Adult Intelligence Scales (Wechsler, 2008). Examples for included verbal episodic memory tests are the Visual Verbal Learning Test (Brand & Jolles, 1985), the RAVLT, and the CVLT. Logical Memory II and Logical Memory Recognition (remembering a story) from the WMS were included but not Logical Memory I, because this subtest is more related to verbal working memory. If multiple verbal episodic memory parameters were provided (e.g., delayed recall, total recall, recall), we retained the total score; otherwise, the provided scores were kept. Learning in all verbal episodic memory measures had to be intentionally (i.e., incidental learning measures were not included).

(2) For both verbal fluency and episodic memory, we excluded tasks that employed emotional stimuli, as they could be confounded with sex/gender differences in emotional processing (Kret & De Gelder, 2012; Stevens & Hamann, 2012). For example, affective semantic fluency categories such as 'pleasant/unpleasant' or 'joy/fear' (e.g., Gawda & Szepietowska, 2013a, 2013b) were not included.

(3) Verbal fluency/episodic memory stimuli were not presented laterally, that is, to one specific hemisphere. For example, tasks that employed laterality paradigms were not considered because of sex/gender differences in hemispheric asymmetry (Hirnstein et al., 2019).

(4) Verbal fluency/episodic memory tasks were not performed simultaneously with other tasks, as multitasking abilities might vary across men and women (Hirnstein et al., 2018).

(5) The publication contained quantitative, empirical data (i.e., no reviews, study protocols, meta-analyses, etc.), which allowed computation of the effect size, as well as the exact number (or percentages) of male and female participants. Only "pure" verbal fluency and verbal episodic memory measures were included. That is, if covariates such as intelligence had

been factored in, the data was excluded. If only aggregate scores were provided from test batteries comprising both eligible and not eligible tasks, data were excluded. Finally, when studies reported multiple verbal fluency/episodic memory tasks, but only provided statistical parameters to compute effect sizes for tests that found significant sex/gender differences – and insufficient statistical parameters for tests that did not find sex/gender differences – the whole study was discarded to avoid introducing a bias towards significant results.

(6) There were at least 10 male and 10 female participants in the sample to mitigate the effect of spurious findings with very small sample sizes.

(7) Participants were healthy individuals without a mental or other condition that could affect verbal fluency/episodic memory performance (e.g., depression, Alzheimer's disease, learning disability) and were not under the influence of any kind of substance, medicine or other factors that might influence cognitive performance (e.g., sleep deprivation, noise exposure, etc.). Data from control groups could be included, unless controls were selected for specific features (e.g., intelligence, age, socioeconomic status) to match clinical groups.

(8) Participants were not preselected for a specific feature that could potentially be related to verbal fluency/episodic memory performance (e.g., participants with certain gene combination(s), participants who performed better than average on a creativity test, samples with homosexual participants only).

(9) The publication was written in English, German, or any Scandinavian language.

(10) Cohen's *d* was outside the range of -4.0 to 4.0, which we deemed unrealistic. The range of included effect sizes was -1.07 to 1.42.

Where inclusion criteria were met but the study lacked important quantitative information (e.g., number of men/women/boys/girls, means, or *p*-values), authors were contacted with a request to provide the relevant data and other relevant data they have or know of. Out of 45 contacted authors, nine provided relevant data.

In total, 496 effect sizes from 168 references were included for quantitative analysis, comprising data from 355,173 participants (men/boys = 178,409, women/girls = 176,764). For a more detailed overview of the study selection process including reasons that led to exclusion, see Figure 1. For a complete list of all included references and effect sizes, see Supplementary Table S2.

Statistical Analysis

For each relevant measure from the included references above, standardized differences in means (Cohen's *d*) were computed based on the available statistical information. If the male/female distribution was given in percentages, they were converted into integers. The effect direction was set such that positively signed values indicate a female advantage and negatively signed values a male advantage. A value of zero indicates the absence of any male/female advantage. We consistently applied the random-effects model, because (a) we expected substantial between-study variance and (b) we aimed to generalize our findings to the entire population. Moreover, we consistently used subgroups within a reference as the unit of analysis (as opposed to using the whole reference as the unit of analysis). That is, if a study included a verbal episodic memory measure from two age groups (e.g., one 50-59 years and another 60-69 years), those subgroups were treated as separate measures, rather than combining them into one measure.

Several studies reported multiple outcomes for each (sub-)sample. For example, a study could provide data from two different tests that both measure recall. It is likely that those tests were correlated with each other and the magnitude of that correlation impacts the variance, and thus the likelihood of finding statistically significant results (Borenstein et al., 2009). Since these correlations were rarely reported, we ran each analysis twice: once with $r = 0$, assuming perfect *independence* of the outcomes, and once with $r = 1.0$, assuming perfect *correlation* between outcomes. In most cases, the results of both analyses yielded similar results. For ease of reading, we always reported the perfect independence results first. All tables/figures were based on the assumption of perfect independence.

Overall Sex/Gender Effects

First, we computed the overall sex/gender effect, separately for verbal fluency and verbal episodic memory. Then, we computed the overall sex/gender effect for each of the following four verbal ability measures: Phonemic and semantic fluency as measures of verbal fluency as well as recognition and recall as measures of verbal episodic memory. One study had aggregated phonemic and semantic fluency scores into a combined verbal fluency score (DeWan, 2006), while another had aggregated recognition and recall scores into combined verbal episodic memory scores (Rouch et al., 2005). Effect sizes from these studies were thus kept in the overall verbal fluency/episodic memory analysis but excluded from the recognition/recall/phonemic/semantic fluency analysis.

For all these analyses, we provided *Q*-statistic (testing the null hypothesis that all studies in the analysis shared a common effect size), I^2 (the proportion of observed variance that reflects difference in true effect sizes, rather than sampling error), and T^2 (the variance of true effect sizes) as indicators of how much the sex/gender effect varied across studies. To address the issue of publication bias, we reported Egger's regression (two-tailed) (Egger et al., 1997) and funnel plots (see Supplementary Figure S1).

Effects of Publication Year, Publication Type, Age, and Gender of First/last Author

To investigate whether sex/gender differences change with publication year (as an indicator for changes over time), vary across publication type (articles versus PhD/Master theses), age, and the gender of the first/last author, we ran a set of *meta-regressions*. Meta-regressions have the advantage that they allow investigating the effect of one factor while controlling for a set of other factors (Borenstein et al., 2009). Here again, we assumed that the true effect size varied across studies and thus applied a random-effects model (method of moments). All tests were two-sided, based on Z-distribution.

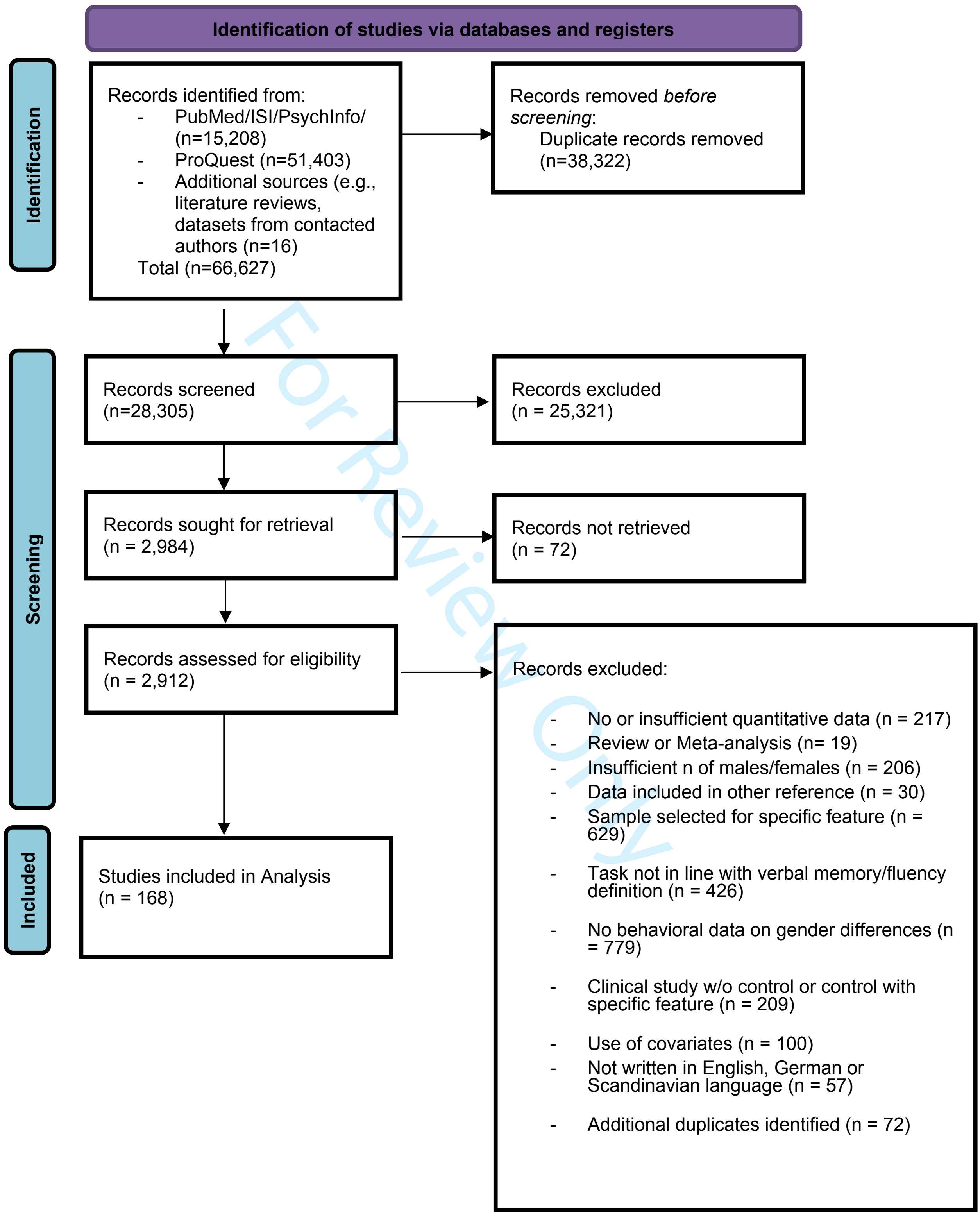
Six covariates were created for the meta-regressions: (1) The continuous covariate “Publication year" simply coded the year when a reference was published. (2) “Publication type” was a categorical covariate that could either be “Published article” or “PhD/Master’s thesis”. (3) Age was analyzed with two covariates: “Mean age” as a continuous variable, which was either obtained directly from the corresponding reference or, in case that information was missing, computed based on the age range (e.g., an age range of 40-60 years would lead to a mean age of 50 years). If age ranges were provided separately for men/boys and women/girls, we took the youngest and oldest age from either sex/gender. If mean ages were provided separately for women/girls and men/boys, we calculated a weighted overall mean. Using “Mean age” alone, however, has two shortcomings. First, several studies only provide age information such as “>70 years”, making it impossible to calculate a mean. Second, many studies have enormous age ranges. For example, ca. 20% of studies had age ranges of 40 years and more, rendering “Mean age” a rather coarse indicator. (4) For this reason, we created a second covariate to examine age effects: “Age groups”. This was a categorical covariate, theoretically grounded in the Medical Subject Heading (MeSh), the standardized vocabulary used in the Medline database for indexing, developed by National Library of Medicine. According to this classification, the following age categories were formed: “Child/Child preschool” (2-12 years), “Adolescent” (13-18 years), “Adult” (19-44 years), “Middle aged” (45-64 years), and “Aged” (65+ years). Effect sizes were grouped into those categories based on the reported age range of the corresponding study. For example, an effect size based on a sample with an age range of 20-27 years, was classified as “Adult”. An effect size based on an age range of 17-40 years was coded blank and excluded from the “Age groups” analysis. As a consequence, the number of effect sizes was substantially higher for “Mean age” (92%, 455/497) than for “Age groups” (51%, 253/497). While both age measures have their respective shortcomings, we combined both as this allows a reasonable estimate of age effects (see also Voyer et al., 2021). Finally, (5) and (6) were the categorical covariates “First author gender” and “Last author gender”, respectively, which was either male or female. In case of single-author studies, this was coded as first author and was not included for analysis of last author effects.

The categorical covariates described above were dummy-coded in order to be entered into the meta-regression. This was done such that “Published articles”, “Males”, and "Adult” served as reference groups for “Publication type”, “First/last author gender”, and “Age groups”, respectively. We did not include language as a covariate, because there were too few non-English reports of data. For comparison: 263 out of 496 effect sizes (53%) were reported in English, while the second most frequent language, Dutch, only comprised 40 effect sizes (8%).

We ran a sequence of meta-regressions for each verbal ability (i.e., recall, recognition, phonemic/semantic fluency) separately. The first meta-regression always included the covariates “Publication year”, “Mean age”, “Publication type”, and “First author gender”. This was done to maximize the number of available effect sizes. “Age groups” was not entered into the first meta-regression because of multicollinearity with “Mean age” and because only half of the effect sizes could be assigned to a specific age group (see above). We thus ran a second meta-regression that included “Age group” and all significant covariates from the first meta-regression as a control (except for “Mean age” due to multicollinearity). “Last author gender” was also not entered into the first meta-regression because of multicollinearity with “Publication type”: None of the PhD/Master’s theses have a last author. Therefore, we ran a third meta-regression for “Published articles” only that included “Last author gender” and all significant covariates from the first meta-regression as a control (except for “Publication type” due to multicollinearity).

Figure 1

PRISMA Flow Diagram Showing the Study Selection Process



Results

Overall Sex/Gender Differences

Effect sizes of the most frequent verbal fluency and verbal episodic memory measures are presented in Table 1.

Verbal Fluency

Assuming perfect independence between multiple outcomes within the same study, the overall effect size was $d = 0.07$ with a 95% confidence interval (CI) of 0.04 to 0.10, based on 290 effect sizes. The female advantage deviated significantly from zero with $Z = 5.10$, $p < .001$. There was substantial heterogeneity among studies ($Q(289) = 2085.1$, $p < .001$, $I^2 = 86.1$ %, $T^2 = 0.02$). Egger's regression intercept of -0.10 was not significant, $t(288) = 0.54$, $p = .591$.

Assuming perfect correlation between multiple outcomes within the same study, all effects remained significant/non-significant: $d = 0.07$, 95% CI [0.04 0.10], $Z = 4.60$, $p < .001$, $Q(209) = 1784.3$, $p < .001$, $I^2 = 88.3$ %, $T^2 = 0.02$, Egger's intercept = -0.13, $t(208) = 0.52$. $p = .602$, 210 effect sizes.

