Is there a bilingual advantage in auditory attention among children? A systematic review and meta-analysis of standardized auditory attention tests

Wenfu Bao¹, ²*, Claude Alain³, ⁴, ⁵, ⁶, Michael Thaut², ⁵, ⁶, ⁷, and Monika Molnar¹, ²

¹ Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada
² Rehabilitation Sciences Institute, University of Toronto, Toronto, Ontario, Canada
³ Rotman Research Institute, Baycrest Health Centre, Toronto, Ontario, Canada
⁴ Department of Psychology, University of Toronto, Toronto, Ontario, Canada
⁵ Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada
⁶ Music and Health Science Research Collaboratory, University of Toronto, Toronto, Ontario, Canada
⁷ Faculty of Music, University of Toronto, Toronto, Ontario, Canada

* Corresponding author

E-mail: wenfu.bao@mail.utoronto.ca (WB)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

A wealth of research has investigated the effects of bilingualism on cognition, especially on executive function. Developmental studies reveal different cognitive profiles between monolinguals and bilinguals in (audio-)visual attention tasks, which might stem from their attention allocation differences. Yet, whether such distinction exists in the auditory domain alone is unknown. In this study, we compared differences in auditory attention, measured by standardized tests, between monolingual and bilingual children. A comprehensive literature search was conducted in three electronic databases: OVID Medline, OVID PsycInfo, and EBSCO CINAHL. Empirical studies using standardized tests to assess auditory attention in monolingual and bilingual participants aged less than 18 years were included. Among the 20 studies identified, we performed a meta-analysis and found that test measure (accuracy vs. response times) was significantly related to differences in effect sizes whereas other factors (e.g., participant age, stimulus type) were not. Specifically, studies reporting accuracy observed a marginal bilingual advantage ($g = 0.10$), but those reporting response times indicated a small monolingual benefit ($g = -0.34$). Therefore, there was little difference between monolingual and bilingual children’s performance on standardized auditory attention tests. We also found that studies tend to include a wide variety of bilingual children but report limited language background information of the participants. This, unfortunately, limits the potential theoretical contributions of the reviewed studies. Recommendations to improve the quality of future research are discussed.
Introduction

Over the past decade, the growing body of research investigating the effects of bilingualism on cognition suggests that bilingual experience can shape the brain and cognitive systems [1, 2]. Some studies [3, 4] demonstrate that bilingual participants outperform their monolingual counterparts on a wide range of cognitive tasks while controlling for other factors, which is often interpreted as a bilingual advantage. For example, bilingual children develop conflict resolution ability earlier than monolinguals and perform better on memory tasks based on executive control [3] (but see [5] for equal performance in inhibitory tasks).

Cognition has many facets, among which executive function (EF) is most frequently measured in bilingual developmental studies [6]. Recent meta-analyses present mixed results about whether bilingualism confers an EF benefit in children. Gunnerud et al. [7] analyzed 143 studies examining different EF components in children aged 18 months through 14.5 years; no bilingual advantage was found in overall EF after adjusting for publication bias. Similarly, Lowe et al. [8] reported a negligible effect of bilingualism on overall EF ($g = -0.04$) in children aged between 3 and 17 years. However, using the Bayesian statistical approach, Grundy [9] revealed that when group differences do appear on EF tasks, bilinguals outperform monolinguals far more likely than chance. Given these discrepant findings, the bilingual advantage appears small at best and might be subject to specific circumstances [10]. However, what these “circumstances” are remains unclear.

Addressing whether a general bilingual advantage exists is beyond the scope of this study. We do not intend to engage in a dichotomous discussion to this debate, as it oversimplifies the effects of bilingualism on cognitive development. Nevertheless, we acknowledge that bilingualism can confer certain cognitive benefits at least in some types of bilinguals, which has
been supported by empirical studies [11, 12]. Further, attention has been hypothesized to be a plausible domain responsible for the bilingualism effects [2]. That is, the habitual use of two languages over years possibly enables bilinguals to become more practiced in managing conflicts and controlling attention. In particular, early bilingual exposure can affect how attention is allocated to the environment. For instance, infants growing up in bilingual homes pay more attention to subtle environmental differences [13, 14], which could improve their attentional processing [15].

However, a definition of attention remains vague in cognitive literature. Here, we conceptualize it as a system with three primary components: sustained attention, selective attention, and executive control [16, 17]. Specifically, based on Petersen and Posner [16], sustained attention (or alerting) refers to maintaining alertness over a long period of time, which is usually measured in the form of vigilance tasks that involve monitoring a target stimulus interspersed with non-targets. Selective attention (or orienting) refers to the ability to select certain input for enhanced processing while suppressing other irrelevant information, and thus is often measured by focusing on the target stimuli and ignoring the distractors. Lastly, executive control deals with resolving conflicts, shifting attention, and regulating thoughts and behavior. These components are also associated with different neural substrates in the human brain [16].

