



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Learning how to communicate: Does exposure to multiple languages promote children's pragmatic abilities? A meta-analytic review

Elise van Wonderen^{*}, Kimberley Mulder, Judith Rispens, Josje Verhagen

Amsterdam Center for Language and Communication, the Netherlands

ARTICLE INFO

Keywords:

Pragmatic abilities
Multilingual children
Multilingual advantage
Meta-analysis

ABSTRACT

Despite the often-reported finding that multilingual children may temporarily possess less advanced lexical or grammatical skills in at least one of their languages than monolingual peers, recent studies have found that exposure to multiple languages benefits children's pragmatic development. To assess the generalizability of these findings, we conducted a meta-analysis of 29 studies that investigated various pragmatic abilities in both multilingual and monolingual children. In addition, we classified the investigated pragmatic abilities into three broad categories: (i) sensitivity to non-verbal information, (ii) understanding non-literal language, and (iii) informativeness in referencing. We found no effect of multilingualism on pragmatic abilities overall and also not in any of the categories separately. However, based on visual inspection of the effect sizes, we cannot rule out that multilingual children sometimes rely more on non-verbal cues during communication compared to their monolingual peers. We recommend future studies to formulate and test more specific hypotheses and to move away from the focus on a multilingual advantage.

1. Introduction

Multilingual children (i.e., children with exposure to *at least* two languages or language varieties) sometimes temporarily lag behind their monolingual peers in terms of their vocabulary and morphosyntactic development in at least one of their languages (Hoff, 2021). Despite this relative delay in language-specific skills, some studies have found that multilingual children outperformed their monolingual peers in their pragmatic ability (e.g., Fan et al., 2015; Genesee et al., 1975; Yow & Markman, 2011a), which we define as the capacity to (i) adjust language or non-verbal communication according to the listener or the social or physical context, or (ii) use information about the speaker or the social and physical context to infer a speaker's communicative intentions, needs or desires (definition adapted from Turkstra et al., 2017). Support for a positive effect of multilingualism on children's pragmatic abilities comes from studies showing that under certain conditions, multilingual children rely more on non-verbal cues (i.e., eye gaze or gestures) during reference resolution than their monolingual peers (Brojde et al., 2012; Groba et al., 2018; Verhagen et al., 2017; Yow, 2015; Yow et al., 2017; Yow & Markman, 2011a; but see Gangopadhyay & Kaushanskaya, 2020, 2021). In addition, multilingual children have been found to more often take a speaker's visual perspective into account (Fan et al., 2015; Genesee et al., 1975; Liberman et al., 2017; Yow & Markman, 2015), repair miscommunications (Wermelinger et al., 2017; but see Comeau et al., 2010), adapt their (non-)

^{*} Correspondence to: Spuistraat 134, 1012 VB Amsterdam, the Netherlands.
E-mail address: e.vanwonderen@uva.nl (E. van Wonderen).

<https://doi.org/10.1016/j.cogdev.2023.101384>

Received 8 June 2023; Received in revised form 2 August 2023; Accepted 3 October 2023

Available online 19 October 2023

0885-2014/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

verbal communication to the needs of their interlocutor (Gampe et al., 2019), and detect violations of Grice's conversational principles (i.e., be truthful, be relevant, avoid ambiguity, and provide as much information as required; Siegal et al., 2009, 2010).

The finding that exposure to multiple languages may promote pragmatic development is important, as pragmatic abilities are essential for successful communication but also relatively difficult to learn (Turkstra et al., 2017). Since language is highly underspecified – meaning that what is said is not sufficient to support a unique interpretation – it is left to the listener to infer the speaker's communicative intent by integrating the verbal message with various sources of extralinguistic information (Ariel, 2010; Carston, 2002). A clear example of underspecified utterances that require pragmatic inferences are indirect requests, which can only be understood by relying on the physical context and assumptions about the speaker's needs and desires (e.g., understanding that "it is cold in here" could be a request to close the window). However, even something as basic as mapping referential expressions like 'this' or 'we' to their appropriate referents requires considering the physical and discourse context (which referents are currently present or previously mentioned?), as well as the speaker's visual perspective (which referents can the speaker see?) and non-verbal cues (e.g., the speaker's eye gaze or pointing gestures). Learning to successfully integrate all this information takes time, as demonstrated by the protracted development of pragmatics relative to vocabulary and morphosyntax (e.g., O'Neill, 2012; Turkstra et al., 2017; Zufferey, 2020). Indeed, even adults sometimes fail to consider or correctly assess the knowledge state or visual perspective of their conversation partner, which sometimes results in miscommunications (e.g., Keysar et al., 2000; Keysar & Henly, 2002; Savitsky et al., 2011).

In the current study we conducted a meta-analysis to test whether, overall, multilingual children show advanced pragmatic abilities compared to monolingual children and, if so, whether this varies for different pragmatic abilities. Before elaborating on the goals and design of the current study, we first review *why* exposure to multiple languages may affect children's pragmatic development and we further examine the evidence to date.

1.1. Why exposure to multiple languages may affect pragmatic development

It has been hypothesized that multilingual children show advanced pragmatic abilities due to the characteristics of the communicative experiences they typically encounter. More specifically, three hypotheses have emerged from the literature, which we call the Monitoring Hypothesis, the Compensation Hypothesis, and the Mediation Hypothesis. We now discuss all three hypotheses in turn to give a complete overview of the literature. However, the information provided in the available studies only allowed us to test the Compensation Hypothesis in the current meta-analysis. We come back to this in the Discussion.

The Monitoring Hypothesis is grounded in the idea that multilingual children must monitor who can speak and understand which language to avoid communication breakdowns. Because of monitoring their interlocutors' language abilities and preferences, multilingual children may pay closer attention to their interlocutors in general, resulting in increased sensitivity to their knowledge state and communicative needs (e.g., Brojde et al., 2012; Yow & Markman, 2011b, 2016). Although it is currently unknown whether this explains the found effects of multilingualism on pragmatic development, multilingual children do seem to infer their interlocutors' language preferences from a very young age onwards and to adapt their language choices accordingly (e.g., Comeau et al., 2003; Genesee, 2006; Genesee et al., 1996; but see Schott et al., 2023).

Alternatively, the Compensation Hypothesis suggests that paying attention to various communicative and contextual cues could serve as a compensatory strategy to mitigate any gaps in multilingual children's vocabularies or morphosyntactic skills (e.g., Siegal et al., 2009; Verhagen et al., 2017; Wermelinger et al., 2020). Multilingual children often possess a smaller vocabulary than monolingual peers in at least one of their languages (Hoff, 2021; Oller et al., 2007; cf. Thordardottir, 2011) and may therefore have to rely more on contextual cues and non-verbal means of communication, both when understanding a speaker and when producing language themselves. Preliminary evidence supporting this hypothesis comes from Verhagen et al. (2017), who showed that multilingual children relied more on non-verbal cues in their weaker language (i.e., the language in which they knew fewer words) than in their stronger language.

Finally, the Mediation Hypothesis proposes that the effect of multilingualism on pragmatic abilities is mediated by multilingual children's advanced executive functions (EF) or Theory of Mind (ToM). EF includes cognitive processes such as attentional control, inhibitory control and working memory, while ToM refers to the ability to attribute mental states (such as thoughts, beliefs or desires) to oneself and others and to predict and explain people's behavior based on these mental states (Premack & Woodruff, 1978). It has often been suggested that mentally representing multiple languages and routinely selecting or inhibiting one of these languages may enhance EF (e.g., Bialystok, 2011, 2017; Bialystok et al., 2012), although robust empirical evidence for this claim is lacking (e.g., Gunnerud et al., 2020; Lowe et al., 2021; Paap et al., 2015). However, if multilingual children show an advantage in EF at some point in their development, this could also benefit their pragmatic development. This is because pragmatic inferences typically require keeping information in memory and integrating multiple cues (e.g., Yow & Markman, 2015), as well as sometimes inhibiting a conflicting interpretation or perspective (e.g., Fan et al., 2015; Yow & Markman, 2011b). Likewise, multilingual children are often considered to possess enhanced ToM (e.g., Farhadian et al., 2010; Kovács, 2009), although the evidence for this claim is based on a limited number of studies (Schroeder, 2018). If multilinguals display better ToM this may in turn enhance their pragmatic abilities, given that many pragmatic inferences recruit ToM to at least some extent (O'Neill, 2012).

Clearly, none of these hypotheses are mutually exclusive. For example, both lexical gaps *and* incorrect assumptions about which languages the interlocutor understands may result in communication breakdowns, which, in turn, may stimulate multilingual children to become more attentive to the communicative cues of others to avoid such breakdowns.

1.2. Mixed evidence

Although there are many reasons why exposure to multiple languages may affect children's pragmatic development, the evidence supporting a multilingual advantage in pragmatic abilities is mixed. For example, several studies on reference resolution found that multilingual children are more sensitive to non-verbal cues such as eye gaze and pointing than monolingual children, but only for some of the experimental conditions tested. Yow and Markman (2011a), for instance, found that multilingual children relied more strongly on non-verbal cues than monolingual children when eye gaze was pitted against body-distal information (i.e., how close the speaker was to each object), but not when cues were consistent or when pointing gestures were used instead of eye gaze to indicate a referent. Furthermore, Verhagen et al. (2017) found that multilingual children relied more strongly on a speaker's pointing gestures than monolingual children, but only when a familiar label was used for a novel object and not when a novel label was used for a familiar object.

Other studies failed to find a difference between monolingual and multilingual children altogether. For example, no effect of multilingualism was found in recent experiments assessing sensitivity to non-verbal cues during word learning using an eye-tracking paradigm (Gangopadhyay & Kaushanskaya, 2020, 2021). In addition, although multilingual children have been found to better detect violations of Gricean maxims (Siegal et al., 2009, 2010), this does not seem to result in better understanding of non-literal language, for which a violation of a Gricean maxim (e.g., be relevant) should lead to a non-literal reinterpretation of the utterance (Antoniou et al., 2020; Antoniou & Katsos, 2017; Schulze et al., 2020). Finally, uncertainty about which pragmatic abilities may be influenced by multilingualism is further exacerbated by large standard errors that accompany most of the reported significant effects, which might, among other things, be caused by small sample sizes, large interindividual variation or noisy measurement instruments (Cumming, 2013; Wilson Van Voorhis & Morgan, 2007). In sum, it is currently unclear whether multilingual children show advanced pragmatic abilities compared to monolingual children, and if so, for which pragmatic abilities and under which circumstances.

1.3. The present study

In this study we conducted a meta-analysis of all published studies that investigated pragmatic abilities in multilingual and monolingual children between the ages of 0 and 18 years. Our main aim was to investigate whether multilingualism enhances children's pragmatic abilities overall, that is, irrespective of the pragmatic ability investigated or other differences in methodology or sample characteristics. Second, we were interested in whether differences between multilingual and monolingual children vary across pragmatic abilities. To investigate this, we classified the pragmatic abilities into three broad categories based on the included studies (see Methods): (i) sensitivity to non-verbal information (including eye gaze, gestures and the interlocutor's visual perspective), (ii) understanding non-literal language (including indirect requests, novel metaphors and irony), and (iii) informativeness in referencing (including level of informativeness and ambiguity of referential expressions).

All three of these categories require adapting the production or comprehension of (non-)verbal communication to the interlocutor or the social or physical context. In the case of sensitivity to non-verbal information, children must consider the visual perspective or non-verbal cues of the interlocutor within a given context to correctly interpret the interlocutor's verbal message. For understanding non-literal language, it is important to consider extralinguistic cues such as prosody and facial expressions as well as the interlocutor's beliefs and desires and possibly the social or physical context to know whether an utterance is to be taken literally. Finally, regarding referential expressions, considering the interlocutor's perspective and knowledge state is crucial for successfully communicating the intended referent. For example, opting for a more informative noun phrase (e.g., *the brown horse*) as opposed to a less informative pronoun (e.g., *he*, *they*) decreases the chance of misunderstandings when the interlocutor does not have the same level of access to the target referent as the speaker.