Verbal Episodic Memory

Assuming perfect independence, there was a significant female advantage with $d = 0.23$, 95% CI [0.19 0.26], $Z = 13.09$, $p < .001$, based on 206 effect sizes. Heterogeneity was substantial ($Q(205) = 1622.7$, $p < .001$, $I^2 = 87.4$ %, $T^2 = 0.04$). Egger's intercept was 1.08, $t(204) = 3.94$, $p < .001$. Assuming perfect correlation, all effects remained significant/non-significant: $d = 0.26$, 95% CI [0.21 0.30], $Z = 11.39$, $p < .001$, $Q(132) = 1194.1$, $p < .001$, $I^2 = 88.9$ %, $T^2 = 0.04$, Egger's intercept = 1.18, $t(131) = 3.45$, $p < .001$, 133 effect sizes.

Phonemic Fluency

There was a significant female advantage with an effect size $d = 0.13$, 95% CI [0.09 0.16], $Z = 6.75$, $p < .001$, based on 135 effect sizes. There was significant heterogeneity ($Q(134) = 272.3$, $p < .001$, $I^2 = 50.8$ %, $T^2 = 0.01$). Egger's intercept was 0.19, $t(133) = 1.04$,

$p = .30$. Assuming perfect correlation, all effects remained significant/non-significant: $d = 0.12$, 95% CI [0.09 0.16], $Z = 6.97$, $p < .001$, $Q(128) = 226.9$, $p < .001$, $I^2 = 43.6$ %, $T^2 = 0.01$, Egger's intercept = 0.20, $t(127) = 1.14$. $p = .25$, 129 effect sizes.]

Semantic Fluency

There was no significant sex/gender difference in semantic fluency with $d = 0.02$, 95% CI [-0.02 0.06], $Z = 1.00$, $p = .315$, based on 147 effect sizes. The effect varied significantly across studies ($Q(146) = 1782.6$, $p < .001$, $I^2 = 91.8$ %, $T^2 = 0.03$) and Egger's intercept -0.61, $t(145) = 1.78$, $p = .078$. Assuming perfect correlation, all effects remained significant/nonsignificant: $d = 0.01$, 95% CI [-0.02 0.05], $Z = 0.70$, $p = .482$, $Q(136) = 1740.1$, $p < .001$, $I^2 = 92.2$ %, $T^2 = 0.03$, Egger's intercept = -0.68, $t(135) = 1.86$. $p = .065$, 137 effect sizes.

Recall

There was a significant female advantage with $d = 0.28$, 95% CI [0.23 0.32], $Z = 12.54$, $p < .001$, based on 136 effect sizes. The effect varied largely between studies ($Q(135) = 1217.0$, $p < .001$, $I^2 = 88.9\%$, $T^2 = 0.04$). Egger's intercept was 1.32, $t(134) = 3.94$, $p < .001$. Assuming perfect correlation, all effects remained significant/non-significant: $d = 0.28$, 95% CI [0.24 0.33], $Z = 11.90$, $p < .001$, $Q(123) = 1155.3$, $p < .001$, $I^2 = 89.4$ %, $T^2 = 0.04$, Egger's intercept = 1.35, $t(123) = 3.85$. $p < .001$, 124 effect sizes.

Recognition

There was a significant female advantage with $d = 0.12$, 95% CI [0.06 0.17], $Z = 4.42$, $p < .001$, 66 effect sizes. The effect varied significantly across studies ($Q(65) = 257.1$, $p < .001$, $I^2 = 74.7$ %, $T^2 = 0.02$). Egger's intercept was 1.27, $t(64) = 3.11$, $p = .003$. Assuming perfect correlation, all effects remained significant/non-significant: $d = 0.17$, 95% CI [0.10 0.24], $Z = 4.78$, $p < .001$, $Q(49) = 164.9$, $p < .001$, $I^2 = 70.3$ %, $T^2 = 0.03$, Egger's intercept = 1.08, $t(48) = 2.42$. $p = .019$, 50 effect sizes.

Table 1

Descriptive overview of sex/gender differences in verbal fluency and verbal episodic memory measures

Verbal ability	Test/Measure	Effect size
Verbal fluency	**Total effect**	***d* = 0.07 (0.04 to 0.10) *k*=290**
Phonemic fluency	**Total effect**	***d* = 0.13 (0.09 to 0.16) *k*=135**
	Generic starting letter(s)	*d* = 0.12 (0.07 to 0.18) *k*=59
	COWAT/F, A, S	*d* = 0.14 (0.08 to 0.20) *k*=55
	Four-word sentences	*d* = 0.03 (-0.20 to 0.26) *k*=5
Semantic fluency	**Total effect**	***d* = 0.02 (-0.02 to 0.06) *k*=147**
	Category: Animals	*d* = -0.13 (-0.16 to -0.09), *k*=58
	Categories Animals & fruits/vegetables/food	*d* = 0.11 (0.03 to 0.18), *k*=26
	Objects with specific color	*d* = 0.19 (0.13 to 0.25), *k*=10
	Categories Animals, fruits/vegetables/food, & action verbs	*d* = 0.25 (-0.03 to 0.53), *k*=8
	Fruits/vegetables/food	*d* = 0.31 (0.16 to 0.47), *k*=8
Verbal episodic memory	**Total effect**	***d* = 0.23 (0.19 to 0.26) *k*=206**
Recall	**Total effect**	***d* = 0.28 (0.23 to 0.32), *k*=136**
	California Verbal Learning Test	*d* = 0.42 (0.32 to 0.52), *k*=28
	Rey Auditory Verbal Learning Test	*d* = 0.39 (0.29 to 0.48), *k*=24
	Generic word list	*d* = 0.17 (0.06 to 0.28), *k*=16
	Delayed Memory for Names/Visual-Auditory Learning from Woodcock Johnson Psycho-Educational Battery - Revised	*d* = -0.13 (-0.27 to 0.01), *k*=12
	10 Word Learning Test from CERAD	*d* = 0.18 (0.07 to 0.28), *k*=10
	Ten-Words Test	*d* = 0.26 (0.13 to 0.39), *k*=7
	Deese, Roediger and McDermott task	*d* = 0.15 (0.02 to 0.28), *k*=7
Recognition	**Total effect**	***d* = 0.12 (0.06 to 0.17), *k*=66**
	Rey Auditory Verbal Learning Test	*d* = 0.22 (0.12 to 0.33), *k*=18
	California Verbal Learning Test	*d* = 0.17 (0.06 to 0.29), *k*=13
	Deese, Roediger and McDermott task	*d* = 0.15 (0.04 to 0.27), *k*=7
	Story telling delayed recognition	*d* = -0.07 (-0.18 to 0.04), *k*=7
	Story telling immediate recognition	*d* = 0.02 (-0.09 to 0.13), *k*=7

Note. Values in parentheses represent 95% CI and k = number of effect sizes included. Effect sizes are provided assuming independence between multiple outcomes within the same study. Effect sizes within each sub-category were combined with a random-effects model, assuming a common among study variance component across sub-categories. That is, T^2 was computed for each age group and then pooled across subgroups. Only tests with at least seven effect sizes are provided, with the exception of phonemic fluency, where the three most frequent tests are provided.

Meta-regressions for Moderator Variables

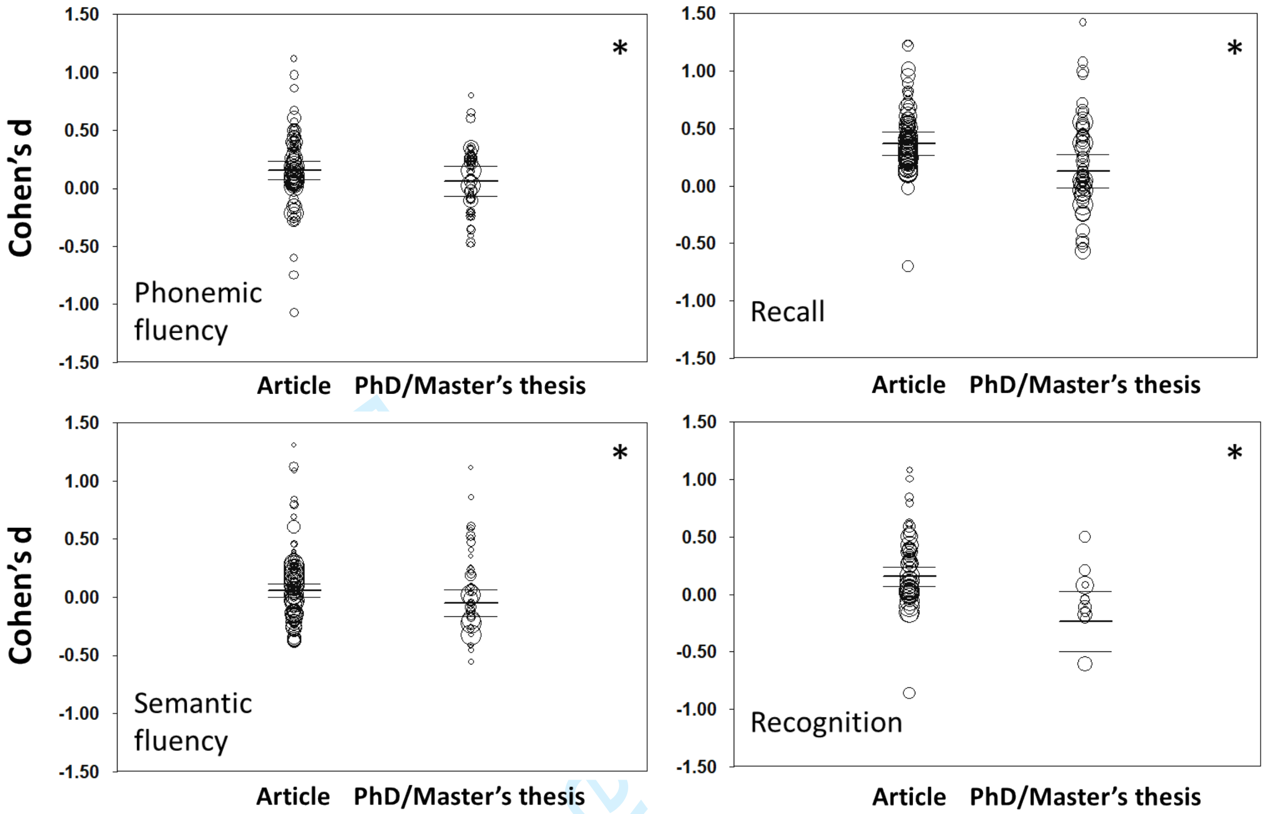
The first set of meta-regressions contained the predictors “Publication year”, “Publication type”, “First author gender”, and “Mean age”. Assuming perfect independence, all four models explained a significant proportion of between-study variance: Phonemic fluency ($Q(4) = 15.75$, $p = .003$, $R^2 = 3.6$, 125 effect sizes), semantic fluency ($Q(4) = 28.94$, $p < .001$, $R^2 = 51.0\%$, 129 effect sizes), recall ($Q(4) = 28.76$, $p < .001$, $R^2 = 23.5\%$, 124 effect sizes), and recognition ($Q(4) = 33.03$, $p < .001$, $R^2 = 31.3\%$, 65 effect sizes). Assuming perfect correlation, all four models remained significant: Phonemic fluency ($Q(4) = 18.04$, $p = .001$, $R^2 = 11.2\%$, 119 effect sizes), semantic fluency ($Q(4) = 35.66$, $p < .001$, $R^2 = 53.2$, 120 effect sizes, recall ($Q(4) = 25.89$, $p < .001$, $R^2 = 23.9$, 111 effect sizes), and recognition ($Q(4) = 23.80$, $p < .001$, $R^2 = 36.2$, 49 effect sizes).

Published Articles versus PhD/Master’s Theses

Published articles consistently reported significantly higher female performance than PhD/Master’s theses: Phonemic fluency ($Z = 2.00$, $p = .045$, $B = -0.093$), semantic fluency ($Z = 2.77$, $p = .006$, $B = -0.108$), recall ($Z = 4.01$, $p < .001$, $B = -0.243$), and recognition ($Z = 4.58$, $p < .001$, $B = -0.390$), see Figure 2. Assuming perfect correlation, all four effects remained significant.

Figure 2

Effect of publication type



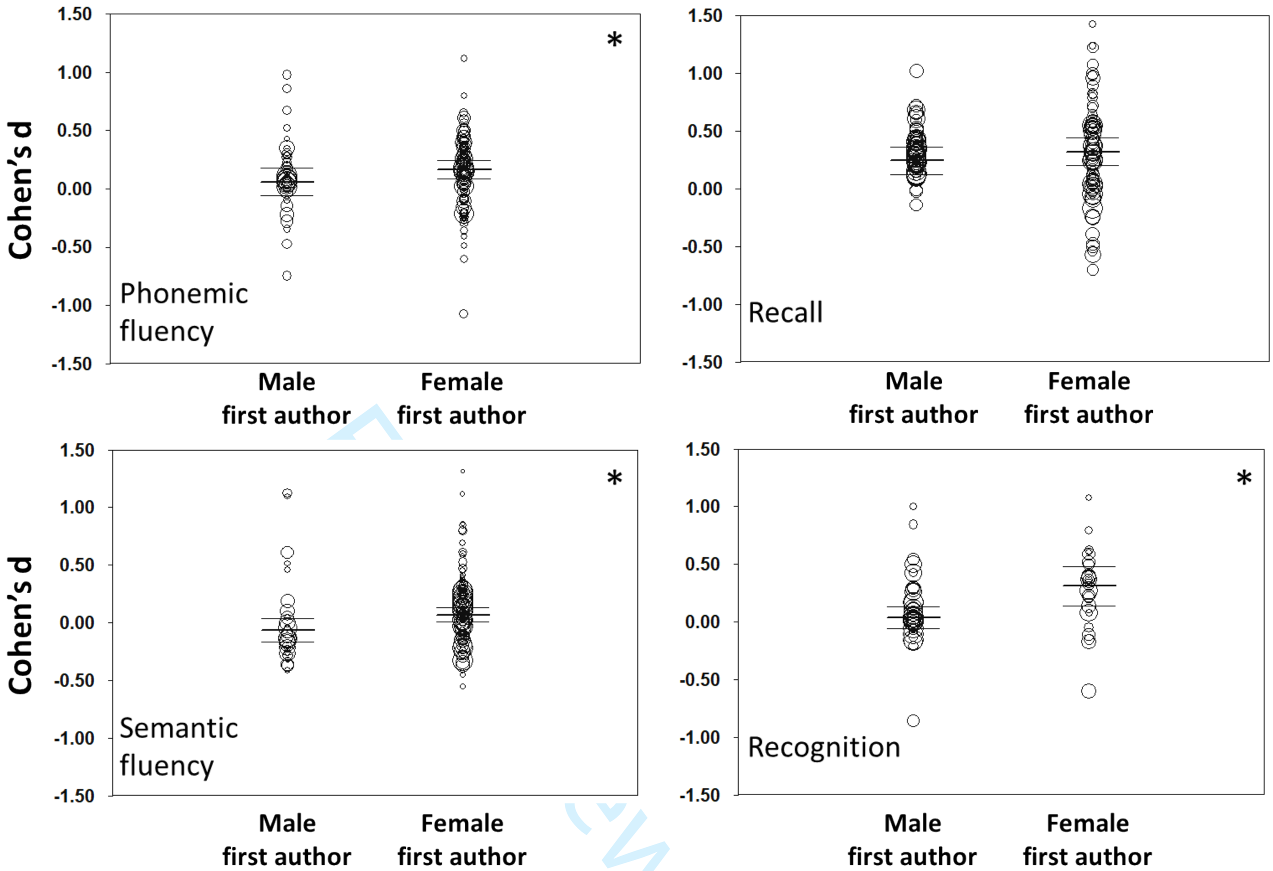
Note. * denotes significant difference between published articles and PhD/Master's theses. Central lines represent means of the respective category, upper and lower lines are confidence intervals. Figures are based on assuming perfect independence between multiple measures from the same (sub-) sample.