Visual or audio-visual attention has been primarily studied in bilingualism research, suggesting that bilingual experience modulates attention allocation [18, 19]. Auditory attention, however, has received little consideration, despite its importance to language processing and development: language is often processed through the auditory domain alone (e.g., speech perception), and infants start learning about language in utero without any visual support [20].
Likewise in the visual domain, bilingualism might also affect how auditory attention is allocated in bilingual children.

The current systematic review and meta-analysis assess whether there are reliable differences in auditory attention, measured by standardized tests, between monolingual and bilingual children. An initial search in our laboratory indicates that standardized tests are often used to assess children’s auditory attention in research and clinical settings [21], and are an ideal comparison across labs given consistent administration guidelines. These tests target different auditory attention components and use different types of auditory stimuli, which are either linguistic or non-linguistic (see Results for further information).

As attention might support monolingual and bilingual development differently [2], investigating auditory attention in monolingual and bilingual children will illuminate different adaptations to their language environments. Additionally, examining the bilingualism effects on auditory attention will address the current literature gap by focusing on the auditory domain, which has been overlooked in prior attention research. Our prediction is that if auditory attention development is shaped by bilingual experience, monolingual and bilingual children would perform differently on standardized tests; considering the bilingual advantage hypothesis, bilinguals might perform better.

An additional goal of our study is to determine whether certain bilingual characteristics (e.g., age of acquisition, language proficiency) mitigate the potential differences between monolinguals and bilinguals. Since cognitive advantages tend to appear in more proficient bilinguals [11, 12], we hypothesize that simultaneous bilinguals (i.e., children who learn both languages before the age of three) would more likely show enhanced auditory attention than sequential bilinguals (i.e., children who learn additional languages after the age of three). Finally,
given that it is unclear whether bilingualism exerts influence beyond linguistic domain and standardized tests vary by stimulus type (linguistic vs. non-linguistic), we will assess if different auditory stimuli affect attention performance in monolingual and bilingual children.

**Materials and methods**

This study is part of a larger systematic review that investigates auditory attention development from infancy to adolescence, in which the population’s language experience (i.e., monolingual vs. bilingual) and research methods are not controlled. The review protocol was registered a priori with OSF [21]. We followed the *Cochrane Handbook for Systematic Reviews of Interventions* [22] as the methods guidance, and developed the search strategy in consultation with a health sciences librarian. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the screening process is presented in Fig 1. The methods described below are specific to our current (more focused) literature search embedded in Bao and Molnar [21].

**Eligibility criteria**

To include the most relevant studies on auditory attention measured by standardized tests, we adhered to the following criteria when determining a study’s eligibility:

- Studies that used standardized tests to measure auditory attention were included.

  *Standardized tests* are norm-referenced tests administered and scored in a consistent...
manner. Studies using non-standardized measures or focusing on visual attention and other cognitive abilities were excluded.

- Studies that tested typically developing individuals below the age of 18 and had a monolingual group and a bilingual group were included. The bilingual group should be raised in bilingual families or study in immersion schools where most of the curriculum is taught in an unfamiliar language. Studies only having monolingual or bilingual participants were excluded.

- Studies that controlled for participants’ age and socio-economic status (SES) were included, as both factors can affect children’s cognitive measures independently [23].

- Published, empirical studies that reported primary results in peer-reviewed articles were included.

- Given the reviewers’ capabilities, only articles written in English were included.

**Information sources and search**

A comprehensive literature search was conducted in three electronic databases: OVID Medline, OVID PsycInfo and EBSCO CINAHL. Search terms were defined after consulting the librarian and tailored to each database (see S1-S3 Tables). The search contained articles from the databases’ start date up to March 8, 2023. Based on the search result of the larger systematic review, we identified all studies that compared monolingual and bilingual performance on standardized auditory attention tests. In addition, a manual search was performed by checking the reference lists of relevant articles, and an updated manual search was undertaken in Google Scholar to identify more studies.
**Study selection**

Records and data were managed using the Covidence systematic review software [24]. Duplicates were identified and removed during reference importing and screening. Two independent reviewers screened studies in two phases: (1) title/abstract screening required to indicate “Yes,” “No” or “Maybe” for relevance; (2) full-text screening required to indicate “Include” or “Exclude” for eligibility and specify the exclusion reasons. If discrepancies occurred, a third reviewer was called to resolve the conflicts.

**Data collection**

For eligible studies, two independent reviewers extracted the following data items on Covidence: (1) sample size (monolingual vs. bilingual); (2) participant age; (3) language background (first and second language), bilingual type (simultaneous vs. sequential), and bilingualism assessment; (4) standardized tests used to assess auditory attention; (5) test measure (accuracy vs. response times or RTs); (6) stimulus type (linguistic vs. non-linguistic); (7) auditory attention components measured; (8) reporting of SES. In the case of unclear or missing items, we contacted the investigators to confirm and obtain additional information. Afterwards, a third reviewer built consensus on extracted data.