In addition to our main analyses, we also coded each study for several design and participant characteristics to allow for exploratory analyses of variables that may moderate the effect of multilingualism on pragmatic abilities. Based on the information available in the included studies, we investigated two additional factors, namely children's receptive vocabulary knowledge and age. If multilingual children pay more attention to extralinguistic cues to compensate for gaps in their vocabulary knowledge, as predicted by the Compensation Hypothesis, then differences in pragmatic ability compared to monolingual children should only emerge when the multilingual children have a smaller vocabulary. We did not have specific hypotheses based on the literature regarding age, but we included this predictor in an exploratory analysis.

2. Methods

2.1. Literature search

Fig. 1 depicts a flow chart of the overall study selection process. Studies included in the analysis were obtained via two methods. First, we performed a database search in the Core Collection of Web of Science. To generate a list of search terms, we extracted key words from two book chapters that provide an overview of children's pragmatic development (McTear & Conti-Ramsden, 1992, Chapter 4; Zufferey, 2020). Search terms that did not yield additional relevant hits were deleted (but see Supplementary Materials S1 at <https://osf.io/ak9vr/> for a complete list), resulting in the following combination of search terms:

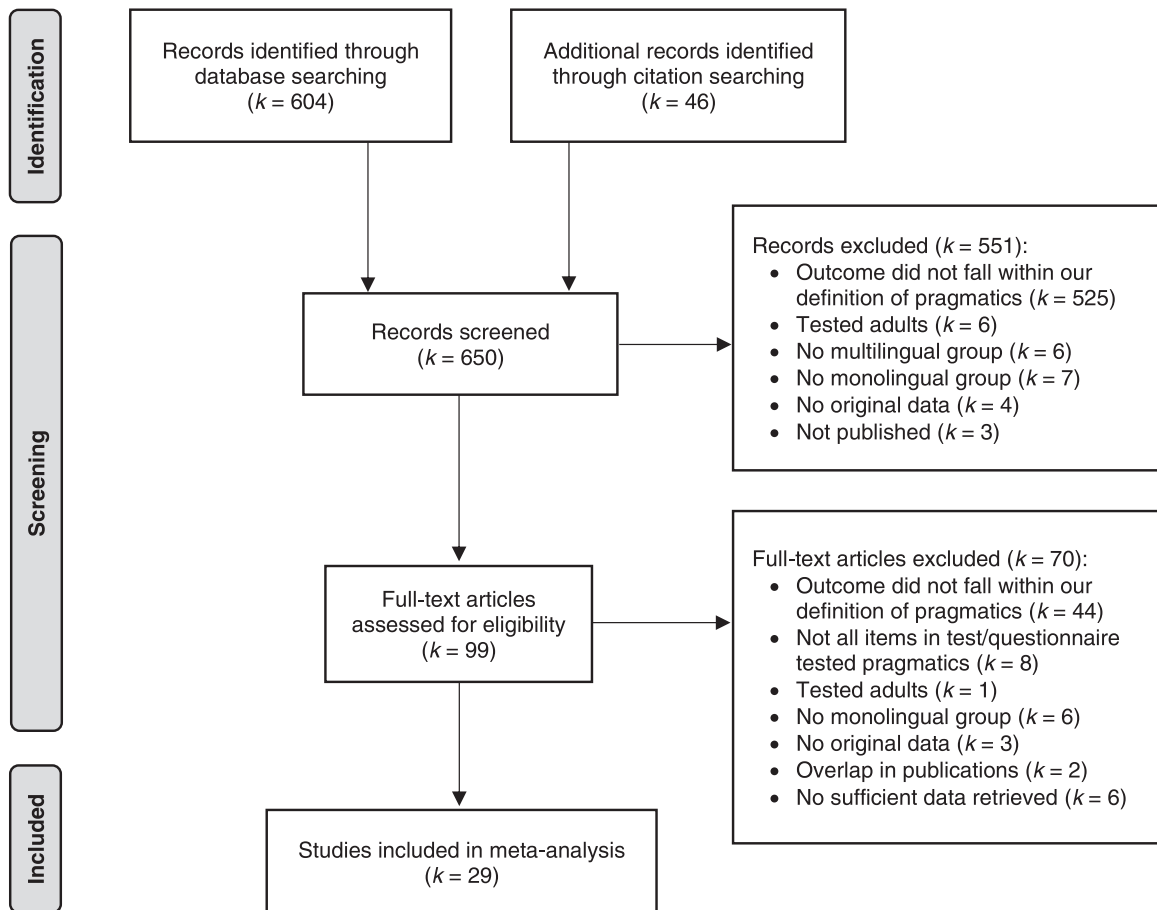


Fig. 1. PRISMA flowchart showing the study selection process (adapted from Moher et al., 2009).

TS= ((pragmatic* OR communicat* OR conversation* OR dialogue OR narrative* OR perspective taking OR theory of mind OR non-verbal cue* OR gesture* OR gaze OR face OR referent* OR definiteness) AND (bilingual* OR multilingual* OR second language) AND (monolingual*) AND (child* OR toddler* OR infant* OR adolescen*))

We included Theory of Mind as a search term, because what we consider to be part of pragmatics (e.g., taking a speaker's visual perspective into account when resolving ambiguous reference; Fan et al., 2015) has sometimes been included as a measure of Theory of Mind by others (c.f. the meta-analysis by Schroeder, 2018). The final search was conducted on March 22, 2023. As a second search strategy, we recursively performed a backward and forward reference search of the included studies, meaning that we scanned the reference lists of all included studies for relevant titles, as well as all academic records that cited the included studies (as indicated by Google Scholar) until no additional relevant studies were identified.

To assess eligibility, the first author read the titles and abstracts of all identified articles as well as the full text in case of doubt. If any doubts remained upon reading the full text, two co-authors read the methods section of the paper and a decision was made after elaborate discussion. To ensure that this yielded a reliable selection process, the last author independently coded 20% of all identified studies ($k = 130$). Only one of these studies initially yielded disagreement, which was resolved after reading the paper more carefully.¹

Studies were included if at least one outcome within the article assessed pragmatic abilities according to our definition in both monolingual and multilingual children (aged 0–18 years). We considered children to be multilingual when they had exposure to at least two languages or language varieties. We used such a broad definition of multilingualism because there are currently no specific hypotheses on which types of multilinguals should or should not show pragmatic benefits (rather, the hypothesis seems to be that any exposure to a second language or language variety may yield this advantage; e.g., Fan et al., 2015). In addition, we included both typically and atypically developing children as we did not have an a priori reason to expect that multilingualism benefits only one of

¹ The one paper where the first and last author initially disagreed was Zvaigzne et al. (2019), which the final author coded as "included." However, upon carefully reading the paper it became clear that results were not reported separately for the crucial manipulation (i.e., whether the tester could see the child's gestures during retelling or not). Therefore, we decided to include Zvaigzne et al. (2008) instead: this paper reports on the same data and does report the results of the crucial manipulation.

these two groups. Studies were excluded if they failed to meet the inclusion criteria above, were not published, or did not contain original data (e.g., meta-analyses, reviews). If publications overlapped because studies were reported both in a conference proceeding as well as in a later journal publication, we only included the journal publication.

Given that pragmatic competence is a latent construct which cannot be measured directly, there are many different ways in which pragmatic competence can be measured. We therefore included a wide range of tasks and outcome measures with the only condition that the outcome had to be an indicator of pragmatic competence according to our definition. To decide whether an outcome matched our definition of pragmatics, we assessed on a case-by-case basis whether the outcome of interest was truly dependent on a specific interlocutor or social or physical context and invoked pragmatic reasoning capabilities rather than being the result of a formulaic chunk or dependent on grammatical knowledge. Examples of phenomena that did not meet these criteria are fixed non-literal expressions (e.g., idioms and conventional metaphors), scripted language (e.g., saying “John speaking” when picking up the phone), conventionalized indirect requests (e.g., “Can you pass me the salt?”), and both conventional and generalized conversational implicatures such as scalar quantifiers (e.g., “some” implying “not all”). The reason we excluded these phenomena is because their interpretation does not *necessarily* rely on pragmatic inferences; rather, they may be learned and stored as formulaic chunks whose meaning can be accessed directly from the lexicon (Ariel, 2010). Likewise, we excluded phenomena such as definiteness and pronoun realization as their expression and interpretation requires language-specific, grammatical knowledge.² Our motivation for excluding specific studies is documented for each study separately in the spreadsheet called *database_search*, and our reasons for excluding individual outcomes within included studies are given under the comments tab in the spreadsheet called *coded_studies* (both available at <https://osf.io/ak9vr/>).

In the end, 35 studies met our criteria. For 18 of these, we needed additional data to compute an effect size. For ten studies, authors sent us the required data and for another two studies we estimated data from figures in the paper using the WebPlotDigitizer program (Rohatgi, 2021). This means there were six studies for which we did not receive sufficient data to compute an effect size and we thus had to exclude those studies. Our final dataset therefore consisted of 29 studies.

2.2. Data coding

Most studies reported multiple comparisons between multilingual and monolingual children. For example, many studies included multiple pragmatic abilities (e.g., understanding indirect speech, novel metaphors, and irony), conditions (e.g., reference resolution based on a speaker’s eye gaze vs. a pointing gesture) or outcome measures (e.g., reaction times and target object selection). In addition, some studies tested multiple age groups or groups with varying levels of exposure to a second or third language. We entered each comparison as a separate row in a spreadsheet, yielding 109 datapoints in total.

Subsequently, the first author coded each datapoint for several design and participant characteristics, including children’s age and language background, the pragmatic ability tested and a description of the outcome of interest, as well as additional tasks administered in the study and for which of these measures there were significant differences between the monolingual and multilingual group.

After coding the data, we classified the outcomes into different categories. Given that there is no consensus on how pragmatic abilities should be classified (see, e.g., O’Neill, 2012), we used a bottom-up approach that, in our view, maximized differences between categories and minimized differences within categories, while also maximizing the number of datapoints per category (see Table 1; for details per study see the *coded_studies* file at OSF). This resulted in the following three categories: (i) sensitivity to non-verbal information, (ii) understanding non-literal language, and (iii) informativeness in referencing. Note that the first two categories focus on comprehension, whereas the third category focuses on production. There were five studies that did not fit any of the categories and were therefore labelled as *other*. For most ($n = 26$) studies, this classification was relatively straightforward, but there were three studies that were difficult to classify (i.e., Wermelinger et al., 2020; Yow & Markman, 2011b; Zvaigzne et al., 2008). For example, in Yow and Markman (2011b), children had to infer a speaker’s emotion based on prosodic cues that mismatched with the literal content of the message. This task is very similar to interpreting irony where what is said is the opposite of what is meant. We therefore classified this study as measuring the understanding of non-literal language. However, one could also consider prosody as a non-verbal cue and categorize the study as measuring sensitivity to non-verbal information. We will return to this issue in the Discussion.

2.3. Effect sizes

For each datapoint, we computed the standardized mean difference Hedges’ g between the scores of the monolingual and multilingual group. Hedges’ g is calculated by first computing Cohen’s d , and then multiplying this value by a correction factor that accounts for the biasing effect of small sample sizes (Borenstein, 2009; Hedges & Olkin, 1985). The effect sizes were calculated such that positive values indicated a multilingual advantage on the pragmatic ability tested, whereas negative values indicated a monolingual advantage.

For most studies ($k = 27$), Cohen’s d was calculated by dividing the mean difference by the pooled standard deviation of the multilingual and monolingual group based on the raw means and standard deviations. If multiple multilingual groups were compared to the same monolingual group, we first computed a pooled standard deviation over all groups and used this as the denominator for Cohen’s d (Pustejovsky, 2015). For one study, we used the reported t -statistic and group sample sizes to compute Cohen’s d because

² Note that we initially included definiteness as a search term, because definiteness was mentioned in the two book chapters that we used to generate a list of search terms. However, due to the reasons mentioned in the main text, we decided that studies on definiteness did not meet our inclusion criteria.