Gender of First Author

Female first authors reported significantly stronger female advantages in phonemic fluency ($Z = 2.44$, $p = .015$, $B = 0.107$), semantic fluency ($Z = 3.69$, $p < .001$, $B = 0.134$), and recognition ($Z = 4.31$, $p < .001$, $B = 0.271$) as compared to male first authors, see Figure 3. No significant difference between male and female first authors emerged in recall ($Z = 1.36$, $p = .175$, $B = 0.076$). Assuming perfect correlation, all effects remained significant/non-significant.

Figure 3

Gender of first author effect



Note. * denotes significant difference between female and male first authors. Central lines represent means of the respective category, upper and lower lines are confidence intervals. Figures are based on assuming perfect independence between multiple measures from the same (sub-) sample.

Publication Year

The female advantage significantly decreased in phonemic fluency ($Z = 2.401$, $p = .016$, $B = -0.004$) and recall ($Z = 2.02$, $p = .044$, $B = -0.005$) with publication year. However, the effect would become non-significant in phonemic fluency, if the oldest study (Elias, 1951) was removed ($Z = 1.91$, p = .057, $B = -0.002$). Neither semantic fluency ($Z = 1.63$, $p = .103$, $B = -0.004$) nor recognition ($Z = 1.43$, $p = .152$, $B = -0.004$) changed significantly with publication year, see Supplementary Figure S2. Assuming perfect correlation, the effect in recall would no longer be significant ($Z = 1.73$, $p = .085$, $B = -0.005$) and all other effects remained non-significant (after removing Elias, 1951).

Mean Age

In phonemic fluency, the female advantage became significantly smaller with increasing mean age ($Z = 2.46$, $p = .014$, $B = -0.002$). By contrast, the female advantage became significantly larger with increasing mean age in recall ($Z = 2.07$, $p = .038$, $B = 0.002$). However, the effect was non-significant ($Z = 1.76$, $p = .078$, $B = 0.002$) after removing the study with the oldest mean age sample that also had an unusually high female advantage (Bleecker et al., 1988). No significant mean age effect emerged in semantic fluency ($Z = 1.94$, $p = .052$, $B = -0.001$) and recognition ($Z = 0.05$, $p = .959$, $B < -0.001$), see Supplementary Figure S3. Assuming perfect correlation, the female advantage decreased significantly with age in semantic fluency ($Z = 2.45$, $p = .014$, $B = -0.002$) and increased significantly in recall, also if Bleecker et al., (1988) is removed ($Z = 2.03$, $p = .043$, $B = 0.002$). All other effects remained significant/non-significant.

Age Groups

A new set of meta-regressions was computed which contained "Age groups" and all significant covariates from the first set of meta-regressions described above. "Mean age" was never retained due to multicollinearity with "Age groups".

The results are presented in Table 2. "Age groups" as a whole (i.e., with all age categories combined) only varied significantly in semantic fluency ($Q(4) = 102.6$, $p < .001$, 77 effect sizes). More specifically, the sex/gender difference in "Middle aged" ($Z = 2.01$, $p = .045$, $B = 0.093$), and "Aged" ($Z = 7.65$, $p < .001$, $B = -0.273$) differed significantly from the reference group "Adults". There was no significant difference between "Child/Child Preschool" as well as "Adolescent" with "Adult" (all $Z \leq 1.57$, all $p \geq .117$). Moreover, there were no significant overall effects "Age groups" in phonemic fluency ($Q(4) = 5.49$, $p = .241$, 63 effect sizes), recall ($Q(4) = 7.54$, $p = .110$, 67 effect sizes), and recognition ($Q(4) = 6.85$, $p = .144$, based on 35 effect sizes). In phonemic fluency (all $Z \leq 1.56$, all $p \geq .119$) also none of the individual age groups differed significantly from the reference group “Adult”. In recall, the “Child/child preschool” group had a significantly smaller female advantage than the “Adult” group ($Z = 2.15$, $p = .032$, $B = 0.200$). In recognition, the “Adolescent” ($Z = 2.11$, $p = .035$, $B = 0.275$) and “Child/child preschool” ($Z = 2.05$, $p = .040$, $B = 0.202$) groups had a significantly higher female advantage than the “Adult” reference group, but in case of adolescents this was based on only three effect sizes.

Assuming perfect correlation, all “Age groups” effects in phonemic fluency (63 effect sizes) and semantic fluency (74 effect sizes) remained significant/non-significant. In recall, “Age groups” as a whole remained non-significant, but now only the “Aged” subsample had a significantly smaller female advantage than “Adult” ($Z = 2.30$, $p = .021$, $B = -0.127$, 62 effect sizes). In recognition, “Age groups” as a whole remained non-significant and none of the individual age groups differed significantly from “Adults” (all $Z \leq 1.78$, all $p \geq .075$, 26 effect sizes).

Gender of Last Author

A third set of meta-regressions was computed for published articles only that contained “Last author gender” and all significant covariates from the respective first set of meta-regressions. “Publication Type” was not included due to multicollinearity. “Last author gender” only became significant in semantic fluency ($Z = 2.50$, $p < .001$, $B = -0.09$, 90 effect sizes), where male last authors reported a stronger female advantage than female last authors. No significant differences between male and female last authors emerged in phonemic fluency ($Z = 1.68$, $p = .0093$, $B = 0.087$, 72 effect sizes), recall ($Z = 0.72$, $p = .474$, $B = 0.031$, 70 effect sizes), and recognition ($Z = 0.35$, $p = .729$, $B = -0.021$, 53 effect sizes), see Supplementary Figure S4. Assuming perfect correlation, all effect remained significant/non-significant.

Table 2

Descriptive overview of age group effects

	Phonemic fluency	**Semantic fluency***	Recall	Recognition
Child/Child preschool (≤12 yrs)	*d* = 0.13	*d* = 0.09	***d* = 0.05***	***d* = 0.13***
	(0.06 to 0.25, *k*=29)	(-0.02 to 0.17, *k*=30)	(-0.06 to 0.17, *k*=15)	(-0.04 to 0.31, *k*=7)
Adolescent (13-18 yrs)	*d* = 0.22	*d* = 0.03	*d* = 0.13	***d* = 0.11***
	(0.03 to 0.41, *k*=5)	(-0.25 to 0.30, *k*=2)	(-0.06 to 0.31, *k*=7)	(-0.14 to 0.35, *k*=3)
Adult (19-44 yrs)	*d* = 0.24	*d* = 0.15	*d* = 0.28	*d* = 0.02
	(0.07 to 0.41, *k*=7)	(0.10 to 0.21, *k*=8)	(0.17 to 0.39, *k*=15)	(-0.10 to 0.13, *k*=9)
Middle aged (45-64 yrs)	*d* = 0.13	***d* = 0.25***	*d* = 0.34	*d* = 0.13
	(0.03 to 0.23, *k*=7)	(0.17 to 0.32, *k*=6)	(0.24 to 0.45, *k*=9)	(-0.04 to 0.28, *k*=6)
Aged (≥ 65 yrs)	*d* = 0.06	***d* = -0.10***	*d* = 0.17	*d* = 0.06
	(-0.03 to 0.15, *k*=15)	(-0.14 to -0.07, *k*=31)	(0.09 to 0.24, *k*=21)	(-0.09 to 0.21, *k*=10)

Note. Values in parentheses represent 95% CI and *k* = number of effect sizes included. Individual age groups marked with asterisk/bold print differed significantly from the reference group "Adult". Verbal ability measures marked with asterisk/bold print indicate that the sex/gender difference varied significantly across all age groups. This table may contain more effect sizes than the meta-regression because the meta regression only includes studies with info on all covariates. Figures are based on assuming perfect independence between multiple measures from the same (sub-)sample.

Discussion

Using a meta-analytical approach, we investigated whether women/girls perform better than men/boys in verbal fluency and verbal episodic memory with neutral stimuli that were memorized intentionally and which factors moderated the female advantage.

Small but Robust Female Advantage in Phonemic but not Semantic Fluency

Women/girls performed significantly better in phonemic fluency than men/boys ($d = 0.13$), but there was no significant female advantage in semantic fluency ($d = 0.01$ to 0.02). When combined into a single verbal fluency score, a significant female advantage remained ($d = 0.07$) but more by virtue of the large number of included effect sizes ($k = 290$). The female advantage is thus limited to phonemic fluency, and even here it is markedly lower than in the landmark meta-analysis by Hyde and Linn (1988), who reported a small effect ($d = 0.33$). This discrepancy might be partly due to a different definition of verbal fluency used in the present meta-analysis which also included a much larger number of studies (168 vs. 14), thereby providing higher precision.

The overall effect size for phonemic fluency ($d = 0.12$ to 0.13) is practically identical with both the COWAT/F-A-S ($d = 0.14$), the most frequently used test/starting letter combination, and when generic starting letters or combination of generic starting letters are combined ($d = 0.12$). To illustrate the magnitude of the female advantage: If men/boys report $M = 36$ words, an effect of $d = 0.14$ would translate into an advantage of roughly one and a half words for women/girls ($M = 37.4$), assuming a realistic standard deviation of 10 words.

The large number of studies and effect sizes in the present meta-analysis allowed testing whether the observed sex/gender difference in *semantic fluency* depended on the specific category participants were tasked with. The results revealed that men/boys generally named more "animals" ($d = -0.13$), while women/girls named more "fruits/food/vegetables" ($d = 0.31$). When both categories were combined, which several studies did, the effects size was slightly positive (*d* = 0.11), indicating a slight female advantage. These findings support the view that there is no overall female advantage in semantic fluency and that sex/gender differences are category-dependent (e.g., Laws, 2004; Sokołowski et al., 2020). Category-dependency is also likely to account in part for the enormous heterogeneity in semantic fluency: The proportion of observed variance that reflects difference in true effect sizes (rather than sampling error) was 92 %. Yet, further research is needed to study those categories in more detail.

Small but Robust Female Advantage in Verbal Episodic Memory

We found a significant female advantage for verbal episodic memory, in general, with effect sizes between *d* = 0.23 and *d* = 0.26. Further, the female advantage was stronger in recall (*d* = 0.28) than in recognition (*d* = 0.12 to 0.17). Both findings are in line with Asperholm et al. (2019) who reported an overall female advantage of *g* = 0.28 for episodic memory with verbal content as well as a female advantage for recall (*g* = 0.28 to 0.31) and recognition (*g* = 0.17). Note that the studies included in both meta-analyses had only little overlap, highlighting the robustness of the female advantage. Recognition is generally considered easier than recall (e.g., Postman et al., 1948). Therefore, the female advantage might be smaller in the less difficult recognition tasks.

The strongest female advantage arose for the CVLT (*d* = 0.42) and the RAVLT (*d* = 0.39). By contrast, the two combined tasks “Delayed Memory for Names” and “Visual-Auditory Learning” from the Woodcock Johnson Psycho-Educational Battery – Revised showed a male advantage (*d* = -0.13). However, since all twelve effect sizes were taken from the same study (Cotten, 1991), generalization of these findings is questionable. In recognition, the CVLT (*d* = 0.17) and RAVLT (*d* = 0.22) also demonstrated a female advantage. The only task showing a male advantage (i.e., “Story telling delayed recognition”) (*d* = -0.07) was not significant (confidence bands comprise zero), and again all seven effect sizes were from the same study (Murre et al., 2013). To illustrate the magnitude of the female advantage in verbal episodic memory, imagine a hypothetical study with the CVLT, in which participants need to memorize a list with 16 nouns. Assuming a realistic standard deviation of three words and M =10 for men, Cohen's d = 0.42 (the largest effect size found for verbal episodic memory) translates into a female advantage of roughly one single word (M = 11.26).

While the present meta-analysis, together with Asperholm et al. (2019), suggest a small but robust female advantage for verbal *episodic* memory, Voyer et al. (2021) demonstrated that the female advantage in verbal *working* memory is practically zero. The largest female advantage reported by the authors was g = 0.15 for free recall. This may be due to the fact that certain tasks, which showed a reliable female advantage in the present study, for example the CVLT, were also included in Voyer et al. (2021). The distinction between episodic long-term and working memory is not always clear cut, and there are good arguments why the CVLT taps into both memory processes. In general, however, the findings from all three meta-analyses suggest that the female advantage in verbal memory is not universal and emerges especially when information needs to be transferred to long-term memory while it is very small or absent in working memory.

The Female Advantage is Small But Relevant

By comparison, the female advantage in verbal episodic memory and phonemic fluency is smaller than in other verbal abilities, such as reading achievement (d = 0.23 to 0.68 Reilly, 2012; Stoet & Geary, 2013) or writing abilities (d = 0.53 to 0.61 Hedges & Nowell, 1995). In general, medium to large sex/gender differences were the exception which is in line with the 'gender similarity hypothesis' (Hyde, 2005, 2014), according to which most sex/gender differences are in the small to medium range.

Verbal episodic memory and phonemic fluency tasks are frequently used for assessing psychological impairments (Barker-Collo & Feigin, 2006; Collie & Maruff, 2000; Pennington

& Ozonoff, 1996). Given that the present study corroborates previous findings that standard tests such as CVLT (Kramer et al., 2003), RAVLT (Bleecker et al., 1988) and COWAT (Halari et al., 2005) reliably showed a female advantage, this implies that sex/gender should be taken into account when phonemic fluency and verbal episodic memory are used in the clinical/diagnostic context.

