Developmental Review 64 (2022) 101023

Contents lists available at ScienceDirect

Developmental Review

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

Review

The relationship between infant pointing and language development: A meta-analytic review

Elizabeth Kirk [a,*], Seamus Donnelly [b,g,h], Reyhan Furman [c], Meesha Warmington [d], Julie Glanville [e], Adam Eggleston [f]

[a] School of Psychology and Sport Science, Anglia Ruskin University, East Road, Cambridge CB4 1PT, UK
[b] Research School of Psychology, Australian National University, Canberra, Australia
[c] School of Psychology and Computer Science, University of Central Lancashire, Preston, Lancashire PR1 2HE, UK
[d] School of Education, University of Sheffield, Sheffield S10 2GW, UK
[e] York Health Economics Consortium, University of York, York YO10 5NQ, UK
[f] Department of Psychology, University of York, York YO10 5DD, UK
[g] ARC Centre of Excellence for Dynamics of Language, Canberra, Australia
[h] Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

A B S T R A C T

Infant pointing has long been identified as an important precursor and predictor of language development. Infants typically begin to produce index finger pointing around the time of their first birthday and previous research has shown that both the onset and the frequency of pointing can predict aspects of productive and receptive language. The current study used a multivariate meta-analytic approach to estimate the strength of the relationship between infant pointing and language. We identified 30 papers published between 1984 and 2019 that met our stringent inclusion criteria, and 25 studies (comprising 77 effect sizes) with samples ≥10 were analysed. Methodological quality of the studies was assessed to identify potential sources of bias. We found a significant but small overall effect size of $r = 0.20$. Our findings indicate that the unique contribution of pointing to language development may be less robust than has been previously understood, however our stringent inclusion criteria (as well as our publication bias corrections), means that our data represent a more conservative estimate of the relationship between pointing and language. Moderator analysis showed significant group differences in favour of effect sizes related to language comprehension, non-vocabulary measures of language, pointing assessed after 18 months of age and pointing measured independent of speech. A significant strength of this study is the use of multivariate meta-analysis, which allowed us to utilise all available data to provide a more accurate estimate. We consider the findings in the context of the existing research and discuss the general limitations in this field, including the lack of cultural diversity.

Introduction

Before infants learn to speak, they can use their hands to gesture, signalling to their caregivers both their wants and their interests (e.g. Tomasello, Carpenter & Liszkowski, 2007). Typically, infants begin to produce the index finger point around the time of their first birthday (e.g. Butterworth & Morissette, 1996), but as with many aspects of child development, there is great individual variability with some studies reporting the onset up to 15 months (Camaioni, Perucchini, Bellagamba, Colonnesi, 2004; Desrochers, Morisette & Ricard, 1995). Between the ages of one and two, the rate of pointing will increase and will take on an enhanced role as infants start to complement, and then later, supplement their one-word utterances with a point. Thus, by saying "mine" and pointing at a toy, the child

* Corresponding author at: Anglia Ruskin University, School of Psychology and Sports Sciences, East Road, Cambridge CB4 1PT, UK.
E-mail addresses: elizabeth.kirk@anglia.ac.uk, elizabeth.kirk@aru.ac.uk, lizabeth.kirk@aru.ac.uk (E. Kirk), seamus.donnelly@mpi.nl (S. Donnelly), rfurman@uclan.ac.uk (R. Furman), m.warmington@sheffield.ac.uk (M. Warmington), julie@glanville.info (J. Glanville), adam.eggleston@york.ac.uk (A. Eggleston).

https://doi.org/10.1016/j.dr.2022.101023
Received 1 October 2020; Received in revised form 5 March 2022
Available online 18 March 2022

can convey two units of information with just one word combined with a gesture (Özçalışkan & Goldin-Meadow, 2005). This is an important precursor to two-word speech, which typically emerges between 18 and 24 months of age. Studies have demonstrated that the production of these gesture-speech combinations predicts the onset of two-word speech (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Iverson et al. 2008).

There is clear support from the wealth of studies that has been published over the last few decades that infant language and gesture are "close family" (Bates & Dick, 2002). Studies report evidence that infant pointing can predict the size of their subsequent spoken vocabularies (Rowe & Goldin-Meadow, 2009), and also the individual words they will acquire (Iverson & Goldin-Meadow, 2005).

Thus, the question which arises is what mechanism underpins this relationship? The predominant view is that pointing elicits labelling responses from caregivers and that it is via this pathway that pointing is associated with language proficiency. Caregivers are more likely to respond to infants' points with labels than they are to reaching gestures or object extensions (i.e., extension of the arm while holding an object) (Kishimoto et al., 2007; Masur, 1982; Olson & Masur, 2015). There is evidence demonstrating that labelling responses to infants' points helps explain the relationship between pointing and word learning, in both typically developing and developmentally disordered children (Dimitrova, Özçalişkan, & Adamson, 2016). Goldin-Meadow, Goodrich, Sauer, and Iverson (2007) tested the association between gesture and vocabulary at the level of individual words. Mothers were observed to translate their child's points into words, which in turn resulted in those words subsequently entering the child's vocabulary. Interestingly, Olson and Masur (2015) found that mothers' object labelling responses to infants' gestures fully mediated both the concurrent and longitudinal relationship between infant gesture and expressive noun lexicons. Thus, there is support for the view that the relation between pointing and language is socially mediated, such that pointing elicits contingent talk from caregivers (i.e. the caregiver talks about the referent of the infant's point) thus increasing the child's exposure to word-object relations (for further theoretical discussion on the motivation of pointing we refer the reader to Tomasello, Carpenter, & Liszkowski, 2007; Liszkowski, & Ruether, 2021). In 2010, Colonnesi, Stams, Koster and Noom published a *meta*-analysis of 25 studies that had examined the association between pointing and language published between 1978 and 2009. Of these, 12 had considered the concurrent association between pointing and language, and this was found to be a strong effect ($r = 0.52$). The longitudinal association was assessed by 18 studies and this was reported to be a medium effect ($r = 0.35$). The strength of the association was found to be moderated by age, with the largest effect sizes observed when pointing was measured when infants were 15 months and older. The majority of the studies included US and UK samples ($n = 16$), with the remainder being European ($n = 8$). Only one study was drawn from a non-western sample. Interestingly, in this study of Japanese infants, no association was found between infant pointing and language (Blake et al. 2003). This raises the question of whether the association between pointing and language transcends cultures and languages. If the association between pointing and language is socially mediated, then we might anticipate that not all cultures respond in the same way to infants' points and so the relationship may be culturally limited. Indeed, we know from cross-cultural research that there is variation in the type and amount of communicative interaction between parents and infants (e.g. Abels, 2020; Keller, 2007; Keller et al., 2008). Furthermore, none of the studies included samples of bilingual infants. An interesting question is whether bilingual infants demonstrate robust associations between pointing and language across both their languages, or whether this varies depending on the typology of language that they are learning (i.e. verb framed or satellite framed as defined by Talmy (1985)).

The aim of this meta-analysis was to test the hypothesis that infant pointing is a significant predictor of language and to assess the strength of this association. We were motivated to extend the work of Colonnesi and colleagues (2010), with a focus on comparing evidence from studies conducted across a range of languages and cultures and to test whether a series of moderators impacted upon the strength of the relationship between pointing and language. Furthermore, we aimed to establish whether any work had been conducted to examine the association between pointing and language in bilingual infants. Specifically, our review questions were: (1) what is the strength of the relationship between pointing and language development in monolingual and bilingual infants? and (2) is this association universal across languages or is it language specific?

Typically, studies of infant gesture and language report several different associations between various measures of language outcome and longitudinal analyses of the associations between pointing and language. The conventional approach is to aggregate these effect sizes to provide one average per paper (e.g. Colonnesi et al. 2010). However, this approach can lead to missed opportunities to utilise all the available data (Cheung, 2019), thus we used three-level meta-analysis to allow multiple effect sizes per individual paper (Cheung, 2019), whilst handling the non-independence of sampling error between effect sizes with cluster robust variance estimation (Hedges, Tipton & Johnson, 2010).

Several moderator variables were identified as factors that could impact upon the strength of the association between pointing and language. This relationship has been reported under a range of different conditions. For example, the pointing variable may refer to the age of onset of pointing, the frequency of pointing, the age of onset of gesture-speech combinations. Similarly, language is measured in different ways and these associations are tested both concurrently and longitudinally. There are various environmental factors that might impede on this association. Thus, the use of a meta-analysis allows for a robust assessment of whether the pointing-language relationship is reliable when these moderators are considered.