Table 1Classification of the pragmatic abilities investigated in the included studies (k = number of studies, n = number of effect sizes).

| Category | k | n |
|--|-----------------|-----|
| 1. Sensitivity to non-verbal information | 14 ^a | 45 |
| Reference resolution or word learning based on eye gaze | 6 | 20 |
| Reference resolution or word learning based on gestures | 6 | 15 |
| Reference resolution based on speaker's visual perspective | 4 | 10 |
| 2. Understanding non-literal language | 5 ^a | 20 |
| Understanding/processing novel metaphors | 2 | 6 |
| Understanding/processing indirect speech | 2 | 8 |
| Understanding/processing irony | 1 | 4 |
| Inferring speaker's emotion based on prosody when literal meaning of the sentence provides conflicting information | 1 | 1 |
| Understanding unconventional figurative language, indirect speech and sarcasm | 1 | 1 |
| 3. Informativeness in referencing | 6 | 19 |
| Proportion of ambiguous referential expressions in narrative tasks | 3 | 12 |
| Informativeness of referential expressions in narrative or referential communication tasks | 2 | 6 |
| Informativeness of iconic gesture in a referential communication task | 1 | 1 |
| 4. Other | 5 | 25 |
| Adaptation to communicative needs of interlocutor (repair communication breakdowns or adapt helping behavior to expressed needs) | 2 | 6 |
| Detection of violations of Gricean maxims | 2 | 18 |
| Pragmatic appropriateness of elicited speech acts | 1 | 1 |

Note. ^aThis is not the sum of the subcategories because some studies tested multiple pragmatic abilities.

raw means and standard deviations were not available (Lipsey & Wilson, 2001). In another study the reported outcome was the proportion of children in each group who showed a certain behavior. For this study, an approximation of Cohen's d was calculated by taking the difference of the arcsine transformed proportions per group (Lipsey & Wilson, 2001). For an overview of the used formulas, see Supplementary Materials S2.

Finally, it is important to note that in some studies ($k = 5$) the outcome of interest was children's differential behavior across two different conditions. For example, Gampe et al. (2019) compared children's helping behavior when a puppet expressed that it did or did not want any help with looking for a missing object. The outcome of interest in this study was whether children adapted their behavior according to the puppet's communicative needs. For studies like this, we first computed the mean difference between conditions for each group separately and requested the standard deviations of the difference scores from the authors or computed these from the raw data when available. We then computed Cohen's d based on the difference between the groups' mean difference scores.

2.4. Statistical analysis

We ran seven separate meta-analytic models: (i) a model on all data to obtain the mean effect size across categories, (ii) three separate models for the different pragmatic categories (i.e., sensitivity to non-verbal information, understanding non-literal language and informativeness in referencing), (iii) two separate models for datapoints where the monolingual and multilingual group were matched versus not matched on their receptive vocabulary knowledge of the language of testing, and (iv) a model on all data with age as a predictor. For each analysis, we used a correlated and hierarchical effects model (Pustejovsky & Tipton, 2022; preregistered at <https://osf.io/vja5x>), using the R-package *metafor* (Viechtbauer, 2010). This model was selected to deal with data dependencies caused by (i) multiple studies coming from the same research group, (ii) studies reporting on multiple subsamples (e.g., multiple age groups), (iii) studies comparing multiple multilingual groups to the same monolingual group, and (iv) studies reporting on multiple outcome measures. The first two sources of dependency create a hierarchical structure in the data, as effect sizes coming from the same studies or research groups may be more similar due to similarities in the abilities investigated, tasks, testing protocols, data analysis procedures, et cetera. The latter two sources of dependency also result in correlated sampling errors between effect sizes, as these effect sizes come from (partly) the same participants (Hedges et al., 2010; Pustejovsky & Tipton, 2022).

Our models dealt with these dependencies by (i) approximating the covariance matrix between effect sizes, and (ii) including random effects for cluster variables. Following Pustejovsky and Tipton (2022), we assumed a correlation of $r = 0.60$ for effect sizes coming from (partly) the same children and conducted sensitivity analyses by varying the assumed correlation between $r = 0.00$ and $r = 0.95$. These analyses showed that choice of r hardly impacted the results, with a maximum change of 0.07 in average effect size estimates and a maximum change of 0.02 in standard error estimates across the investigated range of r (see Supplementary Materials S3). Note that in our pre-registration we expressed our intention to, whenever available, use correlations between task scores as a proxy of the correlation between the sampling errors of effect sizes (Pustejovsky & Tipton, 2022). For 7 of the 29 studies, we could compute or retrieve (some of) these empirical correlations, but inserting them in the covariance matrix resulted in a matrix that was

not positive definite. We therefore did not include empirical correlations, and only varied the assumed correlation between effect sizes as described above.

For the random effects, we initially included random intercepts for effect sizes nested in data collection nested in research group. Studies that (partly) tested the same children were coded as belonging to the same data collection,³ whereas studies were coded as belonging to the same research group when at least one of the authors overlapped (for 15 of the 16 research groups, two or more authors overlapped). However, it turned out that within a research group, data collection did not explain any additional variance (i.e., the variance associated with data collection was zero). Data collection was therefore dropped from the analyses, such that the analyses reported here included random intercepts for effect size nested in research group only. Note that in our preregistration we had not considered research group as a cluster variable. However, after our sample of studies was obtained, it turned out that even though there were 27 data collections (published in 29 studies), these only came from 16 different research groups. Given that this created additional dependencies between effect sizes, we included research group as the highest cluster. The pre-registered analyses with data collection as the highest cluster are given in Supplementary Materials S4 and briefly summarized in the sections below.

Finally, we used cluster-robust variance estimation with clustering at the level of research group (Hedges et al., 2010; Tipton, 2015) using the *clubSandwich* package (Pustejovsky, 2022). Cluster-robust variance estimation ensures that unbiased effect size estimates and robust standard errors are obtained even when the model is slightly misspecified; for example, when there are additional dependencies in the data not captured by the random effects or when the assumed correlations between effect sizes do not reflect the actual correlations.

3. Results

3.1. Descriptive results

The final dataset included 29 articles from 16 research groups, reporting on 39 independent participant samples, yielding a total of 109 effect sizes. These effect sizes compared the pragmatic abilities of 1115 unique multilingual children to 931 unique monolingual children aged 1–12 years (median_{mean age} = 5;3), with sample sizes ranging from 5 to 87 per group (median = 22).⁴ Children's pragmatic abilities were mostly tested via forced-choice ($k = 15$ studies), elicited production ($k = 6$), narratives ($k = 4$), and/or looking-while-listening tasks ($k = 4$).

Regarding participant characteristics, most effect sizes ($n = 99$) came from typically developing children, while a few came from children with autism spectrum disorder ($n = 5$) or developmental language disorder ($n = 5$). Most effect sizes ($n = 85$) involved multilingual children speaking two languages, while a few effect sizes came from children speaking two dialects ($n = 10$) or three or more languages ($n = 4$). The remaining effect sizes came from mixed groups (i.e., groups containing a mix of bilingual, multilingual and bilingual children). Finally, effect sizes came from simultaneous ($n = 39$) or sequential ($n = 12$) multilingual children or a mix of simultaneous and sequential multilingual children ($n = 29$), with information on the age of onset of multilingualism missing for the remaining 29 effect sizes.

3.2. Publication bias

We assessed the presence of publication bias by means of funnel plots (Fig. 2). Funnel plots show the observed effect sizes on the x-axis plotted against a measure of precision – typically the standard error – on the y-axis. Datapoints with a smaller standard error are predicted to be scattered closer to the average effect size (the vertical line) than datapoints with a larger standard error. In the absence of publication bias and heterogeneity, the majority of the datapoints should fall inside the diagonal lines, which represent the pseudo 95% confidence region with bounds $\hat{\theta} \pm 1.96 \text{ SE}$, where $\hat{\theta}$ is the estimated average effect size and SE is the standard error value from the y-axis (Viechtbauer, 2010). If studies that confirm the hypothesis of a multilingual advantage in pragmatic abilities are more likely to get published, this should be reflected in an asymmetrical distribution of datapoints in the funnel plot; that is, there should be more datapoints on the bottom right side of the distribution than on the bottom left side (e.g., Rothstein et al., 2005). We do not see this distribution in Fig. 2. We do find some horizontal scatter of datapoints, which is a signal of heterogeneity in the data (e.g., Sterne et al., 2011). This indicates that there are factors other than random sampling error alone that are causing variation in the effect sizes.

3.3. Effect of multilingualism on pragmatic abilities

Overall, we did not find evidence that multilingual children outperformed monolingual children on pragmatic abilities, Hedges' $g = 0.09$, 95% CI $[-0.16, 0.33]$. Table 2 shows the meta-analytic results per pragmatic category. We did not find evidence in favor of a multilingual advantage in any of the categories, although the effect size for sensitivity to non-verbal information and understanding non-literal language was in the predicted direction. Note, however, that we would need more studies from more independent research

³ If the age range of children across studies from the same authors were very similar, we contacted the authors to verify whether these studies (partly) reported on the same children; if so, these studies were coded as belonging to the same data collection.

⁴ We calculated the number of unique children by summing the number of children per data collection. If different sample sizes were reported for studies from the same data collection (for example when one study reports on a subset of the children in another study), we included the largest sample size only.

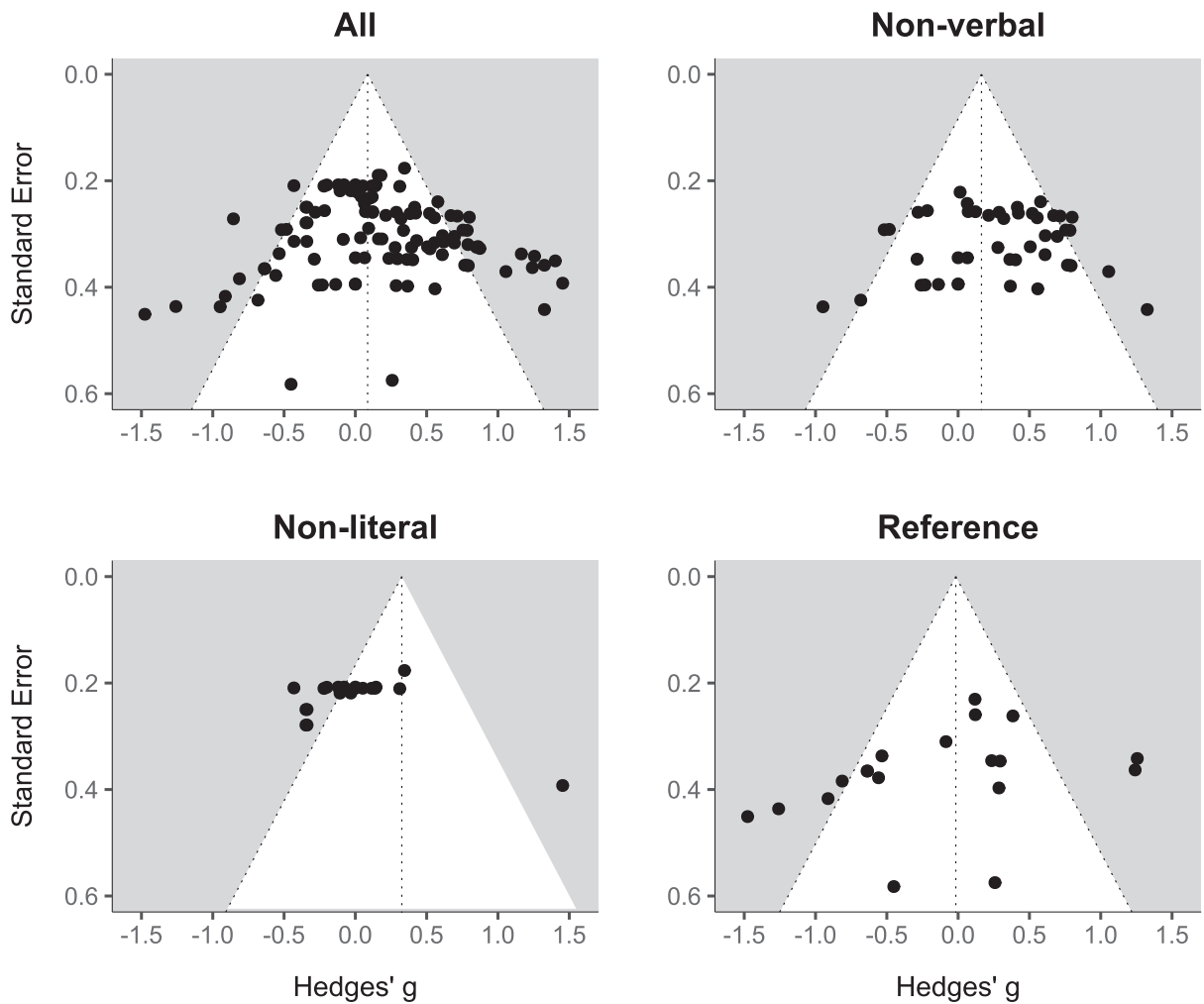


Fig. 2. Funnel plots with observed effect sizes on the x-axis and their corresponding standard errors on the y-axis for all effect sizes (including those from the ‘other’ category) as well as per pragmatic category. We do not show a separate funnel plot for the ‘other’ category, as these studies do not form a coherent set and are therefore not expected to cluster around one mean effect size. The vertical line represents the average effect size, while the diagonal lines represent the boundaries of the pseudo 95% confidence region. Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage.