Stronger Female Advantage in Published Articles Than PhD/Master's Theses

We found support for the notion that the female advantage in verbal fluency and verbal episodic memory is subject to publication bias. Firstly, Egger's regression and the funnel plots (Figure S1) suggest a "small study effect" for verbal episodic memory, in general, as well as recall and recognition. That is, especially small studies with significant results favoring women/girls were more likely to be included in our meta-analysis than small studies favoring men/boys. Egger's regression, however, was not significant for verbal, phonemic, or semantic fluency, suggesting the small study effect is generally stronger in verbal episodic memory.

In addition, we found that the female advantage in all four reported verbal abilities was higher in published articles than in PhD/Master's theses. The difference ranged between $d = 0.09$ to 0.39. In fact, for recognition, the female advantage was not significant in PhD/Master's theses. By using meta-regressions, factors such as publication year, age, or first/last author gender were controlled for. Therefore, it is unlikely that the publication type effect was a mere artifact of, for instance, an overrepresentation of unpublished studies in a particular age group. Similarly, the publication bias is unlikely to arise from lower quality in non-peer-reviewed PhD/Master's theses: If this were the case, we would expect *randomly* weaker or larger sex/gender differences. However, we found *consistently* stronger female advantage in published articles. The most parsimonious explanation is therefore that studies are more likely to be published when they find the anticipated female advantage.

First Authors' Gender Impacts Sex/Gender Difference

The meta-regression further revealed that the first author's gender affects the magnitude of the sex/gender difference in phonemic fluency, semantic fluency, and recognition, but not recall. Both male and female first authors consistently reported stronger performance for members of their own gender. The effect was in the range of $d = 0.11$ to 0.27 and controlled for age, publication type, or publication year. Hyde and Linn (1988) reported a similar first author bias but with smaller effect size ($d = 0.07$) and across a wide range of verbal abilities. We speculate that the first author bias represents an in-group bias where members of one's own group are favored over out-group members. Based on these data, it is not possible to disentangle whether female first authors overreport or male first authors underreport the female advantage.

We also found a last author effect in semantic fluency, where male last authors reported a significantly stronger female advantage than female last authors. This result is difficult to interpret, because the sex/gender effect in semantic fluency is category-dependent, as described above. None of the other three measures (i.e., phonemic fluency, recall, and recognition) yielded significant last author effects, and thus we refrain from speculations regarding last author effects in the present study.

No Clear Cohort or Age Effects

The female advantage decreased significantly with publication year for recall (when assuming perfect independence between multiple outcomes), but the effect was small ($B = -0.004$) and did not emerge when assuming perfect correlation. No significant effect was found for recognition (see also Asperholm et al., 2019). Similarly, the significant publication year effect in phonemic fluency disappeared when one outlier was removed. Overall, sex/gender effects reported here were relatively stable over time.

Age effects were neither in line with the previously reported stronger deterioration in older men as compared to older women (Graves et al., 2017; Kramer et al., 2003; Rodriguez-

Aranda & Martinussen, 2006), nor with an inverted U-shaped curve with smaller sex/gender differences in earlier and later life (Asperholm et al., 2019). When the analysis was based on *mean age*, a significant coefficient (B = -0.002) was only found in phonemic fluency, implying that the female advantage was reduced by d = 0.02 over a ten-year period – a small effect. When the analysis was based on *age groups*, none of the three verbal ability measures that showed a reliable female advantage yielded a significant overall age groups effect. In some cases, certain age groups differed significantly from the adult reference group (see Table 3), but most comparisons with adults were not significant. In general, findings for the three measures that yielded a female advantage, indicated relatively stable sex/gender differences throughout life-span (see also de Frias et al., 2006).

Semantic fluency was the only verbal domain showing a significant overall age group effect: Middle aged participants (45-64 years, d = 0.25) showed the strongest female advantage, followed by adults (19-44 years, d = 0.15) and children (2-12 years, d = 0.09). Participants aged 65 or older even showed a significant male advantage (d = -0.10). However, we refrain from interpretations, as the female advantage was strongly category-dependent.

Limitations

Firstly, the statistical indicators showed considerable variance. The null hypothesis, according to which there is only one true underlying effect size, was violated in all analyses. To include data from very heterogeneous samples can be considered an asset, as it increases the generalizability of our findings. However, although we investigated several moderator variables, there are other potentially relevant factors that we did not examine such as (i) specific categories for semantic fluency, (ii) test language, (iii) mono- versus bilingual participants, and (iv) participants' country/region of origin. The fact that most studies were carried out in the US/UK and used native English-speaking participants might hamper generalizability. For example, while the female advantage in reading comprehension emerges in all countries, its magnitude also varies across countries (Reilly, 2012; Stoet & Geary, 2013).

Secondly, we analyzed age effects with two approaches (age means and age groups) that each have their advantages and disadvantages. "Age means" allowed including more effect sizes at the expense of precision, as the single number of age mean becomes meaningless in samples with large age ranges. "Age groups" allowed examining sex/gender differences in clearly defined developmental periods but at the expense of losing effect sizes that do not fall in an age category. As a result, some of the age groups have very few effect sizes (e.g., two or three) and we thus refrained from interpreting too much into significant differences between specific age groups. Conducting those analyses seemed nevertheless justified and the lack of clear age effects may in part be due to the complex nature of sex/gender differences across age.

Thirdly, we contacted authors whose work we had already identified as suitable for our meta-analysis and where only key statistical parameters were missing for calculating effect sizes. We did not reach out to authors who simply used tests/tasks that we considered as adequate, and we also did not contact forums or researchers in the field of verbal fluency/memory. We further only reached out to authors who provided contact details in published articles, which were unavailable for authors of PhD/Master's theses. Moreover, we did not include data from Google Scholar as the massive numbers of reference (more than 200,000) was simply unfeasible to process. Thus, although the present meta-analysis compiled a large body of data, we might have missed several primary studies.

Conclusion and Future Avenues

Based on data from 168 studies, 496 effect sizes, and 355,173 participants, the present meta-analysis suggests that a small but robust female advantage in verbal fluency and verbal episodic memory exists. With respect to verbal fluency, the female advantage only emerged in phonemic fluency, while sex/gender differences in semantic fluency appeared strongly category-dependent. The female advantage, especially in phonemic fluency, is smaller than previously shown (Hyde & Linn, 1988). However, phonemic fluency as well as verbal episodic memory measures are frequently used in psychological/diagnostic settings, highlighting the need for taking sex/gender effects into account. A discussion of how the female advantage arises and what the underlying brain mechanisms are, is beyond the scope of the present meta-analysis, but as argued for other cognitive sex/gender differences, we propose that the female advantage emerges from an intricate interaction of biological, psychological, and socio-cultural factors (Halpern, 2012; Halpern & Tan, 2001; Hausmann, 2017; Jäncke, 2018).

The female advantage is affected by publication bias in two forms: Published articles reported larger female advantages than unpublished research, and both male and female first authors reported better performance for participants of their own gender. While we found evidence for the existence of publication bias, it did not fully account for the female advantage reported here. Similar investigations of first/last authors effects in cognitive abilities in which men/boys typically excel (e.g., mental rotation) have been largely ignored so far.

In general, meta-analyses focusing on cognitive abilities favouring women/girls are rare (for notable exceptions see Asperholm et al., 2019; Voyer et al., 2007; Voyer et al., 2021; Voyer & Voyer, 2014). Apart from including additional factors listed above, future studies should examine whether similar effects are also present in cognitive domains that are assumed to favour men/boys. Finally, more studies should adopt a biopsychosocial approach and include more routinely sex/gender-related, non-binary factors (e.g., sex hormones, self-efficacy, gender stereotypes), and their interactions, that might explain individual differences in verbal abilities and other cognitive domains better than sex/gender.

Acknowledgments

We would like to thank Kylie Wong and Emily George for their tremendous help with the screening of references.

References

References marked with * are included in the meta-analysis.

*Abdel Aziz, K., Khater, M. S., Emara, T., Tawfik, H. M., Rasheedy, D., Mohammedin, A. S., Tolba, M. F., El-Gabry, D. A., & Qassem, T. (2017, Jul-Aug). Effects of age, education, and gender on verbal fluency in healthy adult Arabic-speakers in Egypt. *Appl Neuropsychol Adult, 24*(4), 331-341. https://doi.org/10.1080/23279095.2016.1185424

*Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., Hurwitz, D. A., Aguero, H., Greenfield, L., & Duara, R. (2000). Category Fluency Test: Normative data for English- and Spanish-speaking elderly. *Journal of the International Neuropsychological Society, 6*(7), 760-769. https://doi.org/undefined

*Agard, C. N. (2008). *Urine creatinine levels and neurocognitive functioning in African-American adults* (Publication Number 3335300) [Ph.D., Howard University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Alexander, J. R. M., & Smales, S. (1997, 1997/11/01/). Intelligence, learning and long-term memory. *Personality and Individual Differences, 23*(5), 815-825. https://doi.org/https://doi.org/10.1016/S0191-8869(97)00054-8

*Alexiou, T. (2005). *Cognitive development, aptitude and language learning in Greek young learners* (Publication Number 10798115) [Ph.D., Swansea University (United Kingdom)]. ProQuest Dissertations & Theses Global. Ann Arbor.

Andreano, J. M., & Cahill, L. (2009, April 1, 2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory, 16*(4), 248-266. https://doi.org/10.1101/lm.918309

*Ardila, A., Rosselli, M., Matute, E., & Guajardo, S. (2005, 2005/08/01). The Influence of the Parents' Educational Level on the Development of Executive Functions. *Developmental Neuropsychology, 28*(1), 539-560. https://doi.org/10.1207/s15326942dn2801_5

Asperholm, M., Hogman, N., Rafi, J., & Herlitz, A. (2019, Aug). What Did You Do Yesterday? A Meta-Analysis of Sex Differences in Episodic Memory [Article]. *Psychological Bulletin, 145*(8), 37. https://doi.org/10.1037/bul0000197

*Auriacombe, S., Fabrigoule, C., Lafont, S., Jacqmin-Gadda, H., & Dartigues, J.-F. (2010). Letter and Category Fluency in Normal Elderly Participants: a Population-Based Study. *Aging, Neuropsychology, and Cognition, 8*(2), 98-108. https://doi.org/10.1076/anec.8.2.98.841

Baldo, J. V., Schwartz, S., Wilkins, D., & Dronkers, N. F. (2006, Nov). Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping [Article]. *Journal of the International Neuropsychological Society, 12*(6), 896-900. https://doi.org/10.1017/s1355617706061078

*Banks, P. G., Dickson, A. L., & Plasay, M. T. (1987, Winter). The verbal selective reminding test: preliminary data for healthy elderly. *Exp Aging Res, 13*(4), 203-207. https://doi.org/10.1080/03610738708259326

Barker-Collo, S., & Feigin, V. (2006, Jun). The impact of neuropsychological deficits on functional stroke outcomes [Review]. *Neuropsychology Review, 16*(2), 53-64. https://doi.org/10.1007/s11065-006-9007-5

*Baxter, L. C. (1998). *Dual-Task Interference Effects on Sex Differences in Verbal Memory* Linguistics and Language Behavior Abstracts (LLBA).

Benton, A. L. (1967). Problems of test construction in the field of aphasia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 3*(1), 32-58. https://doi.org/10.1016/S0010-9452(67)80005-4

*Blair, A. S. (2002). *The effects of acculturation level on verbal learning in a sample of Hispanics of Mexican-American extraction* (Publication Number 3074024) [Psy.D., George Fox University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Bleecker, M. L., Bolla-Wilson, K., Agnew, J., & Meyers, D. A. (1988). Age-related sex differences in verbal memory. *Journal of Clinical Psychology, 44*(3), 403-411. https://doi.org/10.1002/1097-4679(198805)44:3<403::AID-JCLP2270440315>3.0.CO;2-0

*Bolla-Wilson, K., & Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age, and education on the rey auditory verbal learning test. *Developmental Neuropsychology, 2*(3), 203-211. https://doi.org/10.1080/87565648609540342

*Bolla, K. I., Gray, S., Resnick, S. M., Galante, R., & Kawas, C. (1998, 1998/08/01). Category and Letter Fluency in Highly Educated Older Adults. *The Clinical Neuropsychologist, 12*(3), 330-338. https://doi.org/10.1076/clin.12.3.330.1986

*Bolla, K. I., Lindgren, K. N., Bonaccorsy, C., & Bleecker, M. L. (1990, Sep). Predictors of verbal fluency (FAS) in the healthy elderly. *J Clin Psychol, 46*(5), 623-628. https://doi.org/10.1002/1097-4679(199009)46:5<623::aid-jclp2270460513>3.0.co;2-c

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2014). *Comprehensive Meta-analysis.* In Biostat.

Brand, N., & Jolles, J. (1985). Learning and Retrieval Rate of Words Presented Auditorily and Visually. *The Journal of general psychology, 112*(2), 201-210.

*Brandling-Bennett, E. M. (2006). *Categorization during typical development* (Publication Number 3238634) [Ph.D., Washington University in St. Louis]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Brocki, K. C., & Bohlin, G. (2004, 2004/10/01). Executive Functions in Children Aged 6 to 13: A Dimensional and Developmental Study. *Developmental Neuropsychology, 26*(2), 571-593. https://doi.org/10.1207/s15326942dn2602_3

*Brosnan, M. D. (1973). *Developmental changes in memory processes for verbal and Pictorial material* (Publication Number 7415125) [Ph.D., Purdue University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Brownell, W., & Smith, D. R. (1973). Communication patterns, sex, and length of verbalization in speech of four-year-old children. *Speech Monographs, 40*(4), 310-316. https://doi.org/10.1080/03637757309375809

*Brucki, S. M., & Rocha, M. S. (2004, Dec). Category fluency test: effects of age, gender and education on total scores, clustering and switching in Brazilian Portuguese-speaking subjects. *Braz J Med Biol Res, 37*(12), 1771-1777. https://doi.org/10.1590/s0100-879x2004001200002

*Burton, L. A., & Henninger, D. (2013). Sex differences in relationships between verbal fluency and personality. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 32*(2), 168-174. https://doi.org/10.1007/s12144-013-9167-4

*Capitani, E., Laiacona, M., & Basso, A. (1998, 1998/01/01/). Phonetically Cued Word-Fluency, Gender Differences and Aging: A Reappraisal. *Cortex, 34*(5), 779-783. https://doi.org/https://doi.org/10.1016/S0010-9452(08)70781-0

*Carstairs, J. R., Shores, E. A., & Myors, B. (2012). Australian norms and retest data for the Rey Auditory and Verbal Learning Test. *Australian Psychologist, 47*(4), 191-197. https://doi.org/10.1111/j.1742-9544.2012.00086.x

Catani, M., Allin, M. P. G., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., & Jones, D. K. (2007, Oct). Symmetries in human brain language pathways correlate with verbal recall [Article]. *Proc Natl Acad Sci U S A, 104*(43), 17163-17168.