**Data analysis**

Apart from a narrative synthesis of included data, we built meta-regression models to analyze the studies’ effect size using the “metafor” package (version 3.0-2) [25] in R [26]. Specifically, the dependent variable was the unbiased estimate of the population effect size Hedges’ $g$, calculated for each study. The independent variables were test measure (binary:
accuracy vs. RTs), stimulus type (binary: linguistic vs. non-linguistic), participant age (continuous, in years) and attention components (categorical: selective attention, sustained attention, executive control, or auditory attention overall). To evaluate the effects of these predictors, we conducted model comparison in a forward stepwise manner: a likelihood ratio test was performed using the anova function to compare a reduced model and a full model, which had one additional component. If the resulting $p$-value was less than the conventional threshold of 0.05, the full model provided a better fit for the data. Otherwise, the reduced model was favored. Information criterion statistics (e.g., Akaike’s Information Criterion or AIC, Bayesian Information Criterion or BIC) were inspected for both models (with lower values indicating a better model performance), and BIC is preferred over AIC when the heterogeneity is large in the studies [27]. For categorical variables that significantly predicted the effect size, subgroup analyses were conducted using the “meta” package (version 5.2-0) [28] and forest plots were generated to visualize the effects.

Furthermore, publication bias was analyzed for the included studies by means of the funnel plot, created using the “meta” package. The funnel plot displays the studies’ effect size against its standard error. Usually, symmetry indicates the absence of publication bias, which is reflected by data points scattered around the mean effect size forming an upside-down funnel. This symmetry was further quantified through Egger’s regression test.

**Results**

**Synthesis**

The database search yielded 634 records for title/abstract screening. Forty articles remained for full-text review, and three articles qualified for extraction. Our manual search
contributed another nine articles, four of which did not appear in the database search as they were not indexed. In total, 12 articles were extracted, which included 20 studies.

Population characteristics

With an age range between 5 to 14 years, participants can be categorized into preschoolers (5-6 years), primary school-aged children (7-11 years), and young adolescents (12-14 years). This categorization has considered different education systems across countries. For example, while primary school starts at the age of four in the Netherlands, the first two years are comparable to kindergarten. Therefore, the five-year-old children in Boerma et al. [29] and Kwakkel et al. [30] were treated as preschoolers. Table 1 presents a detailed description of all participants. Taking language background for an example, there are a great variety of first and second languages, with French-English and Spanish-English being the most common language combinations within bilinguals.