Table 2

Model results for each pragmatic category, including the estimated variance between research groups and between observations within a research group.

| Model results | | | <i>95% confidence interval</i> | | <i>95% prediction interval^a</i> | |
|--|-----------------------|------------|-----------------------------------|--------------|--|--------------|
| <i>Model</i> | <i>Effect size</i> | <i>df</i> | <i>Lower</i> | <i>Upper</i> | <i>Lower</i> | <i>Upper</i> |
| 1. Sensitivity to non-verbal information | 0.16 | 5.91 | -0.15 | 0.48 | -0.58 | 0.91 |
| 2. Understanding non-literal language | 0.32 | 2.95 | -0.69 | 1.34 | -1.01 | 1.66 |
| 3. Informativeness in referencing | -0.02 | 3.73 | -0.75 | 0.72 | -1.34 | 1.31 |
| Variance components | | | <i>Observation/research group</i> | | | |
| <i>Model</i> | <i>Research group</i> | | <i>n</i> | <i>Var</i> | | |
| 1. Sensitivity to non-verbal information | <i>n</i> | <i>Var</i> | <i>n</i> | <i>Var</i> | | |
| 1. Sensitivity to non-verbal information | 8 | 0.06 | 45 | 0.07 | | |
| 2. Understanding non-literal language | 4 | 0.36 | 20 | 0.01 | | |
| 3. Informativeness in referencing | 5 | 0.19 | 19 | 0.21 | | |

Note. ^aPrediction intervals estimate where the outcome of a new study will fall in 95% of the cases.

groups to obtain reliable estimates for the categories understanding non-literal language and informativeness in referencing, as the degrees of freedom for these categories were below the recommended value of four (Tipton, 2015).

When including data collection as the highest cluster, the results remained largely the same, although the effect size for sensitivity

to non-verbal information was slightly larger with smaller confidence intervals (Hedges' $g = 0.26$, 95% CI = 0.08, 0.45) and the effect size for understanding non-literal language was slightly smaller (Hedges' $g = 0.20$, 95% CI = -0.52, 0.93). Although we did not find evidence for a multilingual advantage for any of the pragmatic categories investigated, the wide confidence intervals for the average effect sizes suggest that there is heterogeneity in effect sizes within each category. This was confirmed by significant tests of heterogeneity (non-verbal: $Q(44) = 104.3$, $p < .001$; non-literal: $Q(19) = 42.3$, $p = .002$; reference: $Q(18) = 67.7$, $p < .001$), indicating that there was variance within each pragmatic category that could not be explained by random sampling error alone.

Given the limited number of studies, we were unable to statistically investigate additional moderators within each category. For this reason, we only visually inspected the distribution of effect sizes to uncover potential patterns in the data by focusing on the direction of the effects and their statistical significance (i.e., whether the confidence intervals did or did not include zero). Fig. 3 presents a forest plot of the effect sizes pertaining to the category *sensitivity to non-verbal information*, where greater dots represent greater precision, the different colors represent different non-verbal cues (i.e., eye gaze, gesture or the interlocutor's visual perspective) and the different shapes represent whether or not there was an experimentally manipulated interfering cue that biased children towards a non-target object (such as object saliency or a verbal label). Two observations can be made. First, we see that multilingual children were typically better at using the interlocutor's visual perspective to constrict possible referents during referential communication tasks. The one study that did not show a positive effect for multilingual children's sensitivity to the interlocutor's visual perspective was the study by Zvaigzne et al. (2008) who investigated whether children adapted their gesture rate and the preciseness of their verbal descriptions based on whether or not the tester could see them. Second, multilingual children seemed to rely more on non-verbal cues during reference resolution than their monolingual peers when the non-verbal cue was more subtle (i.e., eye gaze rather than an indicative gesture) and when this cue was pitted against a conflicting cue that biased children towards a non-target object.

Fig. 4 shows a forest plot for the effect sizes pertaining to the category *understanding non-literal language*, with different colors per subcategory (i.e., indirect speech, novel metaphors, or irony). None of the studies in this category found differences between multilingual and monolingual children, except for Beauchamp et al. (2022) who investigated the understanding of various forms of non-literal language in children with autism spectrum disorder, and Yow and Markman (2011b), in which children had to infer a speaker's emotion based on prosody while the literal content of the sentence provided contradictory information. In these studies, the multilingual children outperformed the monolingual children.

Fig. 5 presents a forest plot of effect sizes pertaining to the category *informativeness in referencing*, with different colors per subcategory (i.e., proportion of ambiguous pronouns, informativeness of children's referential expressions) and different shapes to indicate whether children were typically developing. Studies did not find differences between multilingual and monolingual children in the informativeness of their referential expressions. In contrast, multilingual children sometimes produced *more* (rather than fewer) ambiguous pronouns in their narratives than monolingual children, apart from two notable exceptions where the multilingual children outperformed their monolingual peers (Peristeri et al., 2020; Tsimpli et al., 2017). Interestingly, these effect sizes came from atypically developing children (i.e., children with autism spectrum disorder and developmental language disorder, respectively), whereas the opposite effect was found for the typically developing children in these studies.

Finally, for the sake of completeness, Fig. 6 presents a forest plot of the effect sizes that did not fall within any of the other categories, with different colors per pragmatic ability. Note, however, that we did not compute a mean effect size for this category as these studies do not form a coherent set. In Fig. 6 we see that multilingual children were better at detecting violations of Gricean maxims, but not at producing pragmatically appropriate speech acts or adapting to the communicative needs expressed by the interlocutor (contrary to the original conclusion by Gampe et al., 2019, that was based on separate analyses for the monolingual and multilingual groups rather than a direct comparison between the two groups).

3.4. Role of receptive vocabulary

If multilingual children pay more attention to extralinguistic cues to compensate for gaps in their vocabulary knowledge (according to the Compensation Hypothesis discussed above), then differences in pragmatic ability between multilingual and monolingual children should only emerge when the multilingual group has a smaller vocabulary. However, in the current dataset the difference between the two groups was numerically slightly smaller (rather than larger) when the multilinguals had a smaller receptive vocabulary than the monolinguals (Hedges' $g = 0.09$, $df = 4.98$, 95% CI [-0.25, 0.43], 30 datapoints) compared to when the groups were matched on receptive vocabulary (Hedges' $g = 0.18$, $df = 4.66$, 95% CI [-0.36, 0.71], 17 datapoints). Given the substantial overlap in the confidence intervals for these effect sizes this difference is highly unlikely to be significant.

3.5. Role of age

Fig. 7 presents the distribution of effect sizes as a function of children's mean age in years. Two observations can be made. First, effect sizes did not vary as a function of age, $b = 0.02$, $df = 4.85$, 95% CI [-0.03, 0.06]. Second, we see that different age ranges were investigated for different pragmatic abilities: sensitivity to non-verbal information was typically investigated in children between the ages of 1 and 6, informativeness in referencing was investigated in children between 6 and 10, while understanding non-literal language was typically investigated in older children around the age of 12.

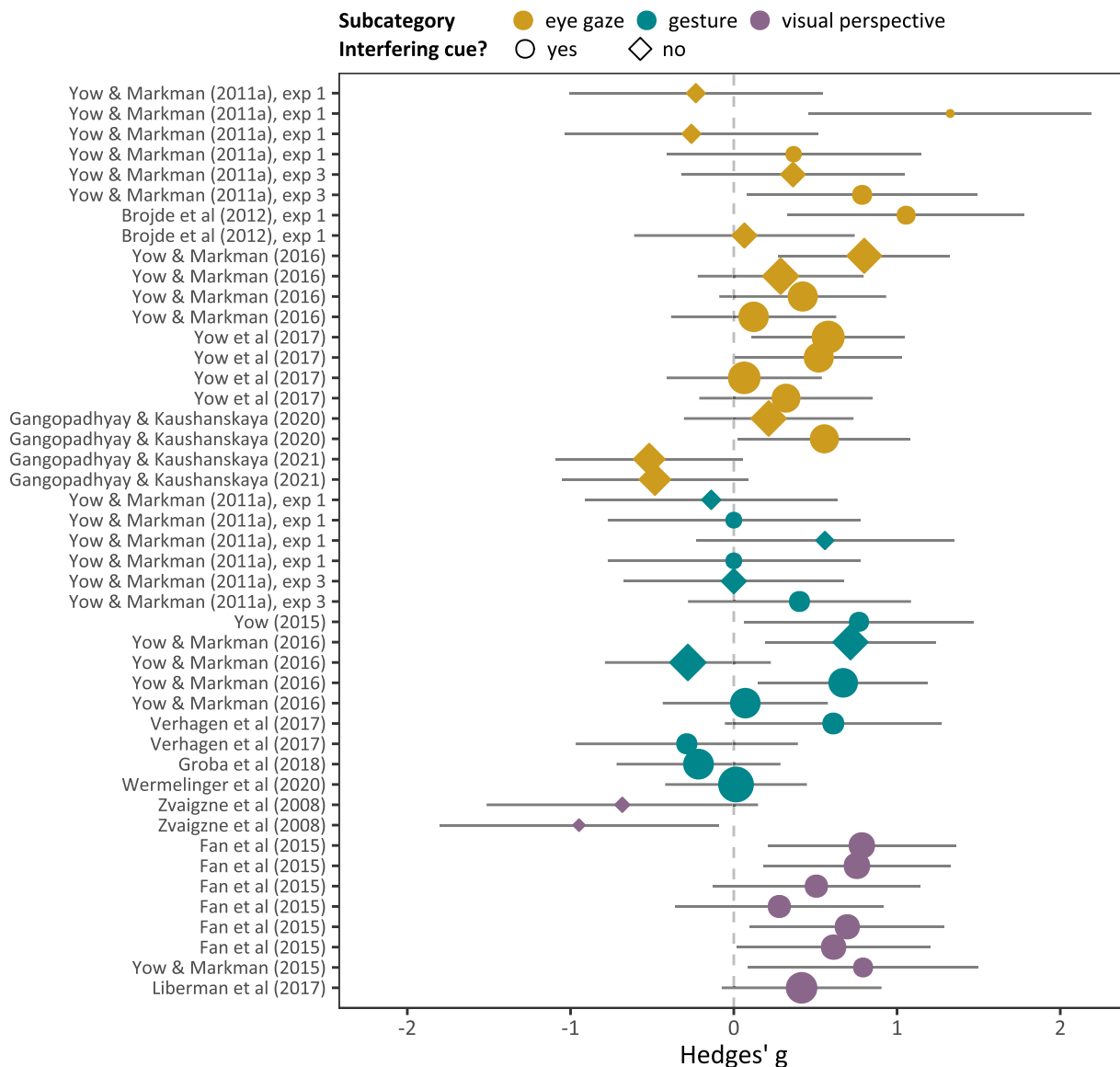


Fig. 3. Effect sizes for the outcomes measuring sensitivity to non-verbal information (14 studies, 45 datapoints) by subcategory and whether or not there was interfering cue. Horizontal lines represent 95% confidence intervals; dot size reflects the inverse variance (i.e., larger dots indicate higher precision). Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage. All effect sizes came from typically developing children. See Fig. S3 at OSF for a plot with additional labels that distinguish between effect sizes from the same study.