*Cerhan, J. R., Folsom, A. R., Mortimer, J. A., Shahar, E., Knopman, D. S., McGovern, P. G., Hays, M. A., Crum, L. D., & Heiss, G. (1998). Correlates of Cognitive Function in Middle-Aged Adults. *Gerontology, 44*(2), 95-105. https://doi.org/10.1159/000021991

*Chan, A. S., & Poon, M. W. (1999, Sep). Performance of 7- to 95-year-old individuals in a Chinese version of the category fluency test. *J Int Neuropsychol Soc, 5*(6), 525-533. https://doi.org/10.1017/s135561779956606x

*Chan, R. C. K., Wong, M., Chen, E. Y. H., & Lam, L. C. W. (2003). Semantic categorisation and verbal fluency performance in a community population in Hong Kong: A preliminary report. *Hong Kong Journal of Psychiatry, 13*(4).

*Chang, Y. J. (1992). *P300 event-related potential in normal young adults: Correlation with memory and imagery variables* (Publication Number 1351382) [M.A., California State University, Long Beach]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Chipman, K. A. (1998). *No sex difference on incidental picture memory, despite better verbal memory in women* (Publication Number MQ32474) [M.A., The University of Western Ontario (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Cohen, D. L. (1975). *Sex differences in the organization of spatial abilities in older men and women* [Ph.D., University of Southern California]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Cole, G. C. (1991). *Alexithymia, Stroop interference, and verbal abilities: Sex differences* [Ph.D., Simon Fraser University (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.

Collie, A., & Maruff, P. (2000, 2000/05/01/). The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neuroscience & Biobehavioral Reviews, 24*(3), 365-374. https://doi.org/https://doi.org/10.1016/S0149-7634(00)00012-9

*Comilang, K. A. (2003). *The influence of internalized cultural and racial socialization on the neuropsychological test performance of Asian Americans* (Publication Number 3103275) [Ph.D., Boston College]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Contador, I., Almondes, K., Fernandez-Calvo, B., Boycheva, E., Puertas-Martin, V., Benito-Leon, J., & Bermejo-Pareja, F. (2016, Dec 1). Semantic Verbal Fluency: Normative Data in Older Spanish Adults From NEDICES Population-Based Cohort. *Arch Clin Neuropsychol, 31*(8), 954-962. https://doi.org/10.1093/arclin/acw071

*Contador, I., Almondes, K., Fernandez-Calvo, B., Boycheva, E., Puertas-Martin, V., Benito-Leon, J., & Bermejo-Pareja, F. (2016, Dec 1). Semantic Verbal Fluency: Normative Data in Older Spanish Adults From NEDICES Population-Based Cohort. *Arch Clin Neuropsychol, 31*(8), 954-962. https://doi.org/10.1093/arclin/acw071

*Corona-LoMonaco, M. E. (2000). *Impact of language and culture on a neuropsychological screening battery for Hispanics* [Ph.D., University of Southern California]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Cory, J. M. (2003). *Sociocultural influences in neuropsychological testing with Latinos of Mexican origin* (Publication Number 3107073) [Ph.D., Colorado State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Cotten, M. A. (1991). *Gender differences on the Woodcock-Johnson Revised Tests of Cognitive Abilities and Tests of Achievement* [Ph.D., Texas Woman's University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Crossley, M., D'Arcy, C., & Rawson, N. S. (1997, Feb). Letter and category fluency in community dwelling Canadian seniors: a comparison of normal participants to those with dementia of the Alzheimer or vascular type. *J Clin Exp Neuropsychol, 19*(1), 52-62. https://doi.org/10.1080/01688639708403836

*Dadin, C. O., Salgado, R., & Fernandez, A. (2009). Ciclos naturales de las hormonas sexuales y diferencias entre sexos en memoria. *Actas Espanolas de Psiquiatria, 37*(2), 68--74.

*de Frias, C. M., Nilsson, L.-G., & Herlitz, A. (2006, 2006/12/01). Sex Differences in Cognition are Stable Over a 10-Year Period in Adulthood and Old Age. *Aging, Neuropsychology, and Cognition, 13*(3-4), 574-587. https://doi.org/10.1080/13825580600678418

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test, 2nd edition*. The Psychological Corporation.

*DeWan, L. K. (2006). *Childhood developmental trends in executive function as measured by the Delis -Kaplan Executive Function System: An exploration of gender differences* [Psy.D., George Fox University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Dias, N. M., Menezes, A., & Seabra, A. G. (2013). Age differences in executive functions within a sample of Brazilian children and adolescents. *Span J Psychol, 16*, E9.
https://doi.org/10.1017/sjp.2013.12

*Egelko, S. E. (1983). *Cognitive sequelae of right cerebrovascular accident: issues of Verbal deficit and sex differential patterns in visuospatial and verbal performance* (Publication Number 8326686) [Ph.D., Fordham University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical research ed.), 315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

*Elias, J. Z. (1951). *Non-intellective factors in certain intelligence and achievement tests: an analysis of factors in addition to the cognitive entering into the intelligence and achievement scores of children at the sixth grade level* (Publication Number 0002503) [Ph.D., New York University]. ProQuest Dissertations & Theses Global. Ann Arbor.
Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*(2), 95-103. https://doi.org/10.1037/0003-066X.43.2.95
*Ernest, C. H. (1983). Imagery and verbal ability and recognition memory for pictures and words in males and females. *Educational Psychology, 3*(3-4), 227-244. https://doi.org/10.1080/0144341830030307
*Farace, E. (1996). *Gender differences in relationships between degree of brain lateralization and cognitive ability* (Publication Number 9701315) [Ph.D., University of Virginia]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Fares, R. (2011). *Normative data of two measures of verbal fluency for Arabic/English bilinguals* Linguistics and Language Behavior Abstracts (LLBA).
*Faul, J. D. (2008). *The effect of lifecourse socioeconomic position and health on trajectories of cognitive function in older adults* (Publication Number 3304967) [Ph.D., University of Michigan]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.
*Findlay, L., Bernier, J., Tuokko, H., Kirkland, S., & Gilmour, H. (2010, Dec). Validation of cognitive functioning categories in the Canadian Community Health Survey--Healthy Aging. *Health Rep, 21*(4), 85-100. https://www.ncbi.nlm.nih.gov/pubmed/21269015
*Fraser, D. A. (1986). *Identifying effective memory strategies: a human experiment in Information science (mnemonics, cognition)* (Publication Number 8623521) [D.L.S., Columbia University]. ProQuest Dissertations & Theses Global. Ann Arbor.
Gaillard, W. D., Sachs, B. C., Whitnah, J. R., Ahmad, Z., Balsamo, L. M., Petrella, J. R., Braniecki, S. H., McKinney, C. M., Hunter, K., Xu, B., & Grandin, C. B. (2003, Mar). Developmental aspects of language processing: fMRI of verbal fluency in children and adults [Article]. *Human Brain Mapping, 18*(3), 176-185. https://doi.org/10.1002/hbm.10091
*Gates, J. K. (1986). *The relation of hemisphere asymmetry and cognitive ability in young children* (Publication Number T-29774) [Ph.D., The University of Chicago]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Gavin, M. R. (1988). *The performance of healthy elderly adults on behavioral tasks associated with frontal lobe functioning* [Ph.D., University of Georgia]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Gawda, B., & Szepietowska, E. M. (2013a). Impact of Unconscious Emotional Schemata on Verbal Fluency - Sex Differences and Neural Mechanisms [Article]. *Neuroquantology, 11*(3), 443-450.
*Gawda, B., & Szepietowska, E. M. (2013b, Aug). Semantic and affective verbal fluency: sex differences. *Psychol Rep, 113*(1), 1258-1268.
*Gerdeman, E. M. (1975). *The contribution of social intelligence to predictive accuracy of interpersonal perception* (Publication Number 7514509) [Ph.D., Loyola University Chicago]. ProQuest Dissertations & Theses Global. Ann Arbor.
*González, H. M., Mungas, D., & Haan, M. N. (2005). A semantic verbal fluency test for English- and Spanish-speaking older Mexican-Americans. *Archives of Clinical Neuropsychology, 20*(2), 199-208. https://doi.org/10.1016/j.acn.2004.06.001
*Gould, K. L. (1972). *Relationships of creativity, reading comprehension, intelligence, and response to a literature selection for fourth grade inner-city children* [Ph.D., The Ohio State University]. ProQuest Dissertations & Theses Global. Ann Arbor.
Graves, L. V., Moreno, C. C., Seewald, M., Holden, H. M., Van Etten, E. J., Uttarwar, V., McDonald, C. R., Delano-Wood, L., Bondi, M. W., Woods, S. P., Delis, D. C., & Gilbert, P. E. (2017). Effects of Age and Gender on Recall and Recognition Discriminability. *Archives of Clinical Neuropsychology, 32*(8), 972-979. https://doi.org/10.1093/arclin/acx024
*Greenstein, Y., Blachstein, H., & Vakil, E. (2010). Interrelations between attention and verbal memory as affected by developmental age. *Child neuropsychology : a journal on normal*

and abnormal development in childhood and adolescence, 16(1), 42-59. https://doi.org/10.1080/09297040903066891

*Gregory, A. M. (2002). *Gender differences in the encoding of occupations: A release from proactive interference method in an aural-verbal modality* (Publication Number 1409474) [M.S., California State University, Fullerton]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Greshner, C. L. (2000). *Relationships between physical and psychological measures of masculinity, femininity, and androgyny, and performance on sexually dimorphic cognitive tests* (Publication Number MQ61434) [M.A., Simon Fraser University (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., Siegel, S. J., & Gur, R. E. (2001, Nov). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology, 25*(5), 766-776. https://doi.org/10.1016/S0893-133X(01)00278-0

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005, Feb). Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins [Article]. *Behavioral Neuroscience, 119*(1), 104-117. https://doi.org/10.1037/0735-7044.119.1.104

*Halari, R. (2003). *The relationship between gonadal hormones and neurocognitive functioning in healthy men and women and patients with schizophrenia* [Ph.D., The City University (London) (United Kingdom)]. ProQuest Dissertations & Theses Global. Ann Arbor.

Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). Psychology Press, Taylor and Francis Group.

Halpern, D. F., & Tan, U. (2001, Apr). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45*(3), 392-414.

Hamson, D. K., Roes, M. M., & Galea, L. A. M. (2016, Jul). Sex Hormones and Cognition: Neuroendocrine Influences on Memory and Learning [Article]. *Comprehensive Physiology, 6*(3), 1295-1337. https://doi.org/10.1002/cphy.c150031

Harris, M. B., & Seibel, C. E. (1976). Effects of sex, occupation, and confidence of model and sex and grade of subject on imitation of language behaviors. *Developmental Psychology, 12*(1), 89-90. https://doi.org/10.1037/0012-1649.12.1.89

*Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000, Jun). Short test of semantic and phonological fluency: normal performance, validity and test-retest reliability. *Br J Clin Psychol, 39 (Pt 2)*, 181-191.

Hausmann, M. (2017). Why sex hormones matter for neuroscience: A very short review on sex, sex hormones, and functional brain asymmetries. *Journal of Neuroscience Research, 95*(1-2), 40-49. https://doi.org/10.1002/jnr.23857

*Hausmann, M., Schoofs, D., Rosenthal, H. E., & Jordan, K. (2009, Apr). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences-A psychobiosocial approach. *Psychoneuroendocrinology, 34*(3), 389-401.

*Havlena, J. E. (1990). *The influence of age, gender, and motivation on the development of memory strategy use* (Publication Number 1341726) [M.A., California State University, Fullerton]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Hazin, I., Leite, G., Oliveira, R. M., Alencar, J. C., Fichman, H. C., Marques, P. d. N., & de Mello, C. B. (2016, 2016-May-10). Brazilian Normative Data on Letter and Category Fluency Tasks: Effects of Gender, Age, and Geopolitical Region [Original Research]. *Frontiers in Psychology, 7*(684). https://doi.org/10.3389/fpsyg.2016.00684

Hedges, L. V., & Nowell, A. (1995, Jul). Sex-differences in mental test-scores, variability, and numbers of high-scoring individuals [Article]. *Science, 269*(5220), 41-45. https://doi.org/10.1126/science.7604277

*Heister, G. (1982, January 01). Sex differences in verbal fluency: a short note [journal article]. *Current Psychology, 2*(1), 257-260. https://doi.org/10.1007/bf03186768

*Held, M. M. (2013). *The effects of gender, culture, and acculturation on verbal fluency in Filipino Americans and Caucasian European Americans* [Ph.D., Palo Alto University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Hendrawan, D., Hatta, T., & Ohira, H. (2015, Mar). Do the letters F, A and S represent Indonesian letter fluency stimuli? *Asia Pac Psychiatry, 7*(1), 64-71. https://doi.org/10.1111/appy.12082

*Herlitz, A., Airaksinen, E., & Nordström, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology, 13*(4), 590-597. https://doi.org/10.1037/0894-4105.13.4.590

*Herlitz, A., Nilsson, L. G., & Backman, L. (1997, Nov). Gender differences in episodic memory [Article]. *Memory & Cognition, 25*(6), 801-811. https://doi.org/10.3758/bf03211324

*Herlitz, A., Reuterskiöld, L., Lovén, J., Thilers, P. P., & Rehnman, J. (2013, 2013/04/01). Cognitive Sex Differences Are Not Magnified as a Function of Age, Sex Hormones, or Puberty Development During Early Adolescence. *Developmental Neuropsychology, 38*(3), 167-179. https://doi.org/10.1080/87565641.2012.759580

Hirnstein, M., Hugdahl, K., & Hausmann, M. (2019). Cognitive sex differences and hemispheric asymmetry: A critical review of 40 years of research. *Laterality: Asymmetries of Body, Brain and Cognition, 24*(2), 204-252. https://doi.org/10.1080/1357650X.2018.1497044

Hirnstein, M., Larøi, F., & Laloyaux, J. (2018, July 02). No sex difference in an everyday multitasking paradigm [journal article]. *Psychological Research, 83*(2), 286-296. https://doi.org/10.1007/s00426-018-1045-0

*Hirnstein, M., Coloma Andrews, L., & Hausmann, M. (2014, Jun 13). Gender-Stereotyping and Cognitive Sex Differences in Mixed- and Same-Sex Groups. *Archives of Sexual Behavior, 43*(8), 1663-1673. https://doi.org/10.1007/s10508-014-0311-5

*Hirnstein, M., Freund, N., & Hausmann, M. (2012, 2012). Gender Stereotyping Enhances Verbal Fluency Performance in Men (and Women). *Zeitschrift Fur Psychologie-Journal of Psychology, 220*(2), 70-77. https://doi.org/10.1027/2151-2604/a000098

*Holland, L. J. (1969). *Profile types of the urban junior high school student: utilizing measuresof cognitive factors, achievement, aptitude, and background information* [Ph.D., Wayne State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Hurks, P. P. (2013). Administering design fluency tests in school-aged children: analyses of design productivity over time, clustering, and switching. *Clin Neuropsychol, 27*(7), 1131-1149. https://doi.org/10.1080/13854046.2013.821170

Hyde, J. S. (2005, Sep). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology, 65*(1), 373-398. https://doi.org/doi:10.1146/annurev-psych-010213-115057

Hyde, J. S., & Linn, M. C. (1988, Jul). Gender differences in verbal-ability - a meta-analysis [Review]. *Psychological Bulletin, 104*(1), 53-69.