Table 1. Descriptive statistics of included studies.
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Monolingual ((n))</th>
<th>Bilingual ((n))</th>
<th>Age (mean; years)</th>
<th>First language; second language</th>
<th>Bilingual type</th>
<th>Bilingualism assessment</th>
<th>Standardized test</th>
<th>Test measure</th>
<th>Stimulus type</th>
<th>Attention component</th>
<th>SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbu et al. 2019</td>
<td>Belgium</td>
<td>57</td>
<td>59</td>
<td>6</td>
<td>French; English</td>
<td>Sequential* (French children enrolled in an immersion program since the age of 5)</td>
<td>Customized language background questionnaire</td>
<td>KiTAP: The Owls</td>
<td>RT</td>
<td>Non-linguistic: animal sounds</td>
<td>Selective attention</td>
<td>Matched for two groups, from diverse levels</td>
</tr>
<tr>
<td>Boerma et al. 2017</td>
<td>The Netherlands</td>
<td>32</td>
<td>32</td>
<td>5</td>
<td>Turkish, Tarifit-Berber, Arabic; Dutch</td>
<td>Sequential* (children learned Dutch as an L2 since the age of 4)</td>
<td>PaBiQ</td>
<td>IVA + Plus: Auditory task</td>
<td>Accuracy</td>
<td>Linguistic: numbers</td>
<td>Sustained attention</td>
<td>No group difference</td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (1)</td>
<td>USA</td>
<td>30</td>
<td>30</td>
<td>5</td>
<td>Spanish; English</td>
<td>Simultaneous* (American children exposed to Spanish since at least 12 months of age)</td>
<td>Customized language background questionnaire; Language Dominance Survey from EOWPVT-SBE</td>
<td>ACPT-P (modified): Verbal Go/No-Go task</td>
<td>RT</td>
<td>Linguistic: speech syllables</td>
<td>Auditory attention (sustained attention, executive control)</td>
<td>Matched for two groups, all from low status</td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Customized language background questionnaire</td>
<td>Parental report</td>
<td>NEPSY-II: Auditory Attention</td>
<td>Accuracy</td>
<td>Linguistic: words</td>
<td>Selective and sustained attention</td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEPSY-II: Response Set</td>
<td>Accuracy</td>
<td>Linguistic: words</td>
<td>Executive control</td>
<td></td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-linguistic: animal and nature sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garratt &amp; Kelly 2007</td>
<td>UK</td>
<td>27</td>
<td>27</td>
<td>7</td>
<td>English; Bengali, Urdu/Punjabi, Malay, Arabic</td>
<td>Not reported</td>
<td>Customized language background questionnaire</td>
<td>NEPSY: Auditory Attention and Response Set</td>
<td>Accuracy</td>
<td>Linguistic: words</td>
<td>Auditory attention</td>
<td>Comparable between two groups, most from low status</td>
</tr>
<tr>
<td>Karlsson et al. 2015 (1)</td>
<td></td>
<td>25</td>
<td>24</td>
<td>7</td>
<td>Swedish; Finish and other (“other” second languages were not specified in the study)</td>
<td>Simultaneous (almost all bilingual children learned both languages by the age of 3)</td>
<td>Parental report</td>
<td>NEPSY-II: Auditory Attention</td>
<td>Accuracy</td>
<td>Linguistic: words</td>
<td>Selective and sustained attention</td>
<td>Matched for two groups</td>
</tr>
<tr>
<td>Karlsson et al. 2015 (2)</td>
<td>Finland</td>
<td>23</td>
<td>27</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>NEPSY-II: Response Set</td>
<td>Accuracy</td>
<td>Linguistic: words</td>
<td>Executive control</td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (3)</td>
<td></td>
<td>25</td>
<td>24</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-linguistic: animal and nature sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (4)</td>
<td></td>
<td>23</td>
<td>27</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krizman et al. 2012</td>
<td>USA</td>
<td>25</td>
<td>23</td>
<td>14</td>
<td>Spanish; English</td>
<td>Simultaneous (children had their first exposure to both languages about the age of 3)</td>
<td>LEAP-Q</td>
<td>IVA + Plus: Auditory task</td>
<td>Accuracy</td>
<td>Linguistic: numbers</td>
<td>Sustained and selective attention</td>
<td>Matched for two groups</td>
</tr>
<tr>
<td>Krizman et al. 2014</td>
<td>USA</td>
<td>27</td>
<td>27</td>
<td>14</td>
<td>Spanish; English</td>
<td>Simultaneous (children learned both languages before/after the age of 3)</td>
<td>LEAP-Q: parental report</td>
<td>IVA + Plus: Auditory task</td>
<td>Accuracy</td>
<td>Linguistic: numbers</td>
<td>Attentional control</td>
<td>44% of monolinguals and 59% of bilinguals from low status Matched for two groups, from high status overall</td>
</tr>
<tr>
<td>Kwakkel et al. 2021</td>
<td>The Netherlands</td>
<td>80</td>
<td>89</td>
<td>5</td>
<td>Dutch; English</td>
<td>Sequential* (Dutch children enrolled in a bilingual program since the age of 4)</td>
<td>Parental questionnaire</td>
<td>ACPT: Go/No-Go task</td>
<td>Accuracy</td>
<td>Non-linguistic: tones</td>
<td>Sustained attention</td>
<td></td>
</tr>
<tr>
<td>Nicolay &amp; Belgium 2014</td>
<td>Belgium</td>
<td>51</td>
<td>53</td>
<td>8</td>
<td>French; English</td>
<td>Sequential* (French children enrolled in a bilingual program since the age of 4)</td>
<td>Not explicitly</td>
<td>KiTAP: The Owls</td>
<td>RT</td>
<td>Non-linguistic: Selective</td>
<td>selective attention</td>
<td>Matched for two groups, from high status overall</td>
</tr>
</tbody>
</table>
AUDITORY ATTENTION IN BILINGUAL CHILDREN

Poncelet 2013

Nicolay & Poncelet 2015
Belgium 50 51 8 French; English Sequential* (French children enrolled in an immersion program since the age of 5) Not explicitly assessed KITAP: The Owls Animal sounds attention two groups, from medium and high levels Matched for two groups, from diverse levels

Simonis et al. 2020 (1)
Belgium 129 156

Simonis et al. 2020 (2)

Simonis et al. 2020 (3)

Strydom et al. 2022 South Africa 20 20 7 Non-English (first languages were not specified in the study); English Simultaneous* (children learned English as a second language before the age of 3) Not explicitly assessed SAAT: Words Intelligibility by Picture Identification in quiet Linguistic: words Selective attention Same for two groups


Numbers following the names signify individual studies from the same article.

*Asterisk indicates that the bilingual type was not directly reported by the investigators but inferred from the original articles.
However, there was not sufficient information about bilingual characteristics in the reviewed papers. For instance, bilingual type (simultaneous vs. sequential) was rarely reported, except that Karlsson et al. [31] and Krizman et al. [32, 33] recruited “early bilinguals.” Usually, bilingual children were from immigrant families learning the societal language as their second language (L2) [29, 34], or enrolled in immersion schools acquiring a foreign language [35-38]. Note that for longitudinal studies, we considered participants’ language condition when they were tested. For example, Nicoley and Poncelet [37] collected data at two time points; data from the second one (i.e., three years after immersion school enrollment) were included only, because children cannot be deemed bilingual yet during the initial testing when they just started the program. Generally, immersion school children were considered sequential bilinguals, as most of them were from a monolingual environment and did not learn an L2 until they started school. Additionally, bilingual experience was often evaluated using parental questionnaires that focus on one aspect (e.g., language use, exposure, or proficiency), which led to inconsistent definitions of bilinguals and the difficulty of comparing them across studies.