4. Discussion

In this study we systematically reviewed previous research on differences in pragmatic abilities between multilingual and monolingual children by means of a meta-analysis. Previously, it has been hypothesized that exposure to multiple languages promotes pragmatic development, either as a consequence of having to monitor and adapt to interlocutors' language abilities (Monitoring Hypothesis; e.g., Brojde et al., 2012; Yow & Markman, 2011b); as a compensatory strategy to overcome gaps in vocabulary or (morpho)syntactic knowledge (Compensation Hypothesis; e.g., Siegal et al., 2009; Verhagen et al., 2017; Wermelinger et al., 2020); or as a result of enhanced ToM or executive functions (Mediation Hypothesis; e.g., Siegal et al., 2010; Yow & Markman, 2011b). However, the evidence for a multilingual advantage in pragmatic abilities has been mixed. Our aim was therefore to quantitatively synthesize all evidence to date to see whether, on average, multilingual children outperform their monolingual peers. A total of 29 studies met our inclusion criteria, resulting in a total of 109 datapoints. These studies investigated a variety of pragmatic abilities, falling into three broad categories: (i) sensitivity to non-verbal information (including eye gaze, gestures and the visual perspective of the interlocutor), (ii) understanding non-literal language (including indirect requests, novel metaphors and irony), and (iii) informativeness in referencing (including level of informativeness and ambiguity of

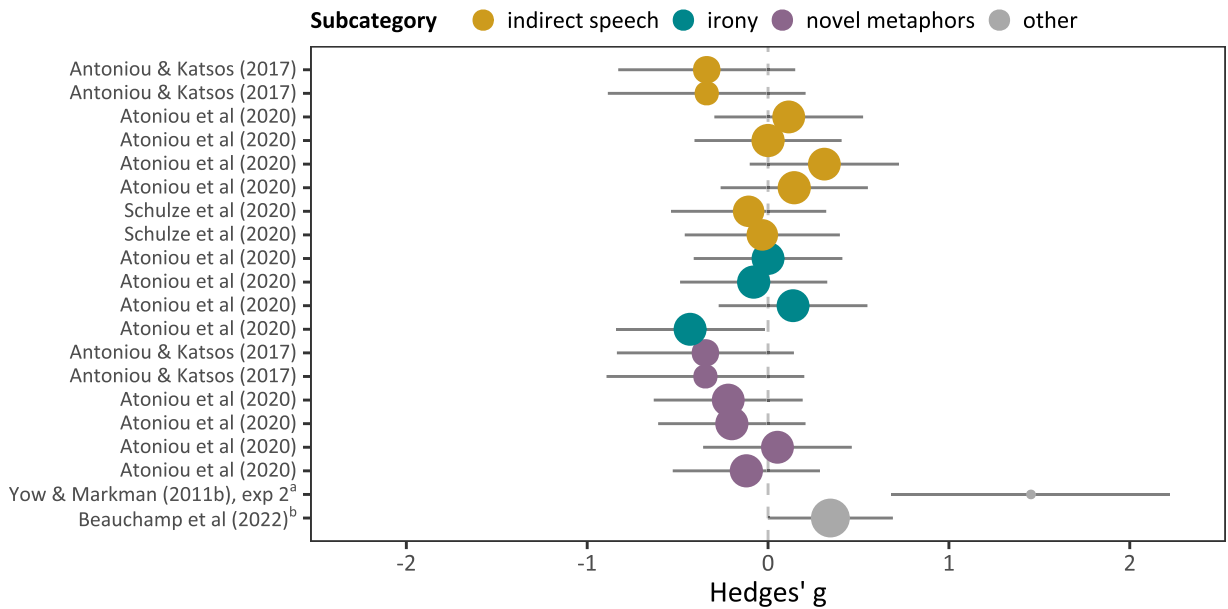


Fig. 4. Effect sizes for the outcomes measuring understanding of non-literal language (5 studies, 20 datapoints) by subcategory. Horizontal lines represent 95% confidence intervals; dot size reflects the inverse variance (i.e., larger dots indicate higher precision). Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage. All effect sizes came from typically developing children except for [Beauchamp et al. \(2022\)](#) who tested children with autism spectrum disorder. See Fig. S4 at OSF for a plot with additional labels that distinguish between effect sizes from the same study. *Note.* ^aInfer speaker’s emotion based on prosody (conflicting cue = literal interpretation). ^bNonliteral Language subtest of the Comprehensive Assessment of Spoken Language (CASL).

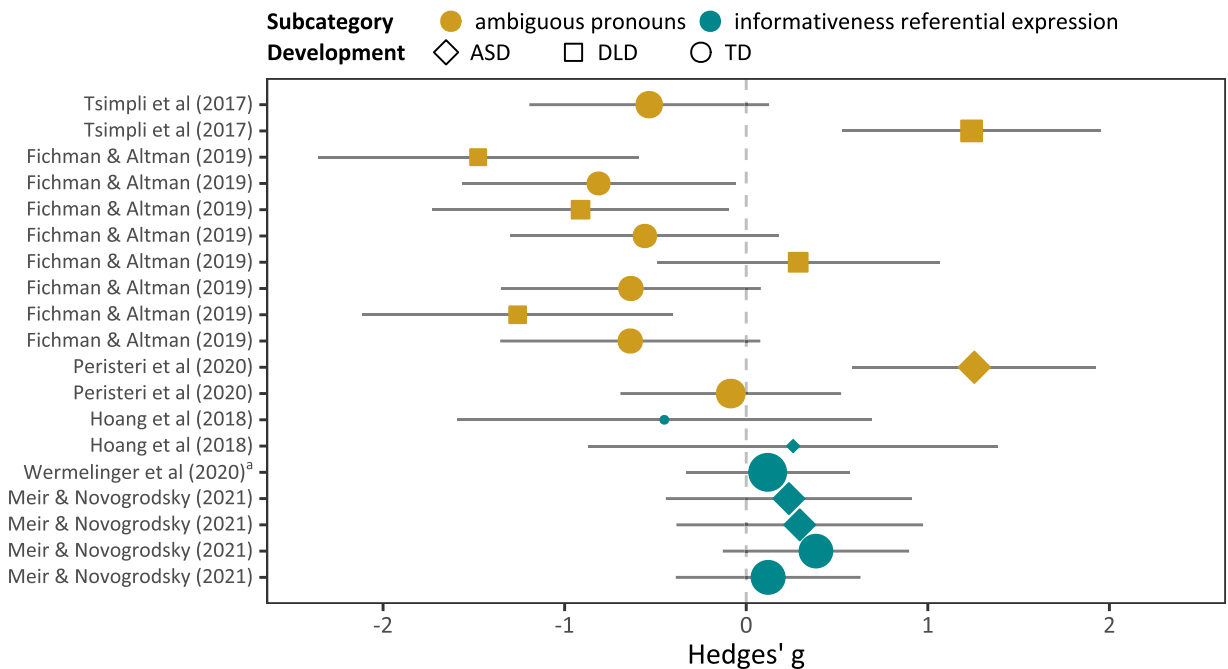


Fig. 5. Effect sizes for the outcomes measuring informativeness in referencing (6 studies, 19 datapoints) by subcategory and children’s development (ASD = autism spectrum disorder, DLD = developmental language disorder, TD = typical development). Horizontal lines represent 95% confidence intervals; dot size reflects the inverse variance (i.e., larger dots indicate higher precision). Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage. See Fig. S5 at OSF for a plot with additional labels that distinguish between effect sizes from the same study. *Note.* ^aThis outcome measured the informativeness of a referential iconic gesture.

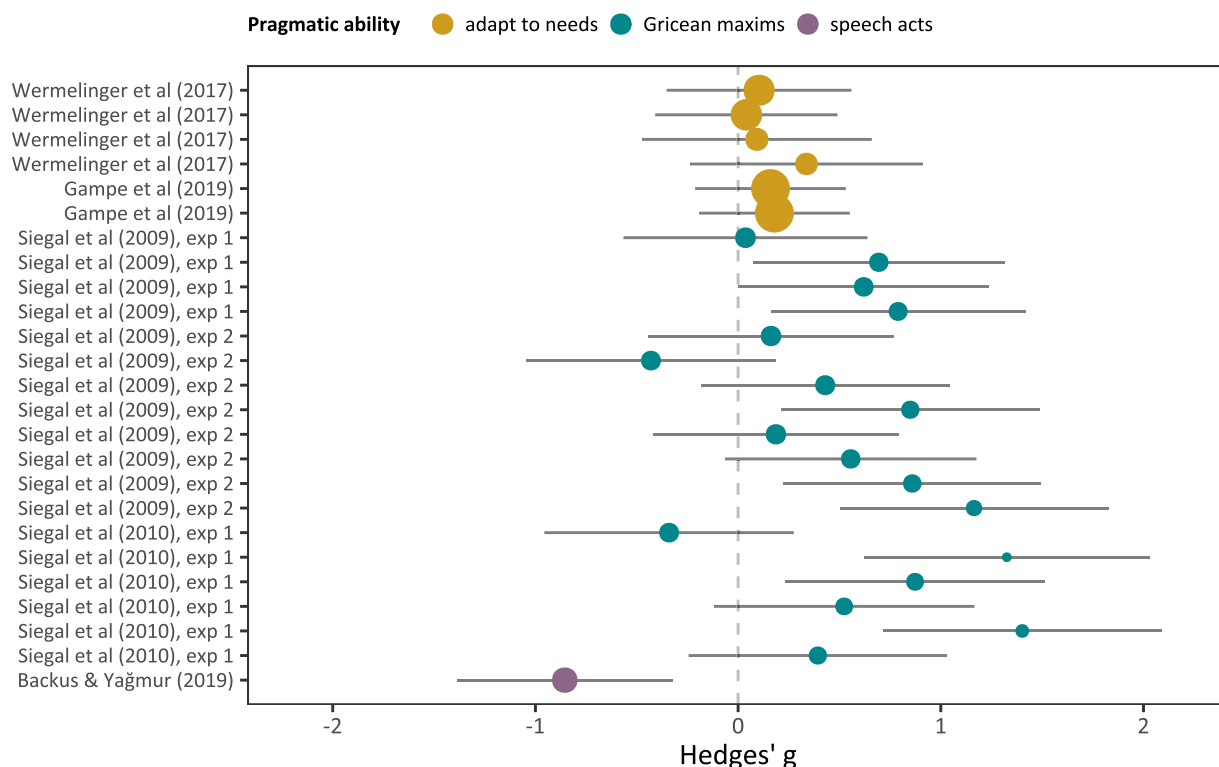


Fig. 6. Effect sizes for the outcomes measuring pragmatic abilities that did not fall within any of the other categories: adaptation to the communicative needs of the interlocutor (2 studies, 6 datapoints), detection of violations of Gricean maxims (2 studies, 18 datapoints), and the pragmatic appropriateness of elicited speech acts (1 study, 1 datapoint). Horizontal lines represent 95% confidence intervals; dot size reflects the inverse variance (i.e., larger dots indicate higher precision). Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage. All effect sizes came from typically developing children. See Fig. S6 at OSF for a plot with additional labels that distinguish between effect sizes from the same study.

referential expressions). The five studies that did not fall within one of these categories tested the detection of violations of Gricean maxims, adaptation to the interlocutor's communicative needs (e.g., repairing communication breakdowns), or the pragmatic appropriateness of children's elicited speech acts. In what follows we first discuss our findings before turning to methodological issues in studying pragmatic abilities, limitations of the current study and recommendations for future research.

4.1. Effect of multilingualism on pragmatic abilities

Generalizing over different pragmatic abilities and research designs, we did not find evidence that there is an overall advantage in pragmatic abilities in multilingual children. In fact, most effect sizes ($n = 75/109$) did not show a difference between the two groups, regardless of pragmatic category. When including 'data collection' instead of 'research group' as the highest cluster (as we had initially preregistered before noting that many studies were conducted by the same research groups), all confidence intervals became smaller and the effect size for sensitivity to non-verbal communication became larger. Based on this analysis, we would have concluded that there is evidence for a multilingual advantage in children's sensitivity to non-verbal information, but not for the other two categories. The smaller confidence intervals with 'data collection' rather than 'research group' as the highest cluster are likely due to this model's incorrect assumption that there are more independent datapoints, which decreases the estimated uncertainty around model estimates. This underscores the importance of accounting for the appropriate highest level of dependency in the data when conducting meta-analyses (Pustejovsky & Tipton, 2022).