*Isomura, A. J. (2002). *Verbal and spatial explicit memory performance among Japanese Americans and European Americans: Gender and ethnic differences* (Publication Number 3055184) [Ph.D., Pacific Graduate School of Psychology]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Iverson, G. L., Brooks, B. L., & Ashton Rennison, V. L. (2014). Minimal gender differences on the CNS vital signs computerized neurocognitive battery. *Appl Neuropsychol Adult, 21*(1), 36-42. https://doi.org/10.1080/09084282.2012.721149

*John, S., & Rajashekhar, B. (2014). Word retrieval ability on semantic fluency task in typically developing Malayalam-speaking children. *Child Neuropsychology, 20*(2), 182-195.

*John, S., Rajashekhar, B., & Guddattu, V. (2016, 2016/10/01). Word Retrieval Ability on Phonemic Fluency in Typically Developing Children. *Applied Neuropsychology: Child, 5*(4), 252-263. https://doi.org/10.1080/21622965.2015.1050099

*Johnston, J. O. (1965). *Relationships between intelligence and personality variables* [Ph.D., Oklahoma State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Jordan, L. M. (2014). *Verbal fluency: Norms for the Lakota population in semantic and phonemic fluency tasks* [M.A., The University of North Dakota]. ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

Jäncke, L. (2018). Sex/gender differences in cognition, neurophysiology, and neuroanatomy *F1000Research, 7*, 805. http://openr.es/boa

*Kavé, G. (2005, 2005/08/01). Phonemic Fluency, Semantic Fluency, and Difference Scores: Normative Data for Adult Hebrew Speakers. *Journal of Clinical and Experimental Neuropsychology, 27*(6), 690-699. https://doi.org/10.1080/138033990490918499

*Kemmotsu, N. (2010). *Performance of Japanese Americans on selected cognitive instruments* (Publication Number 3412246) [Ph.D., University of California, San Diego and San Diego State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Kempler, D., Teng, E. L., Dick, M., Taussig, I. M., & Davis, D. S. (1998, Nov). The effects of age, education, and ethnicity on verbal fluency. *J Int Neuropsychol Soc, 4*(6), 531-538.

*Kesse-Guyot, E., Andreeva, V. A., Jeandel, C., Ferry, M., Hercberg, S., & Galan, P. (2012, May). A healthy dietary pattern at midlife is associated with subsequent cognitive performance. *J Nutr, 142*(5), 909-915. https://doi.org/10.3945/jn.111.156257

*Khalil, M. S. (2010, Nov). Preliminary Arabic normative data of neuropsychological tests: the verbal and design fluency. *J Clin Exp Neuropsychol, 32*(9), 1028-1035. https://doi.org/10.1080/13803391003672305

*Kim, J. K., & Kang, Y. (1999, Aug). Normative study of the Korean-California Verbal Learning Test (KCVLT). *Clin Neuropsychol, 13*(3), 365-369. https://doi.org/10.1076/clin.13.3.365.1740

Kimura, D. (2000). *Sex and cognition*. MIT Press.

*Kimura, D., & Clarke, P. G. (2002, 2002/12/01). Women's Advantage on Verbal Memory is Not Restricted to Concrete Words. *Psychological Reports, 91*(3_suppl), 1137-1142. https://doi.org/10.2466/pr0.2002.91.3f.1137

*Kimura, D., & Seal, B. N. (2003). Sex Differences in Recall of Real or Nonsense Words. *Psychological Reports, 93*(1), 263-264. https://doi.org/10.2466/pr0.2003.93.1.263

*Knaus, T. A. (2003). *Sex differences in the anatomy of human perisylvian regions: Frontal and temporal cortical language areas* (Publication Number 3084118) [Ph.D., Tulane University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2006, Sep). Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the minimental state examination, the trail making test and semantic fluency measures. *Br J Clin Psychol, 45*(Pt 3), 393-402. https://doi.org/10.1348/014466505x68032

*Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004, Mar). The verbal fluency task in the Greek population: normative data, and clustering and switching strategies. *J Int Neuropsychol Soc, 10*(2), 164-172. https://doi.org/10.1017/S1355617704102014

*Kramer, J. H., Delis, D. C., Kaplan, E., Odonnell, L., & Prifitera, A. (1997, Oct). Developmental sex differences in verbal learning [Article]. *Neuropsychology, 11*(4), 577-584. https://doi.org/10.1037/0894-4105.11.4.577

*Kramer, J. H., Yaffe, K., Lengenfelder, J., & Delis, D. C. (2003, Jan). Age and gender interactions on verbal memory performance [Article]. *Journal of the International Neuropsychological Society, 9*(1), 97-102. https://doi.org/10.1017/s1355617703910113

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals [doi: 10.1016/j.neuropsychologia.2011.12.022]. *Neuropsychologia, 50*(7), 1211-1221. http://www.sciencedirect.com/science/article/pii/S0028393212000024

Kraan, C., Stolwyk, R., & Testa, R. (2013). *The Abilities Associated With Verbal Fluency Performance in a Young, Healthy Population Are Multifactorial and Differ Across Fluency Variants* (Vol. 20). https://doi.org/10.1080/09084282.2012.670157

*LaFevor, M. E. (2017). *Examining Concurrent Validity, Reliability, and Sex and Age Normative Values of the Impact Quick Test-Pediatric Version* (Publication Number 10272206) [Ph.D., Michigan State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Lanting, S., Haugrud, N., & Crossley, M. (2009). The effect of age and sex on clustering and switching during speeded verbal fluency tasks. *Journal of the International Neuropsychological Society, 15*(2), 196-204. https://doi.org/10.1017/S1355617709090237

*Larrabee, G. J., & Crook, T. H. (1993). Do men show more rapid age-associated decline in simulated everyday verbal memory than do women? *Psychology and Aging, 8*(1), 68-71. https://doi.org/10.1037/0882-7974.8.1.68

*Laws, K. R. (2004). Sex differences in lexical size across semantic categories. *Personality and Individual Differences, 36*(1), 23-32. https://doi.org/10.1016/S0191-8869(03)00048-5

*Lero, D. S. (1974). *The effects of timed and untimed assessment on creativity test performance* [Ph.D., Purdue University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15*(2), 165-173. https://doi.org/10.1037/0894-4105.15.2.165

*Liang, H. (2013). *Cognitive Dysfunction and Mental Health Status in Ketamine and Poly-drug Abusers* (Publication Number 3578875) [Ph.D., The Chinese University of Hong Kong (Hong Kong)]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Lindblad, S. G. (1995). *Gender, age, and level of achievement differences on the Woodcock-Johnson revised tests of cognitive abilities and tests of achievement - early development battery* [Ph.D., Texas Woman's University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex-differences in spatial ability - a meta-analysis [Article]. *Child Development, 56*(6), 1479-1498.

*Lowe, C. N. (1998). *The association between frontal lobe development and social perspective taking in children* [M.A., Mount Saint Vincent University (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.

Lowe, P. A., Mayfield, J. W., & Reynolds, C. R. (2003, 12//). Gender differences in memory test performance among children and adolescents. *Archives of Clinical Neuropsychology, 18*(8), 865-878. https://doi.org/http://dx.doi.org/10.1016/S0887-6177(02)00162-2

*Lundervold, A. J., Wollschlager, D., & Wehling, E. (2014, Jun). Age and sex related changes in episodic memory function in middle aged and older adults. *Scand J Psychol, 55*(3), 225-232. https://doi.org/10.1111/sjop.12114

*Mathuranath, P. S., George, A., Cherian, P. J., Alexander, A., Sarma, S. G., & Sarma, P. S. (2003, Dec). Effects of age, education and gender on verbal fluency [Article]. *Journal of Clinical and Experimental Neuropsychology, 25*(8), 1057-1064.

*Maylor, E. A., Reimers, S., Choi, J., Collaer, M. L., Peters, M., & Silverman, I. (2007, April 01). Gender and Sexual Orientation Differences in Cognition Across Adulthood: Age Is Kinder to Women than to Men Regardless of Sexual Orientation [journal article]. *Archives of Sexual Behavior, 36*(2), 235-249. https://doi.org/10.1007/s10508-006-9155-y

*McKay, K. E. (1995). *Implicit and explicit recognition memory function: A developmental study of normal, reading disabled and ADHD children and adults* (Publication Number 9605632) [Ph.D., City University of New York]. ProQuest Dissertations & Theses Global. Ann Arbor.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*(1), 37-45. https://doi.org/http://dx.doi.org/10.1016/j.tics.2013.10.011

*Mohd, S. T. (1997). *Assessing reading-related skills in Arabic-speaking children* (Publication Number 9801124) [Ph.D., University of Florida]. ProQuest Dissertations & Theses Global. Ann Arbor.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The Prisma Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Moulden, D. J. A. (1992). *Development of the dichotic right ear advantage (REA) for the inferred lateralization of language: Dichotic word listening norms* [M.A., Laurentian University (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Mullins, M. R. (1977). *The relationship of verbal abilities to cognitive complexity* [M.A., University of Nebraska at Omaha]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Munnelly, M. (2016). *Gender differences in verbal and visual memory* [M.A., Kean University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Murre, J. M. J., Janssen, S. M. J., Rouw, R., & Meeter, M. (2013, 2013/01/01/). The rise and fall of immediate and delayed memory for verbal and visuospatial information from late

childhood to late adulthood. *Acta Psychologica, 142*(1), 96-107. https://doi.org/https://doi.org/10.1016/j.actpsy.2012.10.005
*Nida, R. E. (1986). *A comparative investigation of young children's recall memory proficiency in naturalistic and laboratory settings* (Publication Number 8718680) [Ph.D., The University of North Carolina at Greensboro]. ProQuest Dissertations & Theses Global. Ann Arbor.
*O'Hara, R., Miller, E., Liao, C. P., Way, N., Lin, X., & Hallmayer, J. (2006, Dec 6). COMT genotype, gender and cognition in community-dwelling, older adults. *Neurosci Lett, 409*(3), 205-209. https://doi.org/10.1016/j.neulet.2006.09.047
Parsons, T. D., Rizzo, A. R., Zaag, C. v. d., McGee, J. S., & Buckwalter, J. G. (2005, 2005/01/01). Gender Differences and Cognition Among Older Adults. *Aging, Neuropsychology, and Cognition, 12*(1), 78-88. https://doi.org/10.1080/13825580590925125
*Pedersen, L. L. (2005). *The relationship between behavioral and performance-based measures of executive function in preschool children* (Publication Number 3168583) [Ph.D., Texas Woman's University]. ProQuest Dissertations & Theses Global. Ann Arbor.
Pennington, B. F., & Ozonoff, S. (1996, Jan). Executive functions and developmental psychopathology [Review]. *Journal of Child Psychology and Psychiatry, 37*(1), 51-87. https://doi.org/10.1111/j.1469-7610.1996.tb01380.x
*Phillips, B. L. (1977). *A comparison of the McCarthy Scales of Children's Abilities to the WPPSI and the Columbia Mental Maturity Scale* (Publication Number EP16129) [M.S., University of Wyoming]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Pino Escobar, G. M. (2017). *Do bilingual and monolingual children differ? Measuring and comparing attentional control skills in the verbal and non-verbal domains* [M.Phil., Western Sydney University (Australia)]. ProQuest Dissertations & Theses Global. Ann Arbor.
Postman, L., Jenkins, W. O., & Postman, D. L. (1948). An experimental comparison of active recall and recognition. *The American Journal of Psychology, 61*, 511-519. https://doi.org/10.2307/1418315
*Prigatano, G. P., Gray, J. A., & Lomay, V. T. (2008). Verbal (animal) fluency scores in age/grade appropriate minority children from low socioeconomic backgrounds. *Journal of the International Neuropsychological Society, 14*(1), 143-147.
https://doi.org/10.1017/S1355617708080089
*Rae, G. (1979). *The role of auditory and visual sensory modalities in reading and verbal memory* (Publication Number U444206) [Ph.D., University of Aberdeen (United Kingdom)]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Rahman, Q., Abrahams, S., & Wilson, G. D. (2003). Sexual-orientation-related differences in verbal fluency. *Neuropsychology, 17*(2), 240-246. https://doi.org/10.1037/0894-4105.17.2.240
*Ratcliff, G., Dodge, H., Birzescu, M., & Ganguli, M. (2003). Tracking cognitive functioning over time: ten-year longitudinal data from a community-based study. *Appl Neuropsychol, 10*(2), 76-88. https://doi.org/10.1207/S15324826AN1002_03
Reilly, D. (2012). Gender, Culture, and Sex-Typed Cognitive Abilities. *PLoS ONE, 7*(7), e39904. https://doi.org/10.1371/journal.pone.0039904
*Renteria, L. (2005). *Validation of the Spanish language Wechsler Adult Intelligence Scale (3rd edition) in a sample of American, urban, Spanish speaking Hispanics* (Publication Number 3180959) [Ph.D., Loyola University Chicago]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.
*Reynolds, A. G. (1967). *Associational fluency and written language expression* [M.A., The University of Western Ontario (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Riva, D., Nichelli, F., & Devoti, M. (2000, 2000/02/01/). Developmental Aspects of Verbal Fluency and Confrontation Naming in Children. *Brain and Language, 71*(2), 267-284. https://doi.org/https://doi.org/10.1006/brln.1999.2166
Rodriguez-Aranda, C., & Martinussen, M. (2006, 2006/10/01). Age-Related Differences in Performance of Phonemic Verbal Fluency Measured by Controlled Oral Word Association Task (COWAT): A Meta-Analytic Study. *Developmental Neuropsychology, 30*(2), 697-717. https://doi.org/10.1207/s15326942dn3002_3