SES was reported in all studies, and mostly assessed through education level of the mother or both parents (parental occupation was also used occasionally, e.g., Garratt and Kelly [34]). Monolingual and bilingual participants had a comparable SES across articles, except in Simonis et al. [38] where the bilingual group had a higher SES than the monolingual group. Given no significant difference on test performance between the two groups, this article was included in final analysis.

**Test characteristics**

Various standardized tests (see Table 1) were used to assess auditory attention.

Considering how they were described in the included studies, we categorized them into the
following three groups: (1) Test for Attentional Performance (TAP) [39], Test of Attentional Performance for Children (KiTAP) [40], and Selective Auditory Attention Test (SAAT) [41]: assess *selective attention*; (2) Integrated Visual and Auditory Continuous Performance Test (IVA + Plus) [42], Auditory Continuous Performance Test for Preschoolers (ACPT-P) [43], and Test of Everyday Attention for Children (TEA-Ch) [44]: assess *sustained attention*; (3) Developmental Neuropsychological Assessment (NEPSY) [45] and its second version (NEPSY-II) [46]: assess *auditory attention overall* (selective and sustained attention, and executive control). For tests evaluating audio-visual attention, we included data from subtests using only auditory stimuli. Further, we found that both linguistic and non-linguistic stimuli were utilized. For example, some tests used non-linguistic stimuli like auditory tones (TAP) and animal sounds (KiTAP; e.g., [35, 38]), whereas others employed linguistic stimuli such as numbers (IVA + Plus; TEA-Ch), speech syllables (ACPT-P), and words (NEPSY, NEPSY-II, SAAT; e.g., [29, 32, 33, 47]). Though it remains unknown whether different types of auditory stimuli affect attentional performance.

Moreover, some researchers used translated versions (e.g., Swedish version of NEPSY-II in [31]) or adapted the tests [38], and reported the assessed attention components differently. For example, modelled after ACPT-P, Foy and Mann [48] used an auditory Go/No-Go task with two blocks. We infer that the first block assessed sustained attention and the second one evaluated executive control, as the authors did not explicitly report the attention components. Using the same task, however, Kwakkel et al. [30] stated that they measured “sustained attention.”

Another source of variance we observed involved the test measure reported. All studies presented either accuracy or RTs, except that Simonis et al. [38] did both. Nevertheless, how accuracy was reported in different studies was not always consistent, even for the same test. For
instance, Garratt and Kelly [34] reported standard scores for the Auditory Attention and
Response Set subtests of NEPSY together, whereas Karlsson et al. [31] reported raw scores for
each subtest of NEPSY-II individually. Simonis et al. [38] used decimal numbers to indicate
accuracy (i.e., 1 = 100% accuracy), whereas Kwakkel et al. [30] used reversed omission scores.
Despite these discrepancies, for all accuracy statistics in this meta-analysis, larger values indicate
better auditory attention. However, we did not include data that were presented in other forms
and could not be converted to accuracy scores or proportions, e.g., misses in Foy and Mann [48]
and errors in Barbu et al. [35]. When sufficient statistical details were not reported in a given
study, we contacted the investigators to obtain them. Alternatively, we used WebPlotDigitizer
[49] to extract data from graphs in Krizman et al. [32].

Meta-analysis

Effect size was analyzed using meta-regression modelling. We first built a mixed-effects
model, with each study’s effect size as the dependent variable and test measure as the moderator.
Model summary suggested that test measure significantly influenced the studies’ effect size ($p = 0.0067$; see S4 Table). Further, results of a subgroup analysis confirmed that there was a
significant difference in effect size between studies reporting accuracy (i.e., accuracy studies)
and studies reporting RTs (i.e., RT studies; $p = 0.0014$; see S5 Table). Fig 2 shows a forest plot
stratified by test measure: a marginal bilingual advantage was found in accuracy studies ($g = 0.10$), whereas a small monolingual benefit was observed in RT studies ($g = -0.34$). However,
results favored monolingual children ($g = -0.09$) in standardized auditory attention tests when
taking both measures together.
Fig 2. Forest plot stratified by test measure, based on random-effects model analysis. The upper and lower panels display the results for accuracy studies \( (n = 12) \) and RT studies \( (n = 8) \), respectively.

To investigate the effects of other variables, we divided our data into two subsets (i.e., accuracy studies and RT studies) and analyzed them separately. For accuracy studies \( (n = 12) \), a meta-regression model with participant age as the predictor was built. Age did not influence the effect size significantly \( (p = 0.1282; \text{see S6 Table}) \). However, a positive trend favoring young bilingual adolescents was observed (see S1 Fig). Following the same approach, we tested the effects of stimulus type and attention components, but neither of them was significant (stimulus type: \( p = 0.9550 \); attention components: \( p = 0.4865 \); see S7 and S8 Tables). For RT studies \( (n = 8) \), none of the variables above was significant (participant age: \( p = 0.2056 \), stimulus type: \( p = 0.5494 \), attention components: \( p = 0.9396 \); see S9-S11 Tables). In addition, there was not substantial publication bias in our data according to the funnel plot (S2 Fig), which was further confirmed by the result of Egger’s regression test \( (p = 0.414; \text{see S12 Table}) \). However, we did not include unpublished research, which is a potential limitation.