Although we did not find statistical evidence for a multilingual advantage for any of the pragmatic categories, at least not when accounting for all data dependencies, significant heterogeneity in effect sizes indicated that there was variation within each category that could not be explained by random sampling error alone. This means there might have been systematic differences in effect sizes associated with between-study differences in research design or participant characteristics. Upon visual inspection of the effect sizes, it seemed that positive effect sizes were mostly obtained for sensitivity to relatively subtle non-verbal cues (i.e., eye gaze, the interlocutor's visual perspective and prosody) that were pitted against conflicting cues such as object saliency, children's own visual perspective or the literal meaning of the sentence. However, this finding remains speculative, as we were not able to test this statistically. It is also important to keep in mind that these effect sizes came from a small number of studies and that they were accompanied by wide confidence intervals. Replication studies are therefore needed to test the robustness and generalizability of these

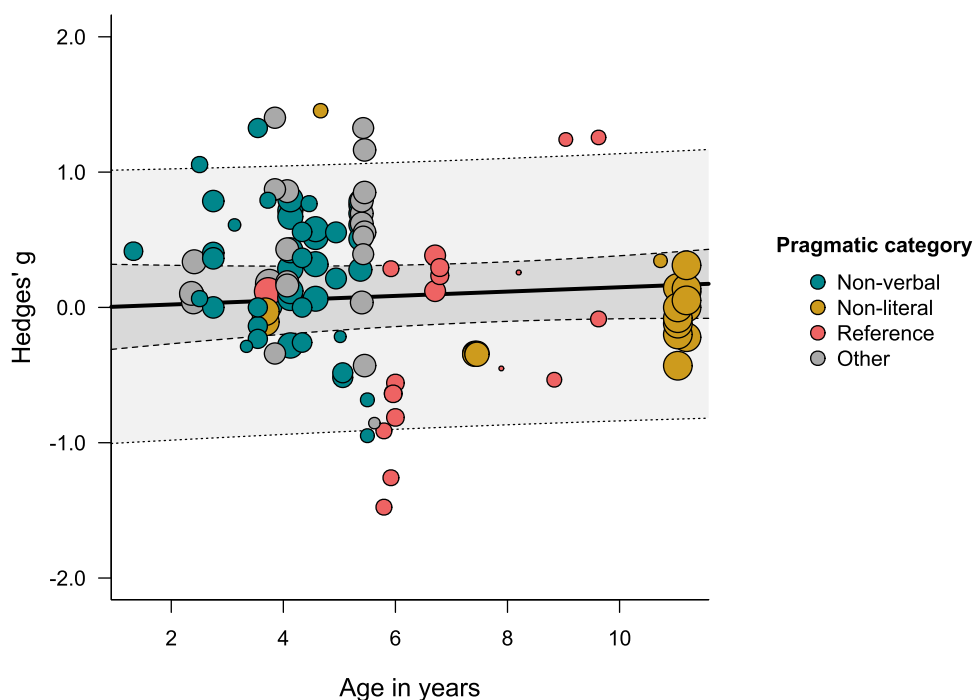


Fig. 7. Effect sizes as a function of children's age (in years) by pragmatic category. The shaded areas represent cluster-robust 95% confidence intervals (dark grey; dashed) and prediction intervals (light grey; dotted). Dot size is proportional to the weight the effect size received in the analysis (with larger dots indicating larger weight). Note that positive effect sizes indicate a multilingual advantage whereas negative effect sizes indicate a monolingual advantage.

findings.

Based on the information available in the included studies, we could conduct exploratory analyses investigating the effect of children's receptive vocabulary knowledge and age. The rationale for testing the effect of receptive vocabulary is that if multilingual children pay more attention to extralinguistic cues to compensate for gaps in their vocabulary knowledge, as predicted by the Compensation Hypothesis, then differences in pragmatic ability between multilingual and monolingual children should only emerge when the former have a relatively smaller vocabulary. We did not find statistical evidence for this hypothesis. Note, however, that this analysis was based on a subset of the datapoints ($n = 47$ out of 109) and was mostly a between-study analysis. This means that the datapoints for which the groups were or were not matched on receptive vocabulary were not necessarily comparable in terms of research design, the pragmatic ability investigated, or participant characteristics other than vocabulary knowledge such as age or socioeconomic status. Regarding age, for which we did not have an a priori hypothesis, we did not find evidence that differences between multilingual and monolingual children varied as a function of children's age. However, we again cannot draw strong conclusions based on this finding as the analysis of age was both cross-sectional and confounded by the fact that different pragmatic abilities were investigated within different age ranges. To investigate whether differences in pragmatic abilities between multilingual and monolingual children become larger or smaller over time, future studies should measure children's pragmatic abilities longitudinally.

In sum, we did not find a multilingual advantage for pragmatic abilities overall. However, we cannot rule out that multilingual children sometimes rely more on non-verbal cues than their monolingual peers. Taken together, these findings indicate that, if multilingualism has an effect on pragmatic abilities, this effect is probably rather small and possibly limited to the domain of non-verbal communication.

4.2. Methodological issues in studying pragmatic competence

Although the pragmatic abilities synthesized in this meta-analysis all met our definition of pragmatics and are often mentioned together in the literature (see, e.g., Gampe et al., 2019; Gangopadhyay & Kaushanskaya, 2021; Yow & Li, 2023), it is not clear whether these abilities form a unitary construct or that we are in fact dealing with multiple inter-related but distinct skills (see also Matthews et al., 2018). Given that the hypotheses in the literature are currently not about specific pragmatic (sub)skills but about pragmatic competence in general, it is only suiting that in this meta-analysis we used a broad definition of pragmatics to investigate whether multilingual children have enhanced pragmatic abilities overall. However, factor analytical work is needed to uncover the different components of pragmatic competence and how they relate to each other. Such research can potentially reveal the extent to which different pragmatic abilities form a coherent construct and can also inform the discussion on how to classify different pragmatic skills.

Apart from the possibility that pragmatic competence is not one single construct, investigating pragmatic competence is also complicated by several confounding factors. For example, virtually any communicative exchange calls on pragmatic reasoning skills as well as grammatical and lexical knowledge (Ariel, 2010; Matthews et al., 2018). Even though we tried our best to exclude phenomena that draw on grammatical or lexical proficiency, there was still a confound with structural language skills in many of the studies. For example, in Backus and Yağmur (2019) the elicited speech acts that were deemed more pragmatically appropriate also tended to be those speech acts that were lexically and syntactically more complex. Likewise, children with larger vocabularies would probably find it easier to understand non-literal language such as novel metaphors, as these children have denser semantic networks to draw on to infer the intended figurative meaning. Given that multilingual children sometimes lag behind their monolingual peers in structural language skills, this potentially masks any pragmatic advantage in tasks that rely heavily on children's language proficiency. In this regard it is interesting to note that the multilingual children in the study by Antoniou et al. (2020) did not perform significantly worse than the monolingual children on understanding non-literal language despite having smaller receptive vocabularies.

Besides language proficiency, tasks measuring pragmatic abilities often draw on ToM and executive function skills. For example, taking the speaker's visual perspective into account during reference resolution requires realizing that their visual perspective is different from one's own and predicting the speaker's behavior based on this knowledge. For this reason, visual perspective taking tasks have sometimes been classified as a measure of ToM (Schroeder, 2018). Likewise, when it comes to producing non-ambiguous referential expressions, many factors come into play other than considering the interlocutor's knowledge state. Whether children produce under- or overspecified referential expressions also depends on their working memory capacity, language experience and proficiency, and possibly the morphological complexity of noun phrases compared to pronouns in the target language (e.g., Fichman et al., 2022; Serratrice & De Cat, 2020; Torregrossa et al., 2021). For example, children may use underspecified pronouns instead of more informative noun phrases as an avoidance strategy when noun phrases require more inflections (Fichman et al., 2022; Fichman & Altman, 2019). Moreover, the context in which children produce referential expressions is crucial: If the child has reason to assume that referents are in the common ground, this eliminates the need to disambiguate referents verbally. This may happen if the child assumes the tester already knows the story or if both the tester and the child can see the pictures that the child needs to describe. Although some of the studies on referencing made sure that the interlocutor was not present when the children first heard the story (Fichman & Altman, 2019; Tsimpli et al., 2017), it was not clear for any of the included studies whether the interlocutor could see the pictures while the children were (re)telling the story.

Due to these confounding factors, it is not always clear whether the found differences between multilingual and monolingual children should be interpreted as differences in pragmatic ability or in another relevant skill. In addition, when no differences were found, this may be because the two groups did not differ in their pragmatic abilities or because the task required another skill for which the difference between the two groups went in the opposite direction (e.g., language proficiency), which may have cancelled out any advantage in pragmatic abilities.

4.3. Limitations of the current study

Apart from the methodological issues in studying pragmatic competence discussed above, an important limitation of the current study was the relatively small number of studies that met our inclusion criteria, which limits the reliability of the meta-analytic results. Moreover, it prevented us from using a more fine-grained classification of pragmatic abilities and from statistically testing additional moderators *within* each category. There are several participant characteristics that may modulate the effect of multilingualism on pragmatic abilities, such as children's development (typical vs. atypical), language distance (Wermelinger et al., 2017), age of onset (i. e., simultaneous vs. sequential multilingualism) and the specific patterns of language use at home and in the wider community (Gullifer & Titone, 2020). In the current meta-analysis, we could not investigate such factors due to the small sample, but also because (i) information on these variables was not always reported in the primary studies, and (ii) these factors often varied between participants within a study. Investigating participant characteristics therefore provides an interesting avenue for future studies.

Second, we used a bottom-up rather than a theoretically motivated approach to classifying pragmatic abilities, because (i) there is currently no consensus on how pragmatic abilities should be classified (Ariel, 2010; Kissine, 2016; O'Neill, 2012), and (ii) existing classifications seem to either not exclusively assign abilities to only one category (e.g., O'Neill, 2012) or make an irrelevant distinction for the purpose of the current study (e.g., Katsos & Andrés-Roqueta, 2021; Kissine, 2016; Recanati, 2004).⁵ We therefore decided to use a classification that, in our view, maximized differences between categories while also yielding sufficient datapoints within each category. Ideally, however, classifications should be both theoretically motivated and supported by empirical data. At the same time, different classifications would not have greatly impacted the results here, given that most studies did not find any differences between multilingual and monolingual children. Nonetheless, other researchers are welcome to investigate different categorizations of our data by using our publicly available dataset and analysis scripts.

Finally, there are a few caveats that apply to meta-analyses more generally. Handbooks of meta-analysis argue that the same effect-size metric can be computed across studies as long as they investigate the same overarching research question (e.g., Lipsey & Wilson,

⁵ Many pragmatists seem to make a distinction between pragmatic inferences that either do or do not require Theory of Mind. For example, Katsos and Andrés-Roqueta make a distinction between linguistic vs. social pragmatics, whereas Kissine speaks of primary vs. secondary meanings, and Recanati of primary vs. secondary pragmatic processes. However, we could not use these categorizations, because we did not include any phenomena that would fall into the first-mentioned categories (such as conventionalized indirect requests), for the reasons explained in the Methods section.

2001). However, given that studies often measure outcome variables on different scales, this still means that for some of the studies the effect-size metric is not optimal for the data. For example, a standardized mean difference as used in the current study is optimal for continuous outcome measures but not for ordinal or binary outcomes. In addition, standardized mean differences are sensitive to non-normality, heteroscedasticity and non-homogeneity (Grissom & Kim, 2001; Sun & Cheung, 2020). Without access to the raw data, the meta-analyst typically does not know whether these assumptions are met; if they are violated, this may introduce bias in the effect size for these studies (Sun & Cheung, 2020). Finally, even small differences in design or sample characteristics result in different population parameters being estimated, posing constraints on the informativeness of an average effect size and effect size comparisons across studies that are not direct replications (Morris & DeShon, 2002).