*Rosen, M. L. (1995). *Concept formation: An integrated developmental model to explain performance differences between ADHD and non-ADHD children* [Ph.D., Illinois Institute of Technology]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Rosselli, M., Ardila, A., Matute, E., & Inozemtseva, O. (2009). Gender differences and cognitive correlates of mathematical skills in school-aged children. *Child Neuropsychology, 15*(3), 216- 231. https://doi.org/10.1080/09297040802195205

*Rosselli, M., Tappen, R., Williams, C., Salvatierra, J., & Zoller, Y. (2009, Nov). Level of education and category fluency task among Spanish speaking elders: number of words, clustering, and switching strategies. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 16*(6), 721-744. https://doi.org/10.1080/13825580902912739

*Rouch, I., Wild, P., Ansiau, D., & Marquié, J. C. (2005, Aug 15). Shiftwork experience, age and cognitive performance. *Ergonomics, 48*(10), 1282-1293. https://doi.org/10.1080/00140130500241670

*Rubin, L. H. (2009). *Effects of sex steroid hormones on cognition in schizophrenia* (Publication Number 3394236) [Ph.D., University of Illinois at Chicago]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Ryan, J. P., Atkinson, T. M., & Dunham, K. T. (2004, Jan). Sports-related and gender differences on neuropsychological measures of frontal lobe functioning. *Clin J Sport Med, 14*(1), 18-24. https://doi.org/10.1097/00042752-200401000-00004

*Sakamoto, M. (2009). *Comparing Alzheimer's disease and vascular dementia profiles on neuropsychological tests among Japanese elders* (Publication Number 3358503) [Ph.D., Drexel University]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Sakurai, H., Hanyu, H., Murakami, M., Kume, K., Takata, Y., Onuma, T., Akai, T., & Iwamoto, T. (2011). The category "animals" is more appropriate than the category "vegetables" to measure semantic category fluency. *Geriatrics & gerontology international*, *11*(3), 374–375. https://doi.org/10.1111/j.1447-0594.2010.00667.x

*Sandel, N. (2016). *High school lacrosse and soccer players' neurocognitive performance and symptoms before and after concussion* (Publication Number 3738377) [Ps.D., Widener University]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Schallmo, M. P., Kassel, M. T., Weisenbach, S. L., Walker, S. J., Guidotti-Breting, L. M., Rao, J. A., Hazlett, K. E., Considine, C. M., Sethi, G., Vats, N., Pecina, M., Welsh, R. C., Starkman, M. N., Giordani, B., & Langenecker, S. A. (2015). A new semantic list learning task to probe functioning of the Papez circuit. *J Clin Exp Neuropsychol, 37*(8), 816-833. https://doi.org/10.1080/13803395.2015.1052732

Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook*. Western Psychological Services.

*Sims, R. C. (2007). *Neurocognitive correlates of cardiovascular risk factors in Blacks* (Publication Number 3299679) [Ph.D., Howard University]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Sinay, R. D. (1967). *Creative aptitude patterns of college honors students* (Publication Number 6807200) [Ph.D., University of Southern California]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Snitz, B. E., Unverzagt, F. W., Chang, C. C., Bilt, J. V., Gao, S., Saxton, J., Hall, K. S., & Ganguli, M. (2009, Dec). Effects of age, gender, education and race on two tests of language ability in community-based older adults. *Int Psychogeriatr, 21*(6), 1051-1062. https://doi.org/10.1017/S1041610209990214

Sokołowski, A., Tyburski, E., Sołtys, A., & Karabanowicz, E. (2020). Sex Differences in Verbal Fluency Among Young Adults. *Advances in cognitive psychology, 16*(2), 92-102. https://doi.org/10.5709/acp-0288-1

*Soleman, R. S., Schagen, S. E., Veltman, D. J., Kreukels, B. P., Cohen-Kettenis, P. T., Lambalk, C. B., Wouters, F., & Delemarre-van de Waal, H. A. (2013, Aug). Sex differences in verbal fluency during adolescence: a functional magnetic resonance imaging study in gender dysphoric and control boys and girls. *J Sex Med, 10*(8), 1969-1977. https://doi.org/10.1111/jsm.12083

*Sosa, A. L., Albanese, E., Prince, M., Acosta, D., Ferri, C. P., Guerra, M., Huang, Y., Jacob, K. S., de Rodriguez, J. L., Salas, A., Yang, F., Gaona, C., Joteeshwaran, A., Rodriguez, G., de la Torre, G. R., Williams, J. D., & Stewart, R. (2009, Aug 26). Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: a cross-sectional survey. *BMC Neurol, 9*, 48. https://doi.org/10.1186/1471-2377-9-48

*Speer, P., Wersching, H., Bruchmann, S., Bracht, D., Stehling, C., Thielsch, M., Knecht, S., & Lohmann, H. (2014). Age- and gender-adjusted normative data for the German version of Rey's Auditory Verbal Learning Test from healthy subjects aged between 50 and 70 years. *J Clin Exp Neuropsychol, 36*(1), 32-42. https://doi.org/10.1080/13803395.2013.863834

Spreen, O., & Benton, A. L. (1977). *Neurosensory center comprehensive examination for aphasia: Manual of directions (revised edition).* . Neuropsychology Laboratory, University of Victoria.

*Stanulis, R. G. (1977). *Hemispheric function and memory: effect of handedness of verbal and imaginal processes* (Publication Number 7805228) [Ph.D., Wayne State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Stevens, J. S., & Hamann, S. (2012, 2012/06/01/). Sex differences in brain activation to emotional stimuli: A meta-analysis of neuroimaging studies. *Neuropsychologia, 50*(7), 1578-1593. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2012.03.011

*Stoddard, E. (2007). *Measuring learning modalities with neuropsychological memory measures in a college population* (Publication Number 3275174) [Ph.D., Drexel University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Stoet, G., & Geary, D. C. (2013, Mar). Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data [Article]. *PLoS ONE, 8*(3), 10, Article e57988. https://doi.org/10.1371/journal.pone.0057988

*Sunderaraman, P., Blumen, H. M., DeMatteo, D., Apa, Z. L., & Cosentino, S. (2013, Jun). Task demand influences relationships among sex, clustering strategy, and recall: 16-word versus 9-word list learning tests. *Cogn Behav Neurol, 26*(2), 78-84. https://doi.org/10.1097/WNN.0b013e31829de450

*Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Mowrey, W., Landau, S., Maki, P. M., & Alzheimer's Disease Neuroimaging, I. (2016, Apr 12). Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. *Neurology, 86*(15), 1368-1376. https://doi.org/10.1212/WNL.0000000000002570

*Swift, P. T. (1999). *Validation of a Spanish language test of verbal learning and memory: The "Perri Test de Aprendizaje Verbal y Memoria"* [Ph.D., University of Connecticut]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Tallberg, I. M., Ivachova, E., Jones Tinghag, K., & Ostberg, P. (2008, Oct). Swedish norms for word fluency tests: FAS, animals and verbs. *Scand J Psychol, 49*(5), 479-485. https://doi.org/10.1111/j.1467-9450.2008.00653.x

*Tanaka, T. R. (2005). *Gender and ethnic differences on select verbal and visuospatial measures among older European and Japanese Americans* (Publication Number 3179491) [Ph.D., Pacific Graduate School of Psychology]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Temple, C. M., & Cornish, K. M. (1993). Recognition memory for words and faces in schoolchildren: A female advantage for words. *British Journal of Developmental Psychology, 11*(4), 421-426. https://doi.org/10.1111/j.2044-835X.1993.tb00613.x

*Thilers, P. P., MacDonald, S. W., & Herlitz, A. (2007, Sep 10). Sex differences in cognition: the role of handedness. *Physiol Behav, 92*(1-2), 105-109. https://doi.org/10.1016/j.physbeh.2007.05.035

*Thomas, L. L., Curtis, A. T., & Bolton, R. (1978). Sex Differences in Elicited Color Lexicon Size. *Perceptual and Motor Skills, 47*(1), 77-78. https://doi.org/10.2466/pms.1978.47.1.77

*Thornburg, H. L. (1973). *An investigation of interrelations of abilities in guilford's structure-of-intellect* (Publication Number 7412212) [Ph.D., University of Illinois at Urbana-Champaign]. ProQuest Dissertations & Theses Global. Ann Arbor.
Thurstone, L., & Thurstone, T. (1962). *Primary mental abilities (Rev.)*. Science Research Associates.
*Tombaugh, T. N., Kozak, J., & Rees, L. (1999, Feb). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch Clin Neuropsychol, 14*(2), 167-177.
*Tuck, S. (2012). *The Development of Verbal Fluency in Children: An Examination of Switching, Clustering, Metastrategic Awareness, and Timing Effects* [Ph.D., York University (Canada)]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Vakil, E., & Blachstein, H. (1997). Rey AVLT: Developmental norms for adults and the sensitivity of different memory measures to age. *The Clinical Neuropsychologist, 11*(4), 356-369. https://doi.org/10.1080/13854049708400464
*van Hooren, S. A., Valentijn, A. M., Bosma, H., Ponds, R. W., van Boxtel, M. P., & Jolles, J. (2007, Jan). Cognitive functioning in healthy older adults aged 64-81: a cohort study into the effects of age, sex, and education. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 14*(1), 40-54. https://doi.org/10.1080/138255890969483
*Vannorsdall, T. D. (2006). *White matter hyperintensities: Neuropsychological correlates in a community -based sample* (Publication Number 3254864) [Ph.D., University of Maryland, Baltimore County]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.
Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007, Feb). Gender differences in object location memory: A meta-analysis [Article]. *Psychonomic Bulletin & Review, 14*(1), 23-38.
Voyer, D., Saint Aubin, J., Altman, K., & Gallant, G. (2021). Sex differences in verbal working memory: A systematic review and meta-analysis. *Psychological Bulletin, 147*(4), 352-398. https://doi.org/10.1037/bul0000320
Voyer, D., Voyer, S., & Bryden, M. P. (1995, Mar). Magnitude of sex-differences in spatial abilities - a metaanalysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174-1204. https://doi.org/10.1037/a0036620
*Wagner, M. D. (1980). *Receptive and expressive language differences between lower and middle-income black children and their relationship to performance on tasks requiring spatial and conceptual inferences* (Publication Number 8021286) [Ph.D., Emory University]. ProQuest Dissertations & Theses Global. Ann Arbor.
*Walburn, K. D. (2014). *Effects of physical activity and stress events on a spatial memory task and stroop task in college students* (Publication Number 1563919) [M.A., University of Nebraska at Omaha]. ProQuest Dissertations & Theses Global. Ann Arbor.
Weber, D., Dekhtyar, S., & Herlitz, A. (2017, 1//). The Flynn effect in Europe – Effects of sex and region. *Intelligence, 60*, 39-45. https://doi.org/http://dx.doi.org/10.1016/j.intell.2016.11.003
Weber, D., Skirbekk, V., Freund, I., & Herlitz, A. (2014, Aug). The changing face of cognitive gender differences in Europe [Article]. *Proc Natl Acad Sci U S A, 111*(32), 11673-11678. https://doi.org/10.1073/pnas.1319538111
Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition*. Pearson.
Wechsler, D. (2009). *Wechsler Memory Scale IV (WMS-IV)* Psychological Corporation.
*Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003, 2003/09/01/). Sex differences in cognitive functions. *Personality and Individual Differences, 35*(4), 863-875. https://doi.org/https://doi.org/10.1016/S0191-8869(02)00288-X
*Weiss, E. M., Ragland, J. D., Brensinger, C. M., Bilker, W. B., Deisenhammer, E. A., & Delazer, M. (2006). Sex differences in clustering and switching in verbal fluency tasks. *Journal of the International Neuropsychological Society, 12*(4), 502-509. https://doi.org/10.1017/S1355617706060656

*Westcott, P. (1969). *Age differences in strategies for free recall* (Publication Number 7017225) [Ph.D., Yale University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Wilkosc, M., Markowska, A., Zajac-Lamparska, L., Skibinska, M., Szalkowska, A., & Araszkiewicz, A. (2016). A Lack of Correlation between Brain-Derived Neurotrophic Factor Serum Level and Verbal Memory Performance in Healthy Polish Population. *Front Neural Circuits, 10*(MAY), 39. https://doi.org/10.3389/fncir.2016.00039

*Willis, M. P. (1997). *Cognitive capacities that contribute to 5- and 6-year-old children's ability to demonstrate nonverbal and verbal understanding of spatial relational polar opposites* (Publication Number 9728323) [Ph.D., Columbia University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954, 1954/12/01). A factor-analytic study of creative-thinking abilities. *Psychometrika, 19*(4), 297-311. https://doi.org/10.1007/BF02289230

*Wolkenberg, F. A. (1999). *Differences between men and women in a lexical decision task and the role of hemispheric processing* (Publication Number 9941846) [Ph.D., Emory University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Wormack, L. (1979). Cognitive Predictors of Articulation in Writing. *Perceptual and Motor Skills, 48*(3_suppl), 1151-1156. https://doi.org/10.2466/pms.1979.48.3c.1151

*Wulandari, S. W., & Hendrawan, D. (2020). Trust your Abilities More than the Stereotype: Effect of Gender-stereotype Threat and Task Difficulty on Word Production, Clustering, and Switching in Letter Fluency. *Pertanika Journal of Social Sciences and Humanities, 28*(4), 2567--2588. https://doi.org/10.47836/pjssh.28.4.05

*Xu, H. (2018). *Association Between Migration and Cognitive Function Among Middle Aged and Older Adults: A Comparison Between China and India* [Ph.D., Duke University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Yeudall, L. T., Fromm, D., Reddon, J. R., & Stefanyk, W. O. (1986). Normative data stratified by age and sex for 12 neuropsychological tests. *Journal of Clinical Psychology, 42*(6), 918-946. https://doi.org/10.1002/1097-4679(198611)42:6<918::AID-JCLP2270420617>3.0.CO;2-Y

*Yi, A. S. (2007). *A methodological approach to gender differences in cognition among older adults* (Publication Number 3262586) [Ph.D., Fuller Theological Seminary, School of Psychology]. Health Research Premium Collection; ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Yonker, J. E., Eriksson, E., Nilsson, L. G., & Herlitz, A. (2003, Jul). Sex differences in episodic memory: minimal influence of estradiol. *Brain Cogn, 52*(2), 231-238. https://doi.org/10.1016/s0278-2626(03)00074-5

*Young, M. C. (2002). *Anterior aphasia as a natural category of acquired cognitive -communicative impairment: Implications for cognitive neurolinguistic theory, experimental methods, and clinical practice* (Publication Number 3089491) [Ph.D., The University of Texas at Austin]. ProQuest Dissertations & Theses Global. Ann Arbor.