**Discussion**

The current systematic review and meta-analysis analyzed 20 studies that compared monolingual and bilingual children’s performance on standardized auditory attention tests. Results suggest that test measure was significantly related to differences in effect sizes, with a marginal bilingual advantage in accuracy studies \( (n = 12) \) and a small monolingual benefit in RT studies \( (n = 8) \). However, no other factors (i.e., participant age, stimulus type, attention
components) resulted in significant differences or interactions. In answer to the question in the title, a bilingual advantage in auditory attention, when measured using standardized tests, likely does not exist or is restricted to certain conditions (e.g., accuracy measure or more proficient bilinguals). To be fair in our interpretation of this finding, two points should be considered: (1) a very heterogeneous bilingual population was synthesized across studies, wherein bilingualism was poorly assessed; (2) most standardized tests are developed for monolingual English speakers, thus they might not be suitable to measure bilinguals. As far as we know, there is no clear theoretical framework where we could place our results directly or interpret possible bilingual advantages [50]. Our work may contribute to building such models by highlighting the relation between bilingualism and auditory attention.

In the reviewed papers, standardized auditory attention tests were administered to participants as young as five years of age, and to children (5-11 years) more often than to young adolescents (12-14 years). Monolingual children seemingly had faster RTs, but bilingual children showed somewhat higher accuracy. Prior work from other domains than auditory attention has presented mixed findings about the RT difference between groups. On one hand, Bialystok et al. [51] used the Simon task—a non-linguistic interference task—and observed faster latency in bilingual children, although this effect was later only found when the demands for inhibitory control were high [52]. On the other hand, some language processing tasks report longer RTs in bilingual adults. For example, bilinguals have slower RTs to target words during a lexical decision task [53], and perform more slowly in a picture naming task than monolinguals [54]. These linguistic tasks specifically measure cross-language interference and switching, and bilinguals’ slower RTs have been often associated with their need to navigate more than one language system.
In our study, bilingual children seemed to have slightly more accurate responses, but at the cost of longer latency relative to their monolingual peers. However, these findings should be interpreted with caution, because most accuracy and RT measures were derived from different studies across the reviewed articles. Only Simonis et al. [38] reported both measures, but neither accuracy nor latency showed group differences between monolingual and bilingual children in their study. To draw a firm conclusion on the relationship between accuracy and RTs, one should only consider studies that report both measures within the same group.

With regards to the measured auditory attention components, selective attention and sustained attention are mostly assessed, followed by executive control. While one may argue that it is hard to disentangle these components and measure them independently, we distinguished them by considering the specific tasks used by investigators and the attention model developed by Posner and colleagues [16]. We also noticed inconsistency across publications when reporting the measured components with a given test. For example, using the auditory section of IVA + Plus, Krizman et al. [33] measured “attentional control,” whereas Boerma et al. [29] assessed “sustained attention.” This observation is in line with Williams et al. [6] who also found variation in targeted cognitive abilities for the same task in bilingualism research.

Moreover, rarely were sufficient details about participants’ bilingual background reported. This is problematic because differences in language experience (e.g., age of acquisition, proficiency; [4, 55]) can affect cognitive performance. For example, more proficient bilinguals tend to exhibit a cognitive advantage. Consistent with this, we observed a trend of the bilingual advantage toward adolescents relative to younger children, which can be explained by the fact that bilingual children become more proficient in their languages when growing older.
In addition, bilingualism was poorly assessed and hardly deemed a continuum. We identified two common themes: (1) most language background questionnaires fail to assess bilingualism comprehensively (i.e., including the use of, exposure to, and proficiency of each language; see [32]); (2) bilingual participants are usually recruited as a comparison group to monolinguals, without explaining inclusion criteria. This practice overlooks the nuances within bilinguals. For example, an English L2 learner from Spain is not comparable to a Spanish heritage speaker from the U.S., although both considered as Spanish-English bilinguals. Thus, variance in different bilingual communities should be considered, captured, and reported in future work.

In future studies, bilingualism should be considered as a continuum, and bilingual variables should be appropriately assessed and reported. This will facilitate our understanding of how various facets of bilingualism interact with auditory attention development. Moreover, further investigation should test monolinguals and bilinguals across age groups and report both accuracy and RTs. Given that most tests are designed for monolingual English speakers, we also advocate for developing tests that are normed to assess both monolinguals and bilinguals. Lastly, there was little consistency regarding the targeted attention components even for the same test across publications, which may be due to the lack of a clear definition of attention in the field itself. Therefore, we need to consider further what attention along with its components refers to in (bilingualism) research and theoretical models.
References


48. Foy JG, Mann VA. Bilingual children show advantages in nonverbal auditory executive


Supporting information

S1 Checklist. PRISMA 2020 checklist.

S1 Fig. Forest plot stratified by participant age groups, based on random-effects model analysis. Green lines represent accuracy studies, and blue lines represent RT studies. A bilingual advantage in young adolescents is shown.