Bilingualism

We aimed to consider the strength of the pointing-language association separately for monolingual and bilingual samples. Bilingual infants typically demonstrate unequal development in their two languages, with one language having more dominance due to greater exposure and other factors (Nicoladis et al., 1999). Thus, these children provide the ideal test sample to examine the pointing-language relationship within the child but across two different language contexts, to elucidate whether gesture more generally signals an underlying cognitive capacity for language, or gesture specifically predicts language development within a particular language (perhaps via the social mechanism of inviting labelling).

Socio-economic status

Maternal SES has been identified to be an important factor for predicting the frequency with which mothers point when interacting with their infants, and in turn the frequency with which infants themselves point, which is predictive of their subsequent vocabularies (Rowe & Goldin-Meadow, 2009). Therefore, variation in pointing and therefore the strength of the relationship between pointing and language is anticipated to be related to SES.

Language modality

Language modality was included as a moderator to assess whether there were stronger associations between pointing and receptive or expressive measures of language. We anticipated that the association between pointing and language would be stronger for measures of receptive language rather than productive (as per Colonessi et al. 2010), because of the assumed pathway with caregivers providing verbal labels to infants' points thus scaffolding their comprehension and subsequently their production.

Language measure

The studies in this field employ a variety of different measures to assess infant language, including parental report measures of infant's vocabulary (the MacArthur Bates Communicative Development Inventory: CDI, Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994), standardised measures of receptive language (e.g. the Peabody Picture Vocabulary Test, Dunn, Dunn, Bulheller, & Häcker, 1965) and standardized measures of children's comprehension and production of language (the Reynell Developmental Language scales, Edwards, Letts & Sinka, 2011; the Mullen Scales, Mullen, 1995; the Pre-School Language Scale, Zimmerman, Steiner & Pond, 2002). Other studies code language measures from videotaped observations of infants, including Mean Length of Utterance and the onset of two-word utterances. Given the range of available measures we anticipated significant heterogeneity which we attempted to control for using moderator analysis.

Language and country

It has been demonstrated that there are cultural differences in triadic joint actions which directly impacts on the frequency of parent, and therefore infant pointing (Salomo & Liszkowski, 2013). It could be that in cultures where the social-interactional input is less focussed on joint action, the link between infant pointing and language development is weaker. It is not yet understood whether the pointing-language association is universal, thus we sought to examine this using meta-analysis.

Mean age of pointing assessment

To consider the impact of the age of pointing assessment, we distinguish between studies that assessed pointing before or after 18 months of age. Around 18 months infants typically attain the milestone of producing 50 words, with some children beginning two-word utterances (Tamis-LeMonda, Bornstein & Baumwell, 2001), thus language development hits its stride after 18 months and therefore the role of pointing is likely to change.

Pointing measure

Within the literature there are studies that examine the relationship between the age of onset of pointing and language outcomes (e.g. Goldin-Meadow & Butcher, 2003), and pointing frequency (measured at various ages) and language (e.g. Igualada et al. 2018). These potentially represent different pathways to language acquisition, with age of onset indexing the start of shared intentionality using gesture, whereas frequency of pointing is more likely to be related to the child's interests in his or her environment and likely a marker for the level of input that the child receives from caregivers. As such, we were interested in whether the strength of the association between pointing and language would change depending on the way pointing was operationalised.

Pointing combination

Children combine pointing with single word-utterances during the transition from telegraphic speech to two-word utterances. Some research focuses specifically on these gesture-speech combinations and how they predict oncoming language (e.g. Özçalışkan and Goldin-Meadow, 2009). Therefore, we considered this as a moderator to test whether the relationships were stronger in these studies compared to studies that assessed pointing independently of its relationship to speech. We anticipated that because these studies typically focus on very close mappings between gesture-speech combinations and the onset of two word speech, that the strength of the association would be stronger compared to studies that consider gesture independently of speech.

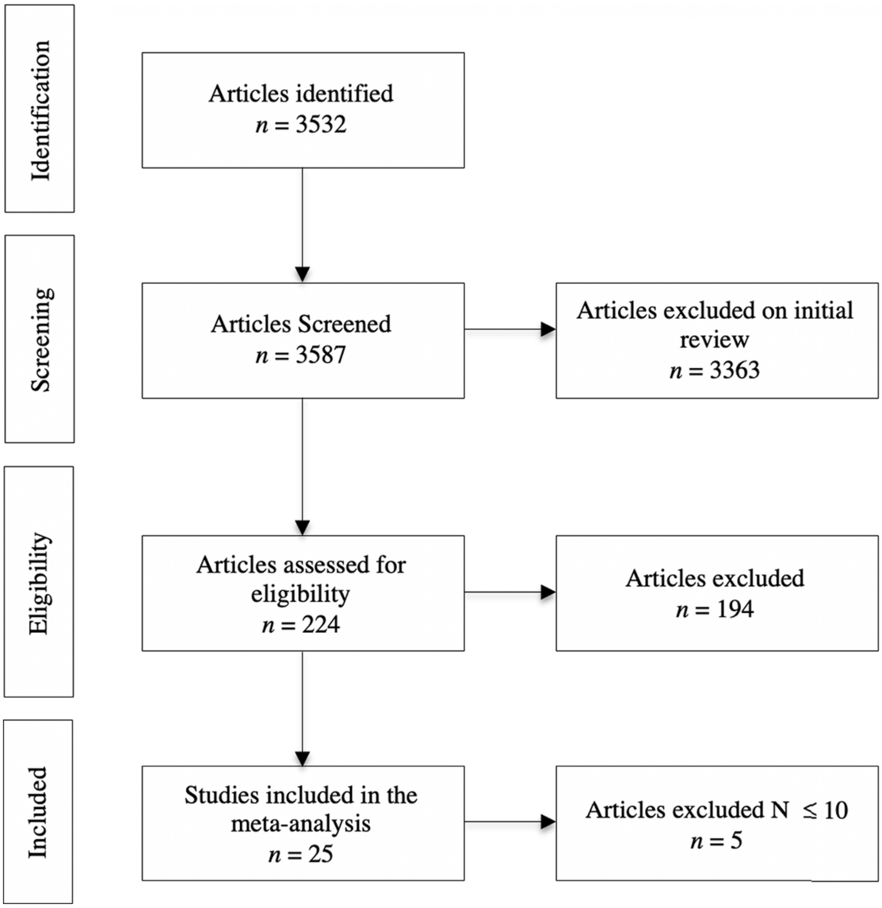


Fig. 1. Prisma Flow Chart of Study Selection.

Design

The two designs commonly employed in this field are concurrent and longitudinal. We anticipated, as per the results of Colonnesi et al (2010), that the strength of the association would be stronger for concurrent measures. The amount of variance captured in longitudinal studies is likely to be increased as other factors exert their influence on infant's language performance.

In summary, the current study aimed to examine the mean magnitude of the gesture-language relationship and consider the possible moderating effects of bilingualism, SES, language modality, language measure, language spoken and country, mean age of pointing assessment, pointing measure, pointing variable, and methodological design.

Method

Study selection

A protocol for this systematic review was developed and registered on PROSPERO before searching commenced (Kirk et al., 2016, Prospero registration number: CRD42016047246). The review's eligibility criteria were specified in the protocol. Studies which reported monolingual and/or bilingual infant samples, aged between 6 and 48 months of age, were eligible. Studies using monolingual only samples, bilingual only samples and monolingual and bilingual samples were all eligible. Infants could be from any country of origin and could be exposed to any language. Studies of infants that were born prematurely (<37 weeks gestation), were deaf, had a hearing impairment, or had any developmental disorders were not eligible. Studies that compared children with and without developmental or other impairments were not eligible.

The review included two types of study design: concurrent and longitudinal. Case studies and studies that experimentally manipulated pointing were excluded. Only studies that had recorded infant pointing using video methodology were eligible. This was to ensure consistency across studies and to enhance reliability in the key variable of interest (infant pointing production). The primary outcome of interest in the review was infant language, both expressive and receptive. This was measured by parental report, observation or standardized measures of language ability. The review considered the contribution of a number of variables to the association

Table 1
Papers by county and year of publication.

Country	1984–1990	1991–2000	2001–2010	2010–2019	Total
Australia	–	–	–	1	1
Canada	–	1	1	–	2
France	–	–	–	1	1
Germany	–	–	–	2	2
Italy	–	–	3	1	4
Spain	–	–	–	2	2
UK	–	–	–	4	4
USA	1	3	7	4	15
Total	1	4	11	15	**31***

* one paper included two samples, one from Italy and one from USA.

between pointing and language, as previously outlined.

For practical reasons, only studies reported in English were eligible. Published journal articles, unpublished manuscripts, conference proceedings, theses (masters or PhD) and end of award reports published on funders' websites were all eligible. Editorials, notes, and news briefings were excluded. No publication date limits were set and the search included papers published up until January 2019.

Searches were conducted in a range of databases and websites to identify published and unpublished studies. The following databases were searched: PsycINFO, MEDLINE, Scopus, Science Citation Index, Social Science Citation Index and ERIC. The full details of the searches can be found in Appendix A. Conference proceedings from the International Society of Gesture Studies (2002 to present) were searched and seven resources were used to search for theses. The websites of two key research funders were also searched to identify end of award reports (RCUK Gateway to Research, NIH Research Portfolio Online Reporting Tools), and ResearchGate and Academia.Edu were searched using the named authors of eligible studies. Google Scholar was searched to identify papers that cite the eligible studies. Systematic and non-systematic reviews identified from the searches and published since 2009 were screened to harvest additional studies. Researchers in the field were also contacted and asked to suggest relevant publications.

Two independent reviewers screened the search results independently against the eligibility criteria using information in the title and abstract, and then from the full text of any documents which seemed likely to be eligible. Screening was conducted using Covidence Systematic Review Software. Any disagreements, at either stage, were discussed with a third reviewer. The study selection was piloted to ensure high rates of agreement and to clarify the eligibility criteria if necessary. The study selection process is shown in Fig. 1.

Data from each of the eligible studies were extracted into a spreadsheet by one reviewer and checked by a second. Any discrepancies were resolved through discussion or by consulting a third reviewer. Details of bibliographic data, population, study design, results and moderators were extracted.

Coding the studies

Studies were coded by two coders as a function of the following characteristics: Infant age at recruitment and age at points of measurement (months); infant gender; Monolingual or Bilingual; Socio-economic status (SES) (information on parental education, income); Country; language spoken; sample size; publication status (e.g. peer-reviewed journal article, unpublished thesis, conference presentation); Year of publication; study design (longitudinal, concurrent, experimental); language modality (language production or comprehension); Language measure (the assessment used, e.g. CDI, PPVT, Reynell Scales, and what aspect of language this measured (e.g. vocabulary, sentence comprehension, syntactical skill); mean age of pointing assessment; The way in which pointing was operationalised (pointing frequency, age of onset of pointing); Whether pointing was coded alone or in combination with speech; study design (concurrent or longitudinal).

Few studies reported sufficient information to allow SES to be coded. There were only two papers (Nicoladis et al. 1999; Nicoladis, 2002) that included bilingual samples therefore it was not possible to conduct a moderator analysis on this variable. A wide range of language measures were used, therefore we coded the language measure as 'vocabulary' (CDI, PPVT) or 'other' (assessment of language not specific to vocabulary, e.g. the Reynell Scales). It had been our intention to consider the impact of language and culture, however the studies identified were largely drawn from English speaking samples from the USA or Canada while the remaining were from a range of different countries with a diverse range of languages spoken with typically one study per country or language therefore precluding moderator analysis. Thus, we simply distinguished between 'North America' (USA and Canada) and 'other' for Country (this included UK, European countries and Australia), and 'English' or 'Other' for language, with some reporting a 'mixed' sample (English and languages other than English). With regards to the way in which pointing was operationalised, studies were coded as either pointing onset (e.g. age of onset) or pointing frequency (e.g. raw frequency, frequency per minute or trials, gesture vocabulary). Please see Table 1 for a summary of papers published by country and year.

Calculation of effect sizes

Effect sizes were transformed to Pearson's *r* and 95% confidence intervals were calculated. For studies that reported Spearman's r

Table 2
Effect sizes and moderators for the 30 eligible studies.

Article	r	n	Language modality	Language measure	Language	Country	Mean age pointing	Pointing measure	Pointing combination	Design
Aureli et al. (2008)*	0.15	18	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Aureli et al. (2008)*	0.57	18	P	Vocab	Other	Other	≤ 18 m	Freq	G	L
Aureli et al. (2008)*	0.16	18	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Aureli et al. (2008)*	0.02	18	P	Vocab	Other	Other	≤ 18 m	Freq	G	L
Aureli et al. (2008)*	0.32	18	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Aureli et al. (2008)*	0.24	18	P	Vocab	Other	Other	≤ 18 m	Freq	G	L
Bidgood et al. (2016)	0.17	48	P	Vocab	English	Other	≤ 18 m	Freq	G	L
Bidgood et al. (2016)	0.25	68	C	Vocab	English	Other	≤ 18 m	Freq	G	L
Brooks et al. (2008)*	0.40	32	P	Vocab	English	North America	≤ 18 m	Onset	G	1
Brooks et al. (2008)*	0.44	20	P	Vocab	English	North America	≤ 18 m	Onset	G	L
Brooks et al. (2008)*	0.39	23	P	Vocab	English	North America	≤ 18 m	Onset	G	L
Brooks et al. (2008)*	0.60	25	P	Vocab	English	North America	> 18 m	Onset	G	L
Carpenter et al. (1998)*	0.40	24	P	Other	English	North America	≤ 18 m	Onset	G	L
Carpenter et al. (1998)*	0.19	24	P	Other	English	North America	≤ 18 m	Onset	G	L
Carpenter et al. (1998)*	0.08	24	P	Other	English	North America	≤ 18 m	Onset	G	L
*Cheong (2015)	−0.19	14	C	Vocab	English	North America	≤ 18 m	Freq	G	C
Cheong (2015)	0.20	12	C	Vocab	English	North America	> 18 m	Freq	G	C
Cheong (2015)	−0.11	14	P	Vocab	English	North America	≤ 18 m	Freq	G	C
Cheong (2015)	0.18	12	P	Vocab	English	North America	> 18 m	Freq	G	C
Cochet & Byrne (2016)	0.77†	13	P	Vocab	English	Other	> 18 m	Freq	G	C
Cochet & Byrne (2016)	0.84†	13	C	Vocab	English	Other	> 18 m	Freq	G	C
Cochet & Byrne (2016)	0.57†	13	P	Vocab	English	Other	> 18 m	Freq	G	C
Cochet & Byrne (2016)	0.55†	13	C	Vocab	English	Other	> 18 m	Freq	G	C
Colonnesi et al. (2008)*	−0.16	35	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Colonnesi et al. (2008)*	−0.06	35	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Colonnesi et al. (2008)*	−0.02	35	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
*Colonnesi et al. (2008)	−0.06	35	C	Vocab	Other	Other	≤ 18 m	Freq	G	L
Desrochers et al. (1995)*	0.43	23	P	Other	English	North America	≤ 18 m	Onset	G	L
Desrochers et al. (1995)*	0.42	23	C	Other	English	North America	≤ 18 m	Onset	G	L
Dobrich & Scarborough (1984)*	0.40	22	P	Other	English	North America	> 18 m	Freq	G	C
Esseily et al. (2011)	0.23	22	P	Vocab	Other	Other	≤ 18 m	Onset	G	C
Esseily et al. (2011)	0.37	22	C	Vocab	Other	Other	≤ 18 m	Onset	G	C

(continued on next page)

Table 2 (*continued*)

Article	*r*	*n*	Language modality	Language measure	Language	Country	Mean age pointing	Pointing measure	Pointing combination	Design
Fasolo & D'Odorico (2012)	0.52	24	P	Vocab	Other	Other	≤ 18 m	Freq	GS	L
Fasolo & D'Odorico (2012)	0.40	24	P	Other	Other	Other	≤ 18 m	Freq	GS	L
Fasolo & D'Odorico (2012)	0.13	24	P	Vocab	Other	Other	≤ 18 m	Freq	GS	L
Fasolo & D'Odorico (2012)	0.24	24	P	Other	Other	Other	≤ 18 m	Freq	GS	L
Goldin-Meadow & Butcher (2003)	0.90	6	P	Other	English	North America	≤ 18 m	Onset	GS	L
Goldin-Meadow & Butcher (2003)	0.46	6	P	Other	English	North America	≤ 18 m	Onset	GS	L
Hall et al. (2013)	0.33	50	P	Vocab	English	Other	> 18 m	Freq	G	C
Igualada et al. (2015)	−0.05	19	P	Vocab	Other	Other	≤ 18 m	Freq	G	L
Igualada et al. (2015)	0.53	19	P	Vocab	Other	Other	≤ 18 m	Freq	GS	L
Iverson & Goldin-Meadow (2005)	0.94†	10	P	Other	English	North America	≤ 18 m	Onset	GS	L
Iverson & Goldin-Meadow (2005)	0.24†	10	P	Other	English	North America	≤ 18 m	Onset	GS	L
Iverson et al. (2008)	0.99	6	P	Other	Mixed	Mixed	≤ 18 m	Onset	GS	L
Iverson et al. (2008)	−0.48	6	P	Other	Mixed	Mixed	≤ 18 m	Onset	GS	L
Kuhn et al. (2014)	0.09	1066	P	Other	English	North America	≤ 18 m	Freq	G	L
Kuhn et al. (2014)	0.11	1066	P	Other	English	North America	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.33	57	C	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.29	55	C	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.20	59	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.17	53	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.20	59	P	Vocab	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.15	59	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke et al. (2017)	0.16	59	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke, et al. (2019)	0.11	37	C	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke, et al. (2019)	0.21	39	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke, et al. (2019)	0.28	38	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke, et al. (2019)	0.05	41	P	Other	Other	Other	≤ 18 m	Freq	G	L
Lüke, et al. (2019)	0.18	39	P	Other	Other	Other	≤ 18 m	Freq	G	L
McGillion et al (2017)	−0.10	46	P	Vocab	English	Other	≤ 18 m	Onset	G	L
McGillion et al (2017)	−0.24	46	C	Vocab	English	Other	≤ 18 m	Onset	G	L
Mumford and Kita (2016)	0.19†	16	P	Vocab	English	Other	≤ 18 m	Freq	G	C
Mumford and Kita (2016)	−0.05†	16	C	Vocab	English	Other	≤ 18 m	Freq	G	C
Murillo et al. (2015)	0.02	30	P	Vocab	Other	Other	> 18 m	Freq	G	C
Murillo et al. (2015)	0.03	27	P	Other	Other	Other	> 18 m	Freq	G	C

(continued on next page)

Table 2 (*continued*)

Article	*r*	*n*	Language modality	Language measure	Language	Country	Mean age pointing	Pointing measure	Pointing combination	Design
Nicoladis (2002)	−0.20	8	C	Vocab	Other	North America	> 18 m	Freq	G	C
Nicoladis (2002)	−0.08	8	C	Vocab	English	North America	> 18 m	Freq	G	C
Nicoladis (2002)	0.34	8	P	Other	Other	North America	> 18 m	Freq	G	C
Nicoladis (2002)	0.32	8	P	Other	English	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	−0.20	5	P	Other	English	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	−0.03	5	P	Other	English	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	−0.64	5	P	Other	English	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	−0.52	5	P	Other	English	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	0.08	5	P	Other	Other	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	0.05	5	P	Other	Other	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	−0.54	5	P	Other	Other	North America	> 18 m	Freq	G	C
Nicoladis et al. (1999)	0.20	5	P	Other	Other	North America	> 18 m	Freq	G	C
Rowe (2000)*	0.64	45	P	Vocab	English	North America	≤ 18 m	Freq	G	C
Rowe & Goldin-Meadow (2009a) Dev Sci	0.49	52	C	Vocab	English	North America	≤ 18 m	Freq	G	L
Rowe & Goldin-Meadow (2009a) Dev Sci	0.35	52	C	Vocab	English	North America	≤ 18 m	Freq	GS	L
Rowe & Goldin-Meadow (2009b) *Science	0.47	50	C	Vocab	English	North America	≤ 18 m	Freq	G	C
Rowe & Goldin-Meadow (2009b)* Science	0.61	50	P	Other	English	North America	≤ 18 m	Freq	G	C
Rowe et al. (2008)	0.41	53	C	Vocab	English	North America	≤ 18 m	Freq	G	L
Rowe et al. (2008)	0.52	53	C	Vocab	English	North America	≤ 18 m	Freq	G	L
Rowe et al. (2008)	0.38	53	C	Vocab	English	North America	≤ 18 m	Freq	GS	L
Tamis-LeMonda et al. (2012)	−0.02	226	C	Vocab	Mixed	North America	≤ 18 m	Freq	G	C
Tamis-LeMonda et al. (2012)	0.16	226	C	Vocab	Mixed	North America	> 18 m	Freq	G	L
Tamis-LeMonda et al. (2012)	−0.11	226	P	Vocab	Mixed	North America	≤ 18 m	Freq	G	C
Tamis-LeMonda et al. (2012)	−0.07	226	P	Vocab	Mixed	North America	> 18 m	Freq	G	L
Tamis-LeMonda et al. (2012)	−0.08	226	P	Other	Mixed	North America	≤ 18 m	Freq	G	L
Tamis-LeMonda et al. (2012)	0.08	226	P	Other	Mixed	North America	> 18 m	Freq	G	C
Tamis-LeMonda et al. (2012)	0.06	226	C	Other	Mixed	North America	≤ 18 m	Freq	G	L
Tamis-LeMonda et al. (2012)	0.20	226	C	Other	Mixed	North America	> 18 m	Freq	G	C
Wu & Gros-Louis (2014)	0.33	51	P	Vocab	English	North America	≤ 18 m	Freq	G	L
Wu & Gros-Louis (2014)	0.12	51	C	Vocab	English	North America	≤ 18 m	Freq	G	L

Note. * indicates articles included in the Colonnesi meta-analysis (2010). P = production, C = Comprehension; Vocab = vocabulary measure, Other = Other measure of language (not vocabulary); Freq – frequency of gesture production, Onset = Onset of gesture production; G gesture alone, GS = gesture produced with speech; L = longitudinal, C = Concurrent. † indicates Spearmans r. Because some studies included more than one effect size, we include multiple effect sizes per study. Other country = UK, Europe or Australia.

Table 3
Methodological quality assessment of the eligible studies.

Papers	Sample Size	Gesture measurement	Outcome measures	Confounds
Aureli et al. (2008)	Low	Medium	High	Low
Bidgood et al. (2016)	Medium	Medium	High	Medium
Brooks & Meltzoff (2008)	Low	High	High	Medium
Carpenter et al. (1998)	Low	High	High	High
Cheong (2015)	Low	Medium	High	Low
Cochet & Byrne (2016)	Low	High	High	Medium
Colonnesi et al. (2008)	Medium	High	High	High
Desrochers et al. (1995)	Low	High	High	Medium
Dobrich & Scarborough (1984)	Low	High	Medium	Low
Esseily et al. (2011)	Low	High	High	Medium
Fasolo & D'Odorico (2012)	Low	High	Medium	Low
Goldin-Meadow & Butcher (2003)	Low	High	Medium	Medium
Hall et al. (2013)	Medium	High	High	Medium
Igualada et al. (2015)	Low	High	High	Medium
Iverson & Goldin-Meadow (2005)	Low	High	Medium	Medium
Iverson et al. (2008)	Low	High	Medium	Medium
Kuhn et al. (2014)	High	High	High	High
Luke et al. (2017)	Medium	High	High	Medium
Luke et al. (2019)	Medium	High	High	High
McGillion et al. (2017)	Medium	High	High	Medium
Mumford & Kita (2016)	Low	High	High	Medium
Murillo et al. (2015)	Low	High	High	Medium
Nicoladis (2002)	Low	High	High	Medium
Nicoladis et al. (1999)	Low	High	Medium	Medium
Rowe (2000)	Medium	High	High	Medium
Rowe & Goldin-Meadow (2009a Dev Sci)	Medium	High	High	Medium
Rowe & Goldin-Meadow (2009b Science)	Medium	High	High	Medium
Rowe et al. (2008)	Medium	Medium	High	High
Tamis-LeMonda et al. (2012)	High	Medium	High	Medium
Wu and Gros-Louis (2014)	Medium	High	High	Medium

these have been included and the data analysed with and without them to check for inconsistencies. Several papers reported multiple effect sizes, and rather than averaging them we use a three-level meta-analysis. The reasons for multiple effect sizes was because studies reported different analyses on the same samples, which may have been longitudinal analyses at different age points, or analyses of different language or gesture measures. Positive effect sizes indicate a positive relationship between pointing frequency and language skill.

Authors were contacted to provide additional details if necessary. Studies were excluded if insufficient detail was reported in the paper and efforts to contact the authors were not responded to ($n = 1$) or authors responded to say they no longer had access to the data ($n = 1$). Table 2 reports a summary of the studies included in the meta-analysis and effect sizes.

Methodological quality

Following data extraction, the methodological quality of each eligible study was assessed. Available measures were deemed not appropriate to the design of the included studies, which were all correlational. Commonly used measures in systematic reviews are designed to assess intervention studies (e.g. the Cochrane Risk of Bias tool; the McMaster University Quality Assessment Tool for Quantitative Studies, National Collaborating Centre for Methods and Tools, 2008). Therefore, to review the potential sources of bias we adopted the approach of Murphy and Unthiah (2015) by developing a rating system for the key methodological features of the studies.

Each publication was evaluated across four dimensions that we identified to reflect important sources of bias: sample size, gesture measurement, language measures, and confounds (see Table B1 in the appendix for detailed description). Sample size referred to the power of the study, gesture measurement assessed the reliability and validity of the gesture measurement, outcome measures referred to the use of clearly defined, standardised measures of language and confounds referred to efforts made to address potential sources of bias. Two reviewers (EK, MW, RF) coded each publication independently on the four categories as *'high'*, *'medium'* or *'low'*. Where there was a discrepancy of one rank we deferred to the more conservative rating (i.e. if reviewer A rated a study as medium and reviewer B rated it as high, the study was coded as medium). If there was a discrepancy of two ranks then the middle ranking was adopted (i.e. if reviewer A coded a study as low, and reviewer B coded it as high, the study was coded as medium). The summary quality assessment is

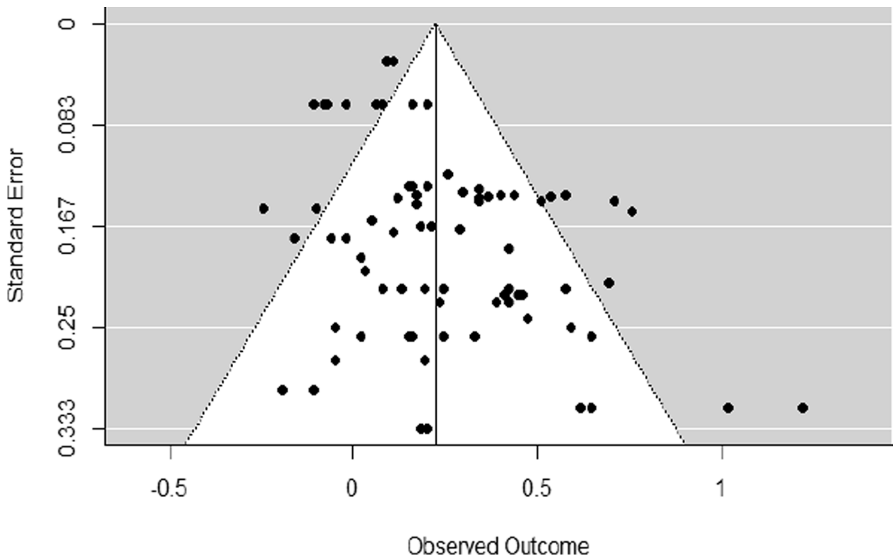


Fig. 2. Funnel plot for effect sizes from eligible studies.

shown in Table 3.

Due to the nature of the research, typically longitudinal studies of infants with high intensity coding of speech and gesture, the sample sizes are usually small. Only two of the papers were rated as 'high' on sample size (Kuhn et al. 2014; Tamis-LeMonda et al. 2012). It was necessary to remove five papers from the meta-analysis due to small sample sizes ($N \leq 10$)[1]. Typically, the measurement of gesture was robust, with papers reporting good inter-rater reliability on the coding of infant's gestures. Language was typically measured with standardised measures, or with well-defined and reliable measures. The papers varied in their handling of sources of bias. While some papers made explicit attempts to control for confounds (e.g. birth order, gender, socio-economic status) in statistical analyses, this was inconsistent across the sample of papers.

Analytic strategy

All analyses were conducted in R 4.0.2 using the metafor package (Viechtbauer, 2010). As some studies contributed multiple effect sizes from the same participants, effect sizes were not independent. To address the non-independence of effect sizes, we used two solutions: First, true effect sizes from the same participants are likely correlated, we used three-level meta-analysis and estimated random intercepts between and within studies. Second, because sampling errors for effect sizes were likely correlated, we used cluster-robust variance estimation within studies (Hedges, Tipton & Johnson, 2010). Publication bias was tested using Egger's regression test, and the PEESE method used by Lehtonen, Soveri, Laine, Järvenpää, de Bruin and Antfolk (2018). All effect sizes were converted to *z* scores before fitting models and model output was converted back to Pearson's *r* when average effect sizes are reported. The full data set and R scripts are available in the Open Science Framework (see https://osf.io/rk4qd/). See https://rpubs.com/sdonnelly85/784197 for all analyses and results in html formatting.

Results

We identified 30 papers published between 1984 and 2019 that met our stringent inclusion criteria, 25 of which (comprising 77 effect sizes) had samples ≥10 and were included in the meta-analysis (see Appendix C for justification for excluding effect sizes based on samples based on 10 or fewer participants). Prior to running substantive analyses, we assessed the influence of publication bias on effect sizes. Egger's regression test indicated significant publication bias ($Z = 3.50$, $p < .001$). Consistent with this, the funnel plot in Fig. 2 shows a clear relationship between effect size magnitude and precision, with standard errors increasing with effect sizes. In order to mitigate the effect of publication bias, we followed the example of Lehtonen et al. (2018) and adapted the PEESE method. We included effect size variances as a predictor variable and treated the intercept, the implied effect size when sampling variance as 0, as the true effect. As noted by Lehtonen et al. this method was developed in situations in which there was not dependence between effect sizes; results should, therefore, be interpreted with caution. We fit all models with and without observations for Spearman's *r*. We

[1] These studies yielded several effect sizes each and were based on sample sizes of 5, 6, 8.and 10. Because of the small sample sizes, they produced extremely large effect size variances (See first funnel plot in R Markdown file), which would strongly affect the publication bias corrections. Because correlations on sample sizes this small are likely unreliable to being with, we removed them from our analyses. Please see Appendix C for a detailed discussion of this issue.

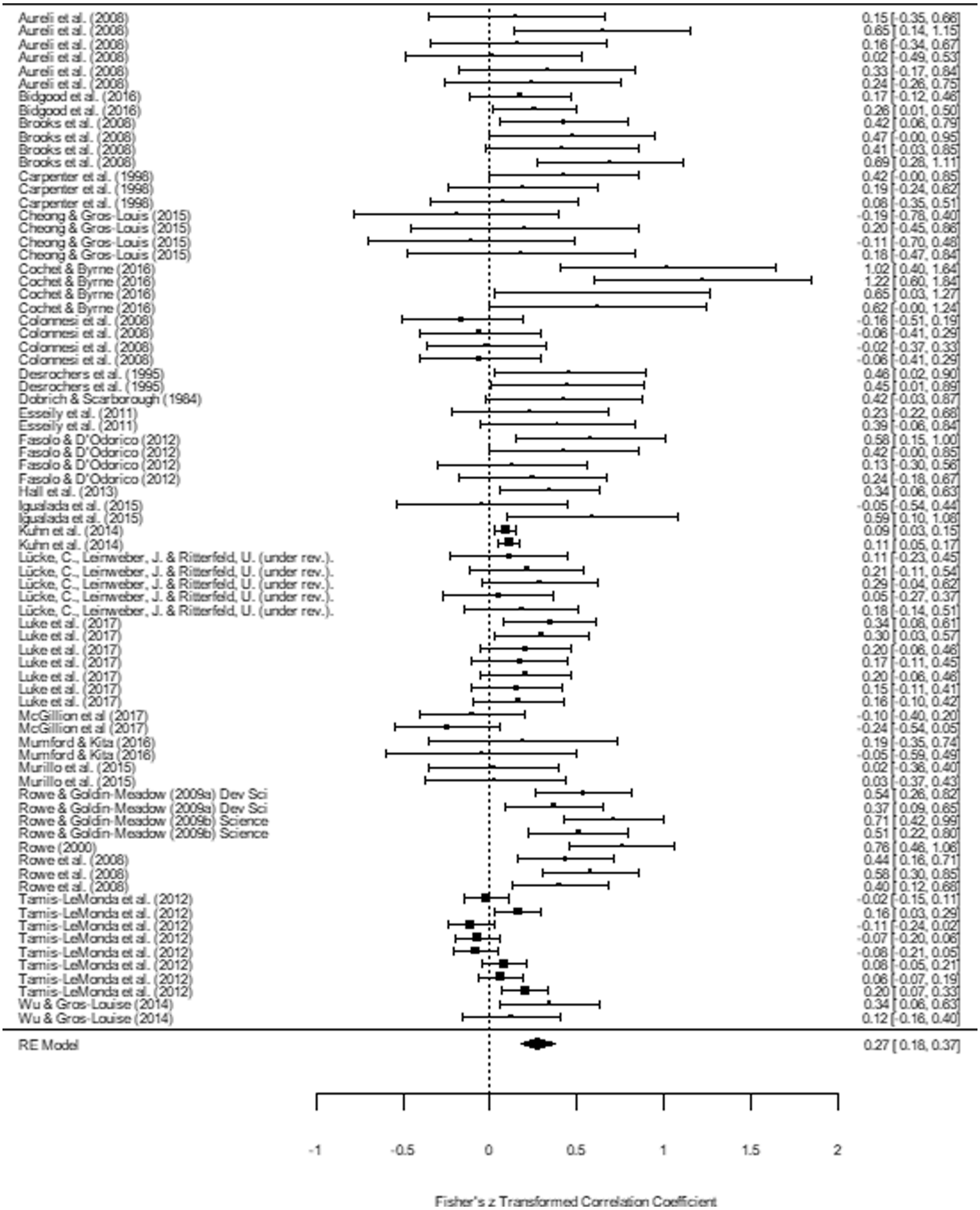


Fig. 3. Forest plot of three-level meta-analysis, prior to correction for publication bias.

Table 4
Estimated effect sizes and confidence intervals by moderators.

		k	*N*	*r*	95 % CI	Contrast F
Language Modality	Comprehension	30	1838	0.27	0.08, 0.43	F(1, 21) = 15.3,
	Production	47	4324	0.14	-0.05, 0.33	*p* <.001
Language Measure	Vocabulary	50	2325	0.17	-0.04. 0.36	F(1, 21) = 8.66,
	Other	27	3837	0.23	0.05, 0.40	*p* =.008
Age of pointing assessment	< 18 months	62	5028	0.22	0.02, 0.40	F(1, 21) = 50.0,
	> 18 months	15	1134	0.32	0.14, 0.47	*p* <.001
Pointing Combination	Pointing alone	70	5942	0.20	0.00, 0.38	F(1, 21) = 15.6,
	Pointing + Speech	7	220	-0.06	-0.25, 0.14	*p* <.001
Language	English	37	3276	0.28	0.04: 0.49	F(1, 20) = 2.9,
	Mixed	8	1808	0.02	-0.00: 0.03	*p* =.077
	Languages Other than English	32	1078	0.02	-0.25: 0.28	
Country	North America	34	4742	0.32	0.07: 0.53	F(1, 21) = 3.9,
	Other	43	1420	-0.04	-0.31: 0.24	*p* =.063
Pointing measure	Pointing Frequency	64	5808	0.20	0.02: 0.36	F(1, 22) = 0.09,
	Pointing Onset	13	354	0.16	-0.17: 0.47	*p* =.777
Design	Concurrent	24	1390	0.22	0.02: 0.41	F(1, 21) = 1.5,
	Longitudinal	53	4772	0.19	0.02: 0.34	*p* =.227

Note. Each row refers to a level of the relevant moderator variable (indicated in the far left corner). k = number of effect sizes at that level of the moderator, N = total sample size across all effect sizes at that level of the moderator, r = model-implied average correlation coefficient for that level of the moderator; 95% CI = confidence interval for r; Contrast F = test of heterogeneity across levels of the relevant moderator. Note that for all moderators with two levels (everything except language), F is the squared t statistic for the difference score between levels of the relevant moderator (b).

report results on the full dataset and note when results differ across the two datasets in footnotes.

After correcting for publication bias, the overall effect size was small and significant ($r = 0.20$, $t(23) = 2.21$, $p = .037$, *CI (Confidence Interval)* = 0.01: 0.36, $k = 77$, *studies* = 25). The relationship between effect size variances and magnitude was positive but non-significant ($b = 2.07$, $t(23) = 0.92$, $p = .366$, *CI* =.-2.57: 6.72). Moreover, there was significant heterogeneity amongst effect sizes ($Q(75) = 161.70$, $p < .001$), justifying the consideration of moderating variables[2]. Fig. 3 illustrates the effect size magnitudes from each study in a forest plot, demonstrating a high level of variance between effect sizes.

We considered each moderator variable (sum coded as −1 or 1) in a separate model. In order to assess the significance of moderator variables, we first tested whether all levels of the moderator variable significantly improved model fit. We then examined individual regression coefficients to determine which levels of the moderator variables differed significantly. Next, to determine whether effect sizes for specific levels of the moderator were significant, we obtained model-implied predicted effect sizes and confidence intervals for each level of the moderator variable. In order to mitigate the effects of differential publication bias across levels of the moderator variables, we allowed moderators to interact with effect size variances (see Table 4 for sumamry of moderator analyses).

The effect of modality was significant ($F(1, 21) = 15.27$, $p < .001$). Examination of model coefficients indicated that effect sizes for comprehension tasks were larger than those for production tasks ($b = 0.06$, $t(21) = 3.91$, $p < .001$, *CI* = 0.03: 0.10). The average effect size for comprehension tasks was small, with a confidence interval that did not overlap with 0 ($r = 0.27$, *CI* = 0.08: 0.43, $k = 30$), while the average effect size for production tasks was very small with a confidence interval that did overlap with ($r = 0.14$, *CI* = -0.05: 0.33, $k = 47$).

The effect of language measure was significant ($F(1, 21) = 8.66$, $p = .008$). Examination of coefficients revealed that effect sizes calculated using other language measures were larger than those using vocabulary ($b = 0.03$, $t(21) = 3.49$, $p = .007$, *CI* = 0.01: 0.05). The average effect size for other language measures was very small with a confidence interval that did not overlap with 0 ($r = 0.23$, *CI* = 0.05: 0.40, $k = 27$), while the average effect size for vocabulary was very small with confidence intervals that did overlap with 0 ($r = 0.17$, *CI* = -0.03: 0.36, $k = 50$). The effect of age of pointing was significant ($F(1, 21) = 49.98$, $p < .001$). Examination of model co-efficients indicated that effect sizes were larger when pointing was measured after 18 months than when pointing was measured prior to 18 months ($b = 0.05$, $t(18) = 7.07$, $p < .001$, *CI* = 0.04: 0.07). The average effect size when pointing was measured after 18 months was small with a confidence interval that did not overlap with 0 ($r = 0.32$, *CI* = 0.14: 0.47, $k = 15$), while the average effect size when pointing was measured before 18 months was very small with a confidence interval that did not overlap with 0 ($r = 0.22$, *CI* = 0.02: 0.39, $k = 62$).

The effect of pointing combination was significant ($F(1, 21) = 15.63$, $p < .001$). Examination of model coefficients revealed that effect sizes for pointing alone were larger than those that used gesture speech combinations ($b = 0.13$, $t(21) = 3.95$, $p < .001$, *CI* = 0.06: 0.20). The average effect size for pointing alone was very small with a confidence interval that did not overlap with 0 ($r = 0.20$, *CI* = 0.00: 0.38, $k = 70$), and the average effect size for gesture-speech combinations was negative with a confidence interval that overlapped with 0 ($r = -0.06$, *CI* = -0.26: 0.14, $k = 7$).

Each of the other moderator variables were non-significant. However, for the purposes of comparisons with other studies, we present model-implied effect sizes for each level of the moderator. Language did not significantly moderate effect sizes ($F(2, 20) =$

[2] The results of the test of heterogeneity are from a three-level l meta-analysis without cluster robust variance estimation.

2.93, p =.077)[3,4]. Effect sizes for English studies were small-to-medium and with a confidence interval that did not overlap with 0 (r = 0.28, CI = 0.04: 0.49, k = 37), while effect sizes for mixed and other languages were very small with confidence intervals that did overlap with 0 (r = 0.02, CI = -0.00: 0.03, k = 8 and r = 0.02, CI = -0.25: 0.28, k = 32, respectively). The effect of country was non-significant (F(1, 21) = 3.86, p =.063. Effect sizes for North American studies were small-to-medium with a confidence interval that did not overlap with 0 (r = 0.33, CI = 0.07: 0.53, k = 34), while the effect sizes for other studies very small and negative with a confidence interval that overlapped with 0 (r = - 0.04, CI = -0.31: 0.24, k = 43). Pointing Measure (frequency or onset) did not significantly moderate effect sizes (F(1, 21) = 0.083, p =.777)[5]. Effect sizes for frequency-based measures was small with a confidence interval that did not overlap with 0 (r = 0.22, CI = 0.03: 0.39, k = 64), while effect sizes for onset-based measures were negative and non-significant with a confidence interval that overlapped with 0 (r = 0.16, CI = -0.17: 0.46, k = 13). Design did not significantly moderate effect sizes (F(1, 21) = 1.54, p =.227). Effect sizes for both concurrent and longitudinal studies were small with confidence intervals that did not overlap with 0 (r =. 22, CI = 0.02: 0.41, k = 24, and z = 0.19, CI = 0.02: 0.34, k = 53, respectively).

Discussion

The close relationship between pointing and language has long been posited to be a causal one, with the prevailing theoretical account explaining this association in terms of the social-interactional context, such that infant pointing elicits timely linguistic input which builds vocabulary (e.g. Goldin-Meadow, Goodrich, Sauer, and Iverson (2007). The current study aimed to use a three-level meta-analytic approach to estimate the strength of the relationship between infant pointing and language development and considered the possible moderating effects of bilingualism, SES, language modality, language measure, language spoken and country, mean age of pointing assessment, pointing measure, pointing variable, and design. We analysed effect sizes from 25 papers (comprising 77 effect sizes) published between 1984 and 2019. We did not identify any eligible papers that included bilingual samples, thus our analyses are conducted on monolingual infants only. Overall, we found a small but significant effect of pointing on language development (r = *0.20*). Our results suggest that the unique contribution of pointing to language development is less substantial than it has been previously thought to be. The presence of significant heterogeneity amongst the studies justified the exploration of moderators, yielding interesting results identifying factors which differentiate between significant and non-significant pointing-language associations.

Moderator analyses revealed that there were significant effects of modality, language measure, age of pointing assessment and pointing combination. There was no significant effect of the other moderators: language, country, pointing measure, or design. We discuss each of the significant moderator analyses in turn. There was a significant effect of modality, such that studies that assessed language comprehension were significant, whereas those that assessed language production were not. This may be explained, in part, because comprehension precedes production in language development. If we view infant gesture as a mechanism to elicit object labelling from caregivers, then it would follow that pointing would have a more direct relationship to infant's comprehension of labels, than their production. Alternatively, considering the inverse causal relationship, it may be infants' underlying cognitive capacity for language that drives the onset and frequency of pointing, such that children with greater understanding of language use gestures more frequently. Indeed, there is evidence that six to nine-month-old infants understand common words (e.g. Bergelson & Swingley, 2015), some months prior to the onset of pointing. Large cohort studies have similarly found that gesture is a better longitudinal predictor of language comprehension than production (Bavin et al., 2008; Fenson et al., 1994; Fenson et al., 2007; Zambrana et al., 2013).

The effect of language measure was significant, such that effect sizes calculated using vocabulary measures were not significant (i.e. had confidence intervals that included zero) and were smaller than the effect sizes calculated using dependent measures other than vocabulary. Because the measures of language were greatly heterogeneous (in our sample of 30 papers>20 different language measures were used), a broad distinction was made between language measures that assessed vocabulary (a significant number of studies used a version of the CDI) or non-vocabulary aspects of language. It was the latter that was significant (i.e. confidence intervals did not include zero), interesting given the mixed bag of measures yet despite the heterogeneity the effect was small but significant. This is surprising given that studies have widely reported correlations between pointing and vocabulary. However, upon examination of the individual effect sizes, these are highly variable (ranging from -0.011 to 0.84) with the majority small to moderate. The importance of pointing as a predictor for vocabulary may have been overemphasised as when averaged over different studies the relationship is reduced.

The effect of age of pointing was significant, with larger effect sizes when pointing was measured after 18 months than when pointing was assessed before 18 months. The effect size for studies of children > 18 months was r = 0.32, which was the highest effect

[3] Language was the only moderator with more than two levels. As a result, we were not able to include the full interaction between Language and effect size variance; as the three interaction terms would be linear combinations of one another. We, therefore, only included one interaction term, not two.

[4] Language was significant when only effect sizes reporting on Pearson's r was considered (p =.038). This was driven by a significant difference between effect sizes in English (r = 0.32) and other languages (r = 0.02). Given that this effect was not significant in the entire set of effect sizes, and for consistency with the relatively stringent criteria applied throughout this paper, we do not interpret the moderating effect of language as significant, but report this difference across data sets for completeness.

[5] Note that when the effect size variance was allowed to interact with Pointing Measure, the resulting effect size estimate for Onset was implausible (r = -0.28), so we only included a main effect for effect size variance. Inferences about the pattern of moderation were not different across model specifications.

size found in our analysis. Pointing production increases in the child's second year, thus studies of children older than 18 months may have greater variability in their gesture variables and therefore more predictive power. Pointing also takes on a different role as infants gain in language proficiency. Children point alongside their speech as they transition to two-word spoken utterances, and while a subset of studies specifically analysed such gesture-speech combinations (e.g. Özçalışkan and Goldin-Meadow, 2009), most studies do not differentiate pointing gestures produced alone or in conjunction with speech. Thus, while we were able to conduct a moderator analysis on the handful of studies that specifically coded gesture-speech combinations, it is highly likely that the studies of children older than 18 months are measuring gesture produced in combination with speech, which may serve a different function to points produced without accompanying speech. Before the onset of spoken vocabulary, infants use their pointing gestures for imperative and declarative functions, to request and share attention (Bates et al., 1975). Once children can talk, this function inevitably changes as pointing alongside speech may serve to complement or supplement their spoken utterance. A recent longitudinal study of over 100 British infants by Donnellan, Bannard, McGillion, Slocombe & Matthews (2020) found that at 11 months of age, pointing was a less important predictor of expressive vocabulary at 24 months than intentionally communicative vocalisations. The authors similarly suggest that gestures may become more important later in development.

The last moderator analysis that was significant was the effect of pointing combination, such that the overall effect size for studies that examined the association between gesture-speech combinations and language was not significant ($r = -0.06$), whereas the average effect size for studies that considered pointing alone was larger and significant ($r = 0.20$). However, this distinction between studies is problematic as there was significant heterogeneity amongst the small sample of seven studies that analysed gesture-speech combinations. These papers differ in their measurement of gesture-speech combinations, including onset of combinations, frequency, gesture tokens, types and proportions. Additionally, the outcome variables included vocabulary size, MLU, the onset of two-word speech and scores on the PPVT. Thus, the overall non-significant effect size likely reflects the methodological diversity amongst this small pool of studies.

We look in detail at the evidence presented in the paper with the highest quality scoring. Kuhn et al. (2014) had the largest sample size of the papers ($n = 1066$) and also likely to be the most diverse as they sampled from low-income families, African American families and the mean level of educational attainment was the equivalent of a high school diploma. They report the longitudinal association between pointing at age 15 months and language at 24 months ($r = 0.09$) and 36 months ($r = 0.11$), small effect sizes but noteworthy given the persistence over a 9–21 month lag between measures. The papers with the highest effect sizes (>0.8, Cochet & Byrne, 2016; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Iverson et al. 2008) tended to have small sample sizes ($n \leq 13$) (two of which are drawn from the same sample). Interestingly these higher effect sizes relate to the relationship between pointing and the onset of two word speech, indicating that this may be a particularly robust relationship, yet warrants investigation in a larger and more diverse sample.

Compared to the one previously published meta-analysis on the relationship between pointing and language, our overall effect size is considerably smaller. Colonnesi et al. (2010) reported an overall effect size of $r = 0.52$ for the association between pointing and language, based on a meta-analysis of 25 studies (and 25 effect sizes). Our meta-analysis samples converge on only 8 studies, since we excluded 17 papers in the Colonnesi review and included 15 papers published after 2010. Our review was exhaustive and we included multiple effect sizes ($k = 77$) from individual studies, thus utilising all available data to form a precise estimate of overall effect size. However, because of our stringent inclusion criteria (as well as our publication bias corrections), our data may represent a more conservative estimate of the relationship between pointing and language. Yet it is of note that our moderator analysis findings converge with those reported by Colonnesi et al (2010) on two main findings. Firstly, that there is a significantly larger effect size when language comprehension is measured compared to production, and secondly, that there is a significantly larger effect for the relationship between pointing and language in older children compared to younger infants (i.e. > 18 months).

How important is pointing for language development?

Gesture is a significant milestone in early language development, the absence of which may indicate developmental disorders such as autism (e.g. Watson, Crais, Baranek, Dykstra, & Wilson, 2013) or language delay/disorders (Capone & McGregor, 2004). The results of our meta-analysis indicate that there is a significant association between pointing and language development, but that this unique effect is small. Thus, while pointing can provide an important marker for forthcoming developments in a child's spoken language skill, we consider infant pointing as one of many important features of dyadic interaction that promote language acquisition. There is good evidence to suggest that, in US-European samples at least, it is the way in which pointing elicits interaction from others which is important. For example, a study by Olson & Masur (2015) found that it was the caregivers' responses to gesture that explained the association between pointing and language. Gesture facilitates joint attention and is an excellent device for eliciting contingent talk, which we know is an important predictor of language development (e.g. Rollins, 2003; McGillion et al. 2013). If we consider pointing in isolation it is a less valuable predictor of language than if we widen the lens to consider the impact of pointing on the infant's communicative partner and also potentially the level of attunement between that partner and the infant which may account for the quality and timing of the response to the infant's pointing acts.

Large cohort studies (which did not meet our inclusion criteria because gesture was measured using parental report) similarly report a minor contribution of gesture to language, once other factors are considered. Zambrana et al. (2013) report data from a large

Norwegian cohort study of 28,000 infants. Pointing at 18 months was found to correlate moderately with language production at 36 months but contributed no unique predictive value of late language production from 18 to 36 months. Similarly, Bavin et al. (2008) report data from the Early Language in Victoria cohort study, an Australian sample of 1447 children. They found children's early action and gesture production at 8 months predicted more variance in language comprehension (22.4%) than production (14.3%) at 12 months. Similar findings are reported by Westerlund, Berglund and Eriksson (2006) in a large Swedish study. Taken together, these findings from large cohorts indicate that the unique contribution of gesture is diminished when early language skills are taken into consideration.

Infants and their communicative partners engage in joint attention not only through pointing but also by alternating their gaze between an object and the other person, or through establishing eye contact with the partner (Carpenter, Nagell, Tomasello, Butterworth & Moore 1998; Tomasello & Farrar, 1986). Although a common socio-cognitive skill such as the ability to understand others' intentions has been suggested to underlie all of these attention directing behaviours, gestures and joint attention separately contribute to an infant's later language ability (Salo, Rowe, & Reeb-Sutherland, 2018). However, practices of language socialisation and attention sharing between infants and others show major differences cross-culturally. In some cultures, preverbal infants are rarely spoken to (Brown, 1998; Schieffelin & Ochs, 1986; Shneidman & Goldin-Meadow, 2012), or engaged in triadic interactions with others (Mastin and Vogt, 2016a; Mastin and Vogt, 2016b; Salomo & Liszkowski, 2013). When infants and caregivers do share attention, this might be achieved through the caregiver redirecting the infant's attention by manipulating the infant's body rather than following their infant's attention and pointing, and also by the infant using their bodily orientation rather than gestures to signal their attention (Abels, 2020). All this variation indicates different pathways to language development where for instance, dyadic interactions with or observation of others play a bigger role in infants' language development compared to joint attention or gestures (Mastin & Vogt, 2016). Thus, more research is needed into different developmental pathways to language and to establish what role, if any, gestures and joint attention play in different communities and cultures.

The initial aim of this meta-analysis was to review the evidence for the pointing-language association in monolingual and bilingual samples. Early on it was apparent that there was a dearth of research in this area, highlighting limitations in the literature. We identified two papers that included bilingual samples, Nicoladis (2002) and Nicoladis et al. (1999), however both were omitted from analysis due to small sample sizes ($n \leq 10$). Both studies, conducted on French-English bilinguals in Canada, reported non-significant correlations between deictic gestures and MLU in French and English between the ages of 24 and 42 months (Nicoladis et al., 1999) and gesture rate and receptive vocabulary in both languages at age four (Nicoladis, 2002). However, these studies were fine-grained longitudinal observations of the lexical development of a small sample of bilingual children, thus these analyses are insightful, but not conclusive. Research on bilingual samples is warranted, to understand the contribution of pointing to language acquisition in these infants who make up more than half of the world's population. We do not yet understand how, for infants exposed to more than one language, pointing contributes to their language development and how infants may use pointing differently depending on the language context.

The studies included in this meta-analysis lacked overall diversity, with the majority of studies drawing on samples from the US and Europe. However, there are studies which have examined cross-cultural differences in the development of pointing in infancy, the results of which are mixed with some finding evidence of universality of the onset and frequency of pointing in infants from different cultures (Liszkowski et al. 2012, Tamis-LeMonda et al., 2012), whereas other studies have noted cultural differences (Salomo & Liszowski, 2013). Only one study, to our knowledge, has considered the impact of culture on caregiver responsiveness to infant gesture. Very recently, Cameron-Faulkner et al. (2020) examined infant gesture in Bengali and Chinese mother-infant dyads living in the UK (with very low levels of English proficiency) compared to an English sample. Interestingly, there were no cultural differences in gesture use or maternal responsiveness, with infant gesture and maternal contingent talk produced in response to infant gesture at 10–12 months, predicting vocabulary at 18 months. This research further emphasises the importance of considering maternal response to gesture when examining the pointing-language association.

Limitations

Our stringent inclusion criteria meant that we excluded studies that measured infant gesture using parent report, a consequence of which was that we were unable to include data from large cohort studies. Additionally, due to practical reasons we excluded papers not reported in English. This means the review may exclude studies that might be eligible but published in languages other than English and their results might impact our findings in ways we cannot predict. Under-reporting of SES meant we could not include it in our moderator analysis however evidence suggests that this could be an important contributing factor (Rowe & Goldin-Meadow, 2009). As earlier identified, there exists significant heterogeneity in the methods used to assess the association between pointing and language, with great diversity in the way in which pointing is operationalised and how language is measured. We attempted to control for this in some way by using moderator analyses, however for some of the moderators it was necessary to make blunt distinctions to broadly categorise effect sizes, e.g. by comparing vocabulary and non-vocabulary dependent variables.

Furthermore, the moderator variables are not independent and there are likely to be confounds. Nevertheless, the moderator analyses are useful in revealing a profile of the study characteristics that yield the strongest effect sizes. A common issue in this field is the use of small sample size, not uncommon in infant research due to practical challenges. Unfortunately, this meant that it was necessary to remove five studies with samples ≤ 10 from the meta-analysis because their relatively large effect size variances would have disproportionately influenced the publication bias corrections (please see Appendix C for a detailed discussion of this issue). One practical implication of this meta-analysis is the provision of a conservative estimate of effect size for the association between pointing and language that can be used for power calculations to ensure future studies are sufficiently powered.

A conceptual question that we were not able to address is whether the intention of the infant point contributes to the association between pointing and language development. Infants' pointing intentions are typically identified as either imperative (a request) or declarative (sharing attention) (e.g. Bates et al. 1975). However, very few of the papers in our sample coded gesture intention, and those that did typically measured the number of points produced in trials designed to elicit imperative or declarative gestures (rather than each point being coded by observers as either declarative or imperative), which raises some concern about reliability and validity. Because the majority of papers did not specify the intention of the point we dropped gesture intention as a moderator because we would not have meaningful groups to compare. In the previous meta-analysis of the association between pointing and language, Colonessi et al (2010) report that the effect size for declarative points ($k = 14$) and papers that did not identify intention ($k = 8$) were both $r = 0.39$, compared to $r = 0.04$ for imperative points ($k = 3$). Notwithstanding our concerns about these groupings, this preliminary analysis may indicate that this is worthy of further investigation to fully understand the contribution of the perlocutionary effect of the gesture on the linguistic response.

Conclusion

Our meta-analytic investigation of the relationship between infant pointing and language development identified a small but significant overall effect size. A significant strength of this study is that we employed three-level meta-analysis, which allowed us to utilise all the available data rather than aggregating multiple effect sizes within each study. Thus, our analysis provides a more accurate estimate of effect size compared to using the conventional approach, additionally allowing us to analyse within-study variations arising from moderators. Our review identifies the significant lack of cultural diversity in the existing research. There was no evidence available to answer whether the association between pointing and language exists beyond monolingual WEIRD populations. Future research should consider testing the role of pointing in language development in more diverse populations, also considering the contribution of caregiver response to gesture to understand whether this proposed mechanism is culturally universal.

Author Note

We have no known conflict of interest to disclose.

The full data set and R scripts are available in the Open Science Framework at https://osf.io/rk4qd/. All analyses are available in html form at https://rpubs.com/sdonnelly85/784197.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link is given in the paper

Acknowledgements

This research was supported by a grant from the University of York Pump Priming Fund awarded to Elizabeth Kirk, and by the Australian Research Council (CE140100041). We would like to thank Amy Bidgood, Rechele Brooks, Malinda Carpenter, Cristina Colonnesi, Julie Gros-Louis, Alfonso Igualada, Carina Lüke, Eva Murillo Meredith Rowe, and Catherine Tamis-LeMonda for replying to requests for information and for providing us with additional data, explanation or clarification.

Appendix A. Searches

The search strategy for PsycINFO is shown in Table A1. The strategy is structured as follows:

Gestures/pointing (sets 1 to 5).

Table A1
Search strategy for PsycINFO (Ovid SP).

1 exp gestures/ (3598)
2 (pointing or gesture or gestures).