Despite the limitations mentioned above, we believe that the current meta-analysis is informative for at least three reasons: (i) it provides an overview of the current state-of-the-art research on effects of multilingualism on children's pragmatic abilities and highlights potential avenues for future research, (ii) it gives us insight into the direction of the effects and, concomitantly into whether, in general, studies provide evidence in favor or against the hypothesis that multilingual children outperform monolingual children, and (iii) it can provide a starting point for generating new hypotheses based on the patterns in the distribution of effect sizes as shown in the forest plots.

4.4. Suggestions for future research

4.4.1. Formulating and testing specific hypotheses

In the Introduction, we presented three hypotheses that predict that multilingualism promotes children's pragmatic development (i.e., the Monitoring Hypothesis, the Compensation Hypothesis, and the Mediation Hypothesis). Although these hypotheses have often been mentioned in the literature, they rarely have been tested directly (exceptions being Verhagen et al., 2017 and Yow & Markman, 2016). In the current meta-analysis, we were able to test the Compensation Hypothesis because (i) this hypothesis makes a straightforward prediction (i.e., that multilingual children show a benefit in pragmatic abilities when they know fewer words than their monolingual peers), and (ii) many studies administered a receptive vocabulary test. In contrast, it is not immediately clear how the Monitoring Hypothesis could be tested. Presumably, the language environment could play a role here, as monitoring and adapting to the language preferences of interlocutors may be more frequent or relevant in certain households or communities as compared to others. For example, there may be a greater need to monitor people's language preferences when the predictability of language choices (i.e., "language entropy"; Gullifer & Titone, 2020) is lower. Specifically, one may hypothesize that children pay closer attention to interlocutors' communicative behaviors if they are surrounded by speakers of multiple languages who do not mutually understand each other. In those situations, it is crucial for them to closely monitor people's language abilities in order to avoid communication breakdowns. Finally, according to the Mediation Hypothesis, one would expect that a multilingual advantage is only found when (i) the multilingual children show an advantage in EF or ToM, and (ii) EF or ToM predict children's scores on the outcome of interest. However, thus far, the few studies that included a measure of EF did not find evidence for either one or both of these requirements (Antoniou et al., 2020; Fan et al., 2015; Siegal et al., 2009, 2010; Yow et al., 2017; Yow & Markman, 2015).

Future studies could test and further specify these hypotheses. For uncovering the specific circumstances under which multilinguals may differ from monolinguals we would like to refer to the roadmap provided by Paap and Greenberg (2013) and Paap et al. (2015), and adapt it to the topic of pragmatic abilities: Studies should start with outlining all skills required for a specific pragmatic ability. Then, studies should identify the aspect (or combination of aspects) of multilingual experience that is most important for influencing one or more of these underlying skills. That critical experience should play the lead role in predicting which multilinguals should show differential performance on this pragmatic ability compared to monolinguals or other multilinguals. Moreover, studies could try to pinpoint when such differences are expected to occur. Perhaps one of the reasons why few studies found differences between monolingual and multilingual children thus far is because most studies employed offline rather than online measures. It is possible that multilingual children arrive at the same interpretations as monolingual children, but use different strategies to get there, for example by relying more on non-verbal cues than on verbal content during communication.

In addition, future studies could test the assumptions that underlie the hypotheses on why certain aspects of multilingual experience may affect pragmatic processing. For example, many studies assume that multilingual children experience more communication breakdowns (for example when children do not know certain words or when they select a language the interlocutor does not understand), which may lead them to pay more attention to extralinguistic information to try and avoid such breakdowns. However, we are unaware of any studies that have directly quantified the total number of communication breakdowns multilingual and monolingual children experience in everyday life.⁶ Similarly, the Monitoring Hypothesis assumes that multilingual children (i) routinely monitor their interlocutors' language abilities to avoid communication breakdowns, (ii) form spontaneous speaker-language associations, and (iii) use these associations to adapt their language choices. These assumptions may not be correct, however, and should be tested empirically (see, e.g., Schott et al., 2023).

4.4.2. Disentangling the influence of confounding variables

To disentangle pragmatic abilities from confounding factors such as formal language skills, ToM and executive functions, future studies could try to match children on these confounding variables or administer different tasks to measure the same pragmatic ability.

⁶ In Comeau et al. (2010), the authors did look at the number of communication breakdowns in bilingual compared to monolingual children but for the bilingual children they did not count breakdowns due to language choice (these breakdowns are analyzed separately in Comeau et al., 2007).

Ideally, these tasks would have different confounds (e.g., language proficiency, working memory, ToM) such that the only characteristic they have in common is the assessment of the pragmatic ability of interest. Combining the scores on these different tasks by means of (latent) factor analysis should then, in theory, get rid of the confounding factors that are not shared across tasks, while retaining the variance on the ability of interest. However, we realize this is easier said than done, as it is very difficult to design tasks that require a certain skill while not requiring other, highly related skills. Still, efforts could be made to minimize the influence of confounding factors. For example, to reduce the influence of language proficiency and working memory on understanding non-literal language, one could design a task that uses simple and short sentences rather than relatively long and complex stories.

This recommendation, however, comes with the caveat that it may turn out to be impossible to separate pragmatic abilities from related skills such as language proficiency, ToM and certain executive functions, as it is possible these skills cannot exist in isolation from each other or because they all rely on the same domain-general abilities (Matthews et al., 2018). We therefore repeat the assertion by Matthews et al. (2018) that an important step for the field will be to investigate to which extent these skills are separable from each other, for example by means of factor analysis.

4.4.3. Moving away from the search for a multilingual advantage

As many of the studies included in this meta-analysis set out to investigate a multilingual advantage in pragmatic abilities, this has also been the approach adopted in this study. However, it is not always clear that higher scores on the outcome measures of interest can be interpreted in terms of an “advantage.” For example, as previously mentioned, multilingual children were sometimes found to be more sensitive to eye gaze or gestures than their monolingual peers when this non-verbal cue was pitted against a conflicting cue. In Verhagen et al. (2017), this conflicting cue was the so-called mutual exclusivity constraint, that is, the assumption that novel labels typically do not refer to familiar objects, and vice versa: that familiar labels do not refer to novel objects. Given that both these cues (i.e., gestures and the mutual exclusivity constraint) are relevant for reference resolution, relying more on one of these cues is not automatically “better.” How different cues should be weighted likely depends on a myriad of contextual factors, and an overreliance on a specific type of cue may result in bias rather than a pragmatic advantage. For example, Champoux-Larsson and Dylman (2019) showed that degree of multilingualism predicted children’s reliance on prosody over semantics in inferring a speaker’s emotion, thus conceptually replicating the findings by Yow and Markman (2011b). However, this was the case regardless of whether children were instructed to base their judgments on tone of voice or word content. The authors therefore conclude that rather than an *advantage* in inferring a speaker’s emotion, multilingual children may display a prosodic *bias*.

Moreover, as we have seen in other areas of research on multilingualism, framing research questions in terms of advantages or disadvantages leads to dichotomous thinking and prevents us from investigating more interesting research questions (e.g., Bialystok, 2021). It fosters research where multilingualism is primarily operationalized as a binary variable, masking the multifaceted nature of multilingualism and disregarding potentially relevant variation within both multilingual and monolingual groups (e.g., Luk & Bialystok, 2013; Rothman et al., 2022). This, in turn, may have prevented researchers from formulating and investigating more specific hypotheses; namely, which aspects of the multilingual experience are predicted to affect which aspects of pragmatic ability and which mechanisms underlie such effects. As many other researchers have argued before us, we believe future research should go beyond evaluative group comparisons by (i) not treating multilingualism as a binary variable but quantifying relevant aspects of the multilingual experience (such as age of acquisition, proficiency, exposure and language entropy; e.g., Gullifer & Titone, 2020; Kremin & Byers-Heinlein, 2021; Marian & Hayakawa, 2021), and (ii) focusing on specific hypotheses about which aspects of the multilingual experience are expected to affect certain pragmatic abilities (e.g., Antoniou et al., 2023; Paap et al., 2015).

5. Conclusion

We did not find evidence for an overall pragmatic advantage in multilingual children. However, visual inspection of the effect sizes suggested that multilingual children may be more sensitive to subtle non-verbal or extralinguistic cues (i.e., eye gaze, speaker’s visual perspective, prosody) when these are pitted against conflicting cues (e.g., lexical cues, object saliency, one’s own visual perspective). It remains to be investigated whether these findings can be replicated and whether they result in more successful communication, rather than a possible overreliance or bias towards these cues. We recommend that future studies move away from the search for a multilingual advantage and start formulating and testing more specific hypotheses on which aspects of pragmatic abilities are affected by which aspects of multilingual experience.

Open practices statement

All data, code and Supplementary Materials for this article are available at <https://osf.io/ak9vr/>. The statistical analysis for this article was preregistered after collecting the studies but prior to conducting any analyses at <https://osf.io/vja5x>.

Authors contribution

Elise van Wonderen: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Kimberley Mulder:** Methodology, Supervision, Writing – review & editing. **Judith Rispens:** Supervision, Writing – review & editing. **Josje Verhagen:** Methodology, Supervision, Writing – review & editing, Funding acquisition.

Acknowledgments

We are very grateful to all the authors who answered our questions about their studies and who sent us additional data. We also thank Ileana Grama and Monique Flecken for their comments on a previous version of this manuscript. This work was supported by a grant from the Netherlands Organisation for Scientific Research (NWO; project number VI.Vidi.191.053) to the last author.

References

Studies included in the meta-analysis are marked with an asterisk.