S2 Fig. Contour-enhanced funnel plot. Effect size (i.e., standardized mean difference or SMD) is plotted against its standard error. Each dot represents an individual study. Grey-shaded areas indicate different \( p \)-value intervals.

S1 Table. Search terms used in the electronic databases under the concept of “auditory attention”.

S2 Table. Search terms used in the electronic databases under the concept of “population”.

S3 Table. Search terms used in the electronic databases under the concept of “study types and methods”.

S4 Table. Mixed-effects model summary, with test measure as the moderator.

S5 Table. Subgroup analysis result, stratified by test measure.

S6 Table. Mixed-effects model summary for accuracy studies, with participant age as the moderator.

S7 Table. Mixed-effects model summary for accuracy studies, with stimulus type as the moderator.

S8 Table. Mixed-effects model summary for accuracy studies, with attention components as the moderator.

S9 Table. Mixed-effects model summary for RT studies, with participant age as the moderator.
S10 Table. Mixed-effects model summary for RT studies, with stimulus type as the moderator.

S11 Table. Mixed-effects model summary for RT studies, with attention components as the moderator.

S12 Table. Egger’s regression test result.
Identification of studies via databases and manual search

Identification:
- Electronic database searches:
  - OVID Medline (n = 8605)
  - OVID PsycInfo (n = 6055)
  - EBSCO CINAHL (n = 5455)
- Records identified in total (n = 20115)

Screening:
- Titles / abstracts screened (n = 16628)
- Irrelevant records excluded (n = 15491)
- Records excluded:
  - Wrong population: not focusing on typically developing children (n = 467)
  - Wrong outcomes: not focusing on auditory attention (n = 500)
  - Wrong study type: not empirical studies (n = 36)
  - Not written in English (n = 49)
  - Study unavailable (n = 13)
  - Duplicate (n = 2)

Included:
- Full-text articles assessed for eligibility (n = 1137)
- Articles included in review (n = 70)
- Articles fitting current inclusion criteria (n = 3)
- Articles identified through manual search (n = 9)
- Articles included in the analysis: (n = 12)
  - Studies included in the analysis: (n = 20)
<table>
<thead>
<tr>
<th>Study</th>
<th>Monolingual Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Hedges’ g</th>
<th>SMD</th>
<th>95%-CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preschoolers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boerma et al. 2017</td>
<td>32</td>
<td>2.33</td>
<td>0.9000</td>
<td>32</td>
<td>2.05</td>
<td>0.8000</td>
<td>-0.32</td>
<td>[-0.82; 0.17]</td>
<td>4.7%</td>
<td></td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (1)</td>
<td>30</td>
<td>1245.79</td>
<td>388.8200</td>
<td>30</td>
<td>1142.52</td>
<td>345.9900</td>
<td>-0.28</td>
<td>[-0.79; 0.23]</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (2)</td>
<td>30</td>
<td>1400.64</td>
<td>344.1700</td>
<td>30</td>
<td>1344.98</td>
<td>455.9200</td>
<td>-0.14</td>
<td>[-0.64; 0.37]</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (3)</td>
<td>30</td>
<td>833.73</td>
<td>381.7700</td>
<td>30</td>
<td>732.98</td>
<td>441.6100</td>
<td>-0.24</td>
<td>[-0.75; 0.27]</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>Foy &amp; Mann 2014 (4)</td>
<td>30</td>
<td>988.54</td>
<td>402.8000</td>
<td>30</td>
<td>788.11</td>
<td>329.6200</td>
<td>-0.54</td>
<td>[-1.05; -0.02]</td>
<td>4.5%</td>
<td></td>
</tr>
<tr>
<td>Kwakkel et al. 2021</td>
<td>80</td>
<td>15.31</td>
<td>11.2300</td>
<td>89</td>
<td>17.22</td>
<td>13.3800</td>
<td>0.15</td>
<td>[-0.15; 0.46]</td>
<td>6.3%</td>
<td></td>
</tr>
<tr>
<td><strong>Random effects model</strong></td>
<td></td>
<td></td>
<td></td>
<td>232</td>
<td>241</td>
<td></td>
<td></td>
<td>-0.18</td>
<td>[-0.45; 0.09]</td>
<td>29.3%</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 27\%, \tau^2 = 0.0279, p = 0.23$