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70*(1), 10-20. https://doi.org/10.1037/a0038208

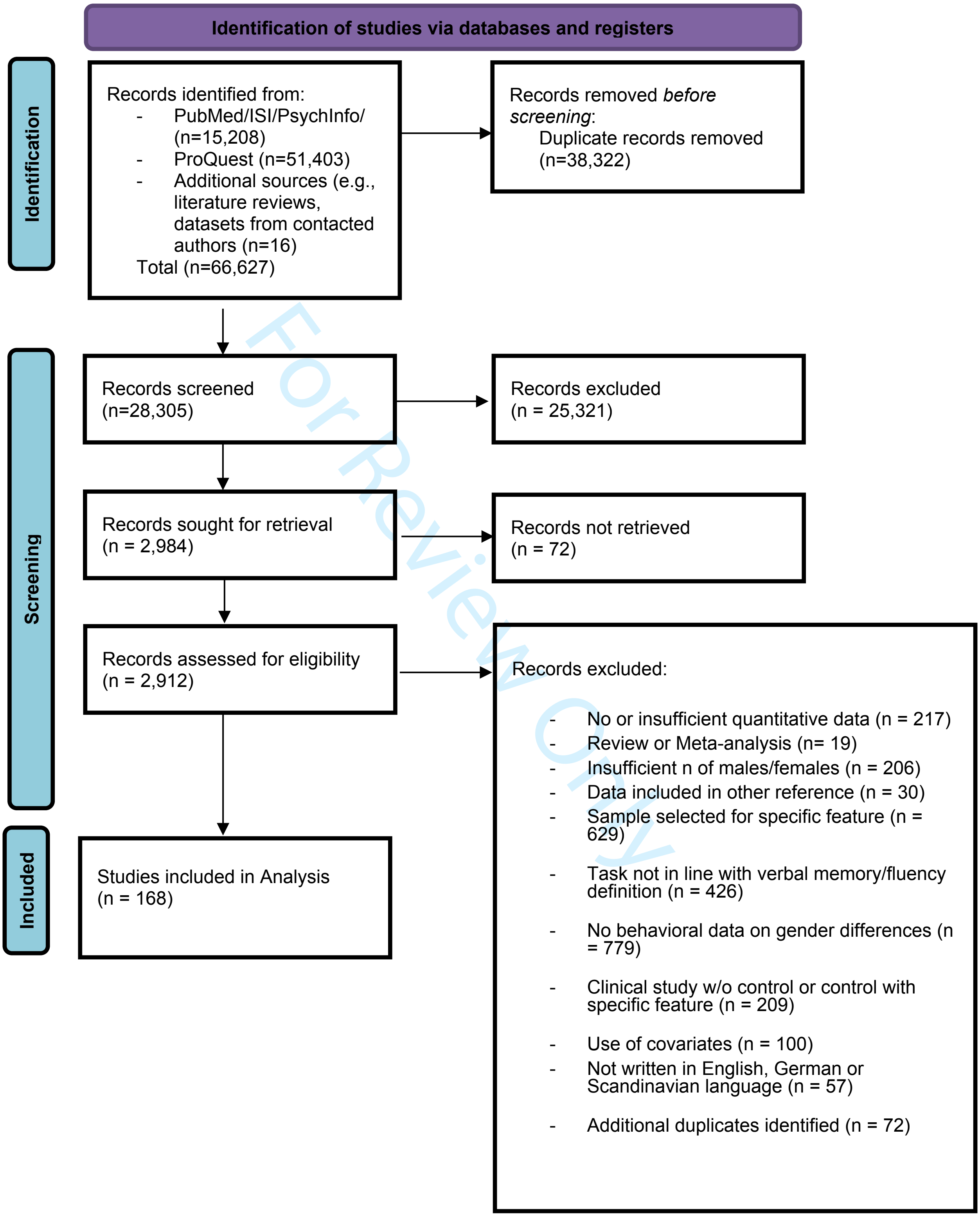
Zhao, Q., Guo, Q., & Hong, Z. (2013, Feb). Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. *Neurosci Bull, 29*(1), 75-82. https://doi.org/10.1007/s12264-013-1301-7

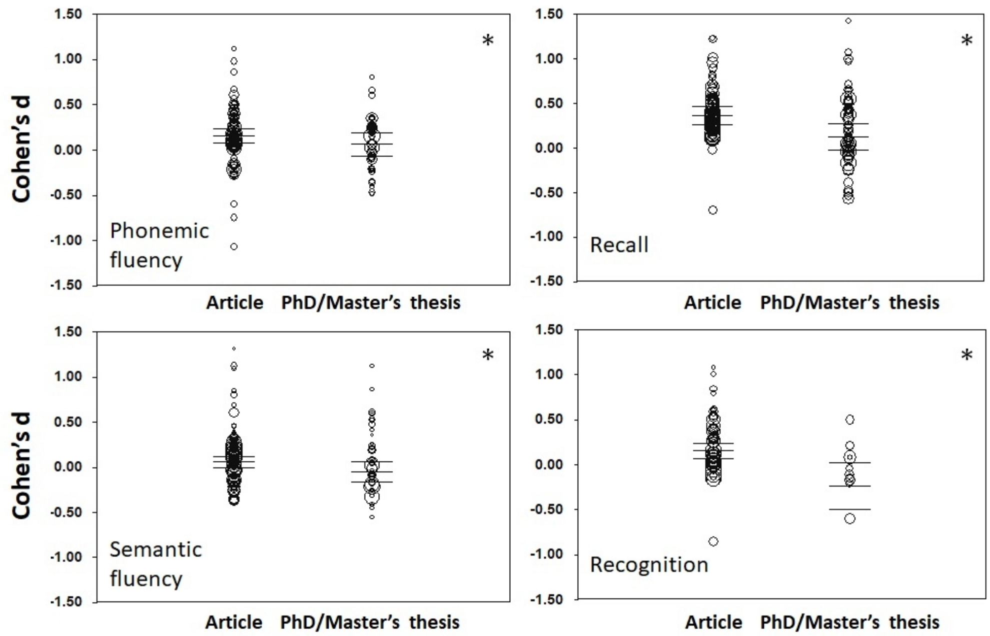
*Zhu, W. (2015). *Association of Objectively Measured Physical Activity with Cognitive Function in Black and White Older Adults: Reasons for Geographic and Racial Differences in Stroke (REGARDS) study* (Publication Number 3701737) [Ph.D., Arizona State University]. ProQuest Dissertations & Theses Global. Ann Arbor.

*Zoccoli, S. L. (2005). *The COWA: An empirical study of phonemic and semantic verbal fluency* [M.A., Southern Methodist University]. ProQuest Dissertations & Theses Global. Ann Arbor.

Figure 1

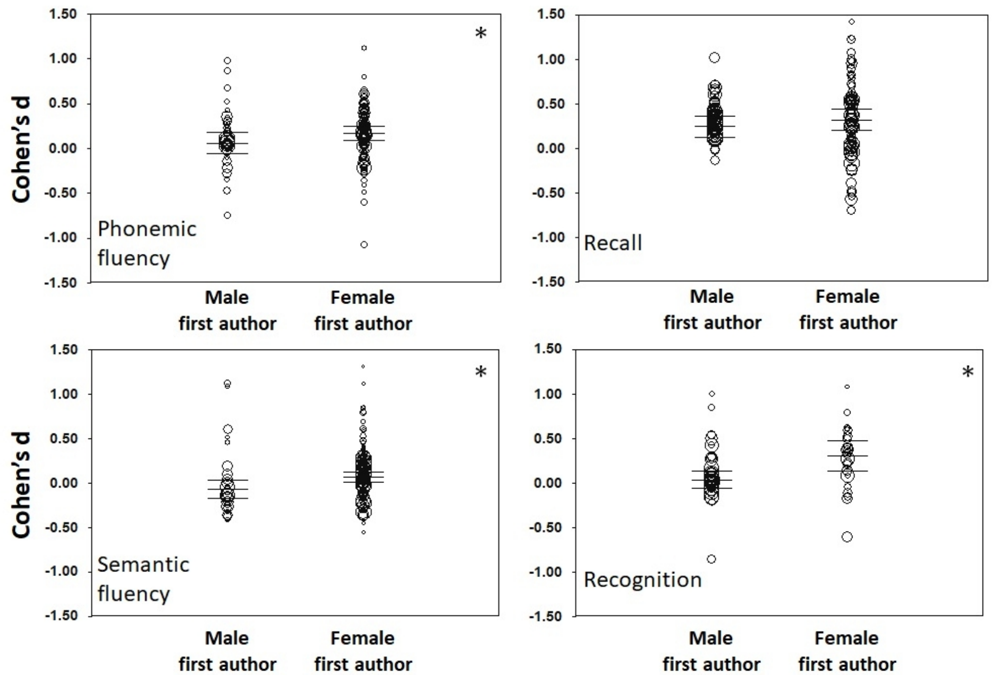
PRISMA Flow Diagram Showing the Study Selection Process





Effect of publication type

208x133mm (120 x 120 DPI)



Gender of first author effect

209x143mm (120 x 120 DPI)

Table 1

Descriptive overview of sex/gender differences in verbal fluency and verbal episodic memory measures

Verbal ability	Test/Measure	Effect size
Verbal fluency	**Total effect**	***d* = 0.07 (0.04 to 0.10) *k*=290**
Phonemic fluency	**Total effect**	***d* = 0.13 (0.09 to 0.16) *k*=135**
	Generic starting letter(s)	*d* = 0.12 (0.07 to 0.18) *k*=59
	COWAT/F, A, S	*d* = 0.14 (0.08 to 0.20) *k*=55
	Four-word sentences	*d* = 0.03 (-0.20 to 0.26) *k*=5
Semantic fluency	**Total effect**	***d* = 0.02 (-0.02 to 0.06) *k*=147**
	Category: Animals	*d* = -0.13 (-0.16 to -0.09), *k*=58
	Categories Animals & fruits/vegetables/food	*d* = 0.11 (0.03 to 0.18), *k*=26
	Objects with specific color	*d* = 0.19 (0.13 to 0.25), *k*=10
	Categories Animals, fruits/vegetables/food, & action verbs	*d* = 0.25 (-0.03 to 0.53), *k*=8
	Fruits/vegetables/food	*d* = 0.31 (0.16 to 0.47), *k*=8
Verbal episodic memory	**Total effect**	***d* = 0.23 (0.19 to 0.26) *k*=206**
Recall	**Total effect**	***d* = 0.28 (0.23 to 0.32), *k*=136**
	California Verbal Learning Test	*d* = 0.42 (0.32 to 0.52), *k*=28
	Rey Auditory Verbal Learning Test	*d* = 0.39 (0.29 to 0.48), *k*=24
	Generic word list	*d* = 0.17 (0.06 to 0.28), *k*=16
	Delayed Memory for Names/Visual-Auditory Learning from Woodcock Johnson Psycho-Educational Battery - Revised	*d* = -0.13 (-0.27 to 0.01), *k*=12
	10 Word Learning Test from CERAD	*d* = 0.18 (0.07 to 0.28), *k*=10
	Ten-Words Test	*d* = 0.26 (0.13 to 0.39), *k*=7
	Deese, Roediger and McDermott task	*d* = 0.15 (0.02 to 0.28), *k*=7
Recognition	**Total effect**	***d* = 0.12 (0.06 to 0.17), *k*=66**
	Rey Auditory Verbal Learning Test	*d* = 0.22 (0.12 to 0.33), *k*=18
	California Verbal Learning Test	*d* = 0.17 (0.06 to 0.29), *k*=13
	Deese, Roediger and McDermott task	*d* = 0.15 (0.04 to 0.27), *k*=7
	Story telling delayed recognition	*d* = -0.07 (-0.18 to 0.04), *k*=7
	Story telling immediate recognition	*d* = 0.02 (-0.09 to 0.13), *k*=7

Note. Values in parentheses represent 95% CI and *k* = number of effect sizes included. Effect sizes are provided assuming independence between multiple outcomes within the same study. Effect sizes within each sub-category were combined with a random-effects model, assuming a common among study variance component across sub-categories. That is, T^2 was computed for each age group and then pooled across subgroups. Only tests with at least seven effect sizes are provided, with the exception of phonemic fluency, where the three most frequent tests are provided.

Table 2

Descriptive overview of age group effects

	Phonemic fluency	**Semantic fluency***	Recall	Recognition
Child/Child preschool (≤12 yrs)	*d* = 0.13	*d* = 0.09	***d* = 0.05***	***d* = 0.13***
	(0.06 to 0.25, *k*=29)	(-0.02 to 0.17, *k*=30)	(-0.06 to 0.17, *k*=15)	(-0.04 to 0.31, *k*=7)
Adolescent (13-18 yrs)	*d* = 0.22	*d* = 0.03	*d* = 0.13	***d* = 0.11***
	(0.03 to 0.41, *k*=5)	(-0.25 to 0.30, *k*=2)	(-0.06 to 0.31, *k*=7)	(-0.14 to 0.35, *k*=3)
Adult (19-44 yrs)	*d* = 0.24	*d* = 0.15	*d* = 0.28	*d* = 0.02
	(0.07 to 0.41, *k*=7)	(0.10 to 0.21, *k*=8)	(0.17 to 0.39, *k*=15)	(-0.10 to 0.13, *k*=9)
Middle aged (45-64 yrs)	*d* = 0.13	***d* = 0.25***	*d* = 0.34	*d* = 0.13
	(0.03 to 0.23, *k*=7)	(0.17 to 0.32, *k*=6)	(0.24 to 0.45, *k*=9)	(-0.04 to 0.28, *k*=6)
Aged (≥ 65 yrs)	*d* = 0.06	***d* = -0.10***	*d* = 0.17	*d* = 0.06
	(-0.03 to 0.15, *k*=15)	(-0.14 to -0.07, *k*=31)	(0.09 to 0.24, *k*=21)	(-0.09 to 0.21, *k*=10)

Note. Values in parentheses represent 95% CI and *k* = number of effect sizes included. Individual age groups marked with asterisk/bold print differed significantly from the reference group "Adult". Verbal ability measures marked with asterisk/bold print indicate that the sex/gender difference varied significantly across all age groups. This table may contain more effect sizes than the meta-regression because the meta regression only includes studies with info on all covariates. Figures are based on assuming perfect independence between multiple measures from the same (sub-)sample.