- (*) Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilingualism on implicature understanding. *Applied Psycholinguistics*, 38(4), 787–833. <https://doi.org/10.1017/S014271641600045X>.
- (*) Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2020). How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings. *Bilingualism: Language and Cognition*, 23(1), 186–203. <https://doi.org/10.1017/S1366728918001189>.
- Antoniou, M., Pliatsikas, C., & Schroeder, S. R. (2023). Editorial: New approaches to how bilingualism shapes cognition and the brain across the lifespan: Beyond the false dichotomy of advantage versus no advantage. *Frontiers in Psychology*, 14. (<https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1149062>).
- Ariel, M. (2010). *Defining pragmatics*. Cambridge University Press.
- (*) Backus, A., & Yağmur, K. (2019). Differences in pragmatic skills between bilingual Turkish immigrant children in the Netherlands and monolingual peers. *International Journal of Bilingualism*, 23(4), 817–830. <https://doi.org/10.1177/1367006917703455>.
- (*) Beauchamp, M. L. H., Rezzonico, S., Bennett, T., Duku, E., Georgiades, S., Kerns, C., Miranda, P., Richard, A., Smith, I. M., Szatmari, P., Vaillancourt, T., Waddell, C., Zaidman-Zait, A., Zwaigenbaum, L., & Elsabbagh, M. (2022). The influence of bilingual language exposure on the narrative, social and pragmatic abilities of school-aged children on the autism spectrum. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-022-05678-w>.
- Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue Canadienne Délévôt Psychologie Expérimentale*, 65(4), 229–235. <https://doi.org/10.1037/a0025406>
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262. <https://doi.org/10.1037/bul0000999>
- Bialystok, E. (2021). Bilingualism as a Slice of Swiss Cheese. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.769323>
- Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240–250. <https://doi.org/10.1016/j.tics.2012.03.001>
- Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- (*) Brojde, C. L., Ahmed, S., & Colunga, E. (2012). Bilingual and monolingual children attend to different cues when learning new words. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00155>.
- Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Blackwell.
- Champoux-Larsson, M.-F., & Dylman, A. S. (2019). A prosodic bias, not an advantage, in bilinguals' interpretation of emotional prosody. *Bilingualism: Language and Cognition*, 22(2), 416–424. <https://doi.org/10.1017/S1366728918000640>
- Comeau, L., Genesee, F., & Lapaquette, L. (2003). The Modeling Hypothesis and child bilingual codemixing. *International Journal of Bilingualism*, 7(2), 113–126. <https://doi.org/10.1177/13670069030070020101>
- Comeau, L., Genesee, F., & Mendelson, M. (2007). Bilingual children's repairs of breakdowns in communication. *Journal of Child Language*, 34(1), 159–174. <https://doi.org/10.1017/S0305000906007690>
- Comeau, L., Genesee, F., & Mendelson, M. (2010). A comparison of bilingual and monolingual children's conversational repairs. *First Language*, 30(3–4), 354–374. <https://doi.org/10.1177/0142723710370530>
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- (*) Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, 26(7), 1090–1097. <https://doi.org/10.1177/0956797615574699>.
- Farhadian, M., Abdullah, R., Mansor, M., Redzuan, M., Gazanizadand, N., & Kumar, V. (2010). Theory of mind in bilingual and monolingual preschool children. *Journal of Psychology*, 1(1), 39–46. <https://doi.org/10.1080/09764224.2010.11885444>
- (*) Fichman, S., & Altman, C. (2019). Referential cohesion in the narratives of bilingual and monolingual children with typically developing language and with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 62(1), 123–142. https://doi.org/10.1044/2018_JSLHR-L18-0054.
- (*) Fichman, S., Walters, J., Melamed, R., & Altman, C. (2022). Reference to characters in narratives of Russian-Hebrew bilingual and Russian and Hebrew monolingual children with Developmental Language Disorder and typical language development. *First Language*, 42(2), 263–291. <https://doi.org/10.1177/0142723720962938>.
- (*) Gampe, A., Wermelinger, S., & Daum, M. M. (2019). Bilingual children adapt to the needs of their communication partners, monolinguals do not. *Child Development*, 90(1), 98–107. <https://doi.org/10.1111/cdev.13190>.
- (*) Gangopadhyay, I., & Kaushanskaya, M. (2020). The role of speaker eye gaze and mutual exclusivity in novel word learning by monolingual and bilingual children. *Journal of Experimental Child Psychology*, 197, Article 104878. <https://doi.org/10.1016/j.jecp.2020.104878>.
- (*) Gangopadhyay, I., & Kaushanskaya, M. (2021). Word learning in monolingual and bilingual children: The influence of speaker eye-gaze. *Bilingualism: Language and Cognition*, 24(2), 333–343. <https://doi.org/10.1017/S1366728920000565>.
- Genesee, F. (2006). Bilingual first language acquisition in perspective. In P. D. McCardle, & E. Hoff (Eds.), *Childhood bilingualism: Research on infancy through school age* (pp. 45–67). Multilingual Matters.
- Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied Psycholinguistics*, 17(4), 427–442. <https://doi.org/10.1017/S0142716400008183>
- Genesee, F., Tucker, G. R., & Lambert, W. E. (1975). Communication skills of bilingual children. *Child Development*, 46(4), 1010–1014. <https://doi.org/10.2307/1128415>
- Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6(2), 135–146.
- (*) Groba, A., De Houwer, A., Mehnert, J., Rossi, S., & Obrig, H. (2018). Bilingual and monolingual children process pragmatic cues differently when learning novel adjectives. *Bilingualism: Language and Cognition*, 21(2), 384–402. <https://doi.org/10.1017/S1366728917000232>.
- Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, 23(2), 283–294. <https://doi.org/10.1017/S1366728919000026>
- Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, 146(12), 1059–1083. <https://doi.org/10.1037/bul0000301>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hoff, E. (2021). Why bilingual development is not easy. In J. J. Lockman (Ed.), *Advances in Child Development and Behavior*, 61 pp. 129–167. JAI. <https://doi.org/10.1016/bs.acdb.2021.03.002>.
- Katsos, N., & Andrés-Roqueta, C. (2021). Where next for pragmatics and mind reading? A situation-based view (Response to Kissine). *Language*, 97(3), e184–e197. <https://doi.org/10.1353/lan.2021.0036>
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38. <https://doi.org/10.1111/1467-9280.00211>

- Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13(3), 207–212. <https://doi.org/10.1111/1467-9280.00439>
- Kissine, M. (2016). Pragmatics as metacognitive control. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.02057>
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12(1), 48–54. <https://doi.org/10.1111/j.1467-7687.2008.00742.x>
- Kremin, L. V., & Byers-Heinlein, K. (2021). Why not both? Rethinking categorical and continuous approaches to bilingualism. *International Journal of Bilingualism*, 25(6), 1560–1575. <https://doi.org/10.1177/13670069211031986>
- (*) Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*, 20(1), Article e12420. <https://doi.org/10.1111/desc.12420>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.
- Lowe, C. J., Cho, I., Goldsmith, S. F., & Morton, J. B. (2021). The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological Science*, 32(7), 1115–1146. <https://doi.org/10.1177/0956797621993108>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a "bilingualism quotient". *Applied Psycholinguistics*, 42(2), 527–548. <https://doi.org/10.1017/S0142716420000533>
- Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, 14(3), 186–223. <https://doi.org/10.1080/15475441.2018.1455584>
- McTear, M., & Conti-Ramsden, G. (1992). *Pragmatic disability in children*. Whurr.
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105. <https://doi.org/10.1037/1082-989X.7.1.105>
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191–230. <https://doi.org/10.1017/S0142716407070117>
- O'Neill, D. K. (2012). 10. Components of pragmatic ability and children's pragmatic language development. In H.-J. Schmid (Ed.), *Cognitive Pragmatics* (pp. 261–288). DE GRUYTER. <https://doi.org/10.1515/9783110214215.261>
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258. <https://doi.org/10.1016/j.cogpsych.2012.12.002>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <https://doi.org/10.1016/j.cortex.2015.04.014>
- (*) Peristeri, E., Baldimtsi, E., Andreou, M., & Tsimpli, I. M. (2020). The impact of bilingualism on the narrative ability and the executive functions of children with autism spectrum disorders. *Journal of Communication Disorders*, 85, Article 105999. <https://doi.org/10.1016/j.jcomdis.2020.105999>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Pustejovsky, J. E. (2022). *ClubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections*. R package version 0.5.8. [Computer software]. (<https://CRAN.R-project.org/package=clubSandwich>).
- Pustejovsky, J. E. (2015, September 17). *Correlations between standardized mean differences*. (<https://www.jepusto.com/correlations-between-smds/>).
- Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. *Prevention Science*, 23(3), 425–438. <https://doi.org/10.1007/s11221-021-01246-3>
- Recanatì, F. (2004). *Literar meaning*. Cambridge University Press.
- Rohatgi, A. (2021). *WebPlotDigitizer* (Version 4.5) [Computer software]. (<https://automeris.io/WebPlotDigitizer>).
- Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Duñabettia, J. A., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., van Osch, B., Pereira Soares, S. M., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2022). Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *Applied Psycholinguistics*, 1–14. <https://doi.org/10.1017/S0142716422000315>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustment*. Wiley.
- Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, 47(1), 269–273. <https://doi.org/10.1016/j.jesp.2010.09.005>
- Schott, E., Tamayo, M. P., & Byers-Heinlein, K. (2023). Keeping track of language: Can monolingual and bilingual infants associate a speaker with the language they speak. *Infant and Child Development*, Article e2403. <https://doi.org/10.1002/icd.2403>
- Schroeder, S. R. (2018). Do bilinguals have an advantage in Theory of Mind? A meta-analysis. *Frontiers in Communication*, 3, 36. <https://doi.org/10.3389/fcomm.2018.00036>
- (*) Schulze, C., Endesfelder Quick, A., Gampe, A., & Daum, M. M. (2020). Understanding verbal indirect communication in monolingual and bilingual children. *Cognitive Development*, 55, Article 100912. <https://doi.org/10.1016/j.cogdev.2020.100912>
- Serratrice, L., & De Cat, C. (2020). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingualism: Language and Cognition*, 23(2), 371–386. <https://doi.org/10.1017/S1366728918000962>
- (*) Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, 110(1), 115–122. <https://doi.org/10.1016/j.cognition.2008.11.002>
- (*) Siegal, M., Surian, L., Matsuo, A., Geraci, A., Iozzi, L., Okumura, Y., & Itakura, S. (2010). Bilingualism accentuates children's conversational understanding. *PLoS One*, 5(2), Article e9004. <https://doi.org/10.1371/journal.pone.0009004>
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343, d4002. <https://doi.org/10.1136/bmj.d4002>
- Sun, R. W., & Cheung, S. F. (2020). The influence of nonnormality from primary studies on the standardized mean difference in meta-analysis. *Behavior Research Methods*, 52(4), 1552–1567. <https://doi.org/10.3758/s13428-019-01334-x>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426–445. <https://doi.org/10.1177/1367006911403202>
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393. <https://doi.org/10.1037/met0000011>
- Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. M. (2021). Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, 24(4), 694–706. <https://doi.org/10.1017/S1366728920000826>
- (*) Tsimpli, I. M., Peristeri, E., & Andreou, M. (2017). Object Clitic production in monolingual and bilingual children with Specific Language Impairment: A comparison between elicited production and narratives. *Linguistic Approaches to Bilingualism*, 7(3–4), 394–430. <https://doi.org/10.1075/lab.15025.tsi>
- Turkstra, L. S., Clark, A., Burgess, S., Hengst, J. A., Wertheimer, J. C., & Paul, D. (2017). Pragmatic communication abilities in children and adults: Implications for rehabilitation professionals. *Disability and Rehabilitation*, 39(18), 1872–1885. <https://doi.org/10.1080/09638288.2016.1212113>
- (*) Verhagen, J., Grassmann, S., & Küntay, A. C. (2017). Monolingual and bilingual children's resolution of referential conflicts: Effects of bilingualism and relative language proficiency. *Cognitive Development*, 41, 10–18. <https://doi.org/10.1016/j.cogdev.2016.10.003>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3). <https://doi.org/10.18637/jss.v036.i03>
- (*) Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, 155, 84–94. <https://doi.org/10.1016/j.jecp.2016.11.005>
- (*) Wermelinger, S., Gampe, A., Helbling, N., & Daum, M. M. (2020). Do you understand what I want to tell you? Early sensitivity in bilinguals' iconic gesture perception and production. *Developmental Science*, 23(5), Article e12943. <https://doi.org/10.1111/desc.12943>

- Wilson Van Voorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43–50. <https://doi.org/10.20982/tqmp.03.2.p043>
- (*) Yow, W. Q. (2015). Monolingual and bilingual preschoolers' use of gestures to interpret ambiguous pronouns. *Journal of Child Language*, 42(6), 1394–1407. <https://doi.org/10.1017/S0305000914000737>.
- Yow, W. Q., & Li, X. (2023). Role of bilingual experience in children's context-sensitive selective trust strategies. *Bilingualism: Language and Cognition*, 1–12. <https://doi.org/10.1017/S1366728923000433>
- (*) Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. P. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental Science*, 20(1), Article e12482. <https://doi.org/10.1111/desc.12482>.
- (*) Yow, W. Q., & Markman, E. M. (2011a). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, 14(4), 562–569. <https://doi.org/10.1017/S1366728910000404>.
- (*) Yow, W. Q., & Markman, E. M. (2011b). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, 12(1), 12–31. <https://doi.org/10.1080/15248372.2011.539524>.
- (*) Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent. *Bilingualism: Language and Cognition*, 18(3), 391–399. <https://doi.org/10.1017/S1366728914000133>.
- (*) Yow, W. Q., & Markman, E. M. (2016). Children increase their sensitivity to a speaker's nonlinguistic cues following a communicative breakdown. *Child Development*, 87(2), 385–394. <https://doi.org/10.1111/cdev.12479>.
- Zufferey, S. (2020). Pragmatic development in a first language: An overview. In K. P. Schneider, & E. Ifantidou (Eds.), *Developmental and Clinical Pragmatics* (pp. 33–60). De Gruyter. <https://doi.org/10.1515/9783110431056-002>.
- (*) Zvaigzne, M., Oshima-Takane, Y., Groleau, P., Nakamura, K., & Genesee, F. (2008). The function of children's iconic co-speech gestures: A study with Japanese-French bilinguals and French monolinguals. *Proceedings of 32nd Annual Boston University Conference on Language Development*, 598–609.
- Zvaigzne, M., Oshima-Takane, Y., & Hirakawa, M. (2019). How does language proficiency affect children's iconic gesture use? *Applied Psycholinguistics*, 40(2), 555–583. <https://doi.org/10.1017/S014271641800070X>