<table>
<thead>
<tr>
<th>Study</th>
<th>Monolingual Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Hedges’ g</th>
<th>SMD</th>
<th>95%-CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary school-aged children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barbu et al. 2019</td>
<td>57</td>
<td>875.36</td>
<td>152.4900</td>
<td>59</td>
<td>817.20</td>
<td>141.9300</td>
<td>-0.39</td>
<td>[-0.76; -0.02]</td>
<td>5.7%</td>
<td></td>
</tr>
<tr>
<td>Garratt &amp; Kelly 2007</td>
<td>27</td>
<td>10.20</td>
<td>2.7000</td>
<td>27</td>
<td>9.60</td>
<td>2.7000</td>
<td>-0.22</td>
<td>[-0.75; 0.32]</td>
<td>4.4%</td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (1)</td>
<td>25</td>
<td>52.70</td>
<td>9.5000</td>
<td>24</td>
<td>51.20</td>
<td>8.5000</td>
<td>-0.16</td>
<td>[-0.72; 0.40]</td>
<td>4.2%</td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (2)</td>
<td>23</td>
<td>58.50</td>
<td>2.0000</td>
<td>27</td>
<td>58.10</td>
<td>2.1000</td>
<td>-0.19</td>
<td>[-0.75; 0.37]</td>
<td>4.2%</td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (3)</td>
<td>25</td>
<td>49.30</td>
<td>11.8000</td>
<td>24</td>
<td>52.70</td>
<td>9.4000</td>
<td>0.31</td>
<td>[0.25; 0.88]</td>
<td>4.2%</td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015 (4)</td>
<td>23</td>
<td>66.00</td>
<td>4.4000</td>
<td>27</td>
<td>64.70</td>
<td>6.1000</td>
<td>-0.24</td>
<td>[-0.80; 0.32]</td>
<td>4.2%</td>
<td></td>
</tr>
<tr>
<td>Nicolay &amp; Poncelet 2013</td>
<td>51</td>
<td>820.63</td>
<td>127.2500</td>
<td>53</td>
<td>751.87</td>
<td>92.4600</td>
<td>-0.62</td>
<td>[-1.01; -0.22]</td>
<td>5.5%</td>
<td></td>
</tr>
<tr>
<td>Nicolay &amp; Poncelet 2015</td>
<td>50</td>
<td>824.50</td>
<td>109.0000</td>
<td>51</td>
<td>752.80</td>
<td>90.8000</td>
<td>-0.71</td>
<td>[-1.11; -0.31]</td>
<td>5.4%</td>
<td></td>
</tr>
<tr>
<td>Strydom et al. 2022</td>
<td>20</td>
<td>96.30</td>
<td>3.2600</td>
<td>20</td>
<td>96.60</td>
<td>3.0500</td>
<td>0.09</td>
<td>[-0.53; 0.71]</td>
<td>3.8%</td>
<td></td>
</tr>
<tr>
<td><strong>Random effects model</strong></td>
<td></td>
<td></td>
<td></td>
<td>301</td>
<td>312</td>
<td></td>
<td></td>
<td>-0.29</td>
<td>[-0.53; -0.04]</td>
<td>41.7%</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 39\%, \tau^2 = 0.0423, p = 0.11$

<table>
<thead>
<tr>
<th>Study</th>
<th>Monolingual Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Hedges’ g</th>
<th>SMD</th>
<th>95%-CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Young adolescents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krizman et al. 2012</td>
<td>25</td>
<td>69.00</td>
<td>21.6500</td>
<td>23</td>
<td>88.00</td>
<td>18.4600</td>
<td>0.93</td>
<td>[0.33; 1.52]</td>
<td>3.9%</td>
<td></td>
</tr>
<tr>
<td>Krizman et al. 2014</td>
<td>27</td>
<td>69.19</td>
<td>21.8800</td>
<td>27</td>
<td>84.37</td>
<td>18.2900</td>
<td>0.74</td>
<td>[0.19; 1.29]</td>
<td>4.3%</td>
<td></td>
</tr>
<tr>
<td>Simonis et al. 2020 (1)</td>
<td>129</td>
<td>0.86</td>
<td>0.3400</td>
<td>156</td>
<td>0.91</td>
<td>0.2800</td>
<td>0.16</td>
<td>[-0.07; 0.40]</td>
<td>6.9%</td>
<td></td>
</tr>
<tr>
<td>Simonis et al. 2020 (2)</td>
<td>153</td>
<td>0.90</td>
<td>0.2900</td>
<td>173</td>
<td>0.92</td>
<td>0.2600</td>
<td>0.07</td>
<td>[-0.14; 0.29]</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td>Simonis et al. 2020 (3)</td>
<td>153</td>
<td>671.88</td>
<td>193.2700</td>
<td>173</td>
<td>679.34</td>
<td>174.1900</td>
<td>0.04</td>
<td>[-0.18; 0.26]</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td><strong>Random effects model</strong></td>
<td></td>
<td></td>
<td></td>
<td>487</td>
<td>552</td>
<td></td>
<td></td>
<td>0.30</td>
<td>[-0.17; 0.77]</td>
<td>29.0%</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 68\%, \tau^2 = 0.0876, p = 0.02$

<table>
<thead>
<tr>
<th>Study</th>
<th>Monolingual Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Hedges’ g</th>
<th>SMD</th>
<th>95%-CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random effects model</strong></td>
<td></td>
<td></td>
<td></td>
<td>1020</td>
<td>1105</td>
<td></td>
<td></td>
<td>-0.09</td>
<td>[-0.27; 0.10]</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 66\%, \tau^2 = 0.0931, p < 0.01$

**Prediction interval**

Heterogeneity: $I^2 = 66\%, \tau^2 = 0.0931, p < 0.01$

Test for subgroup differences: $\chi^2 = 8.78, df = 2 (p = 0.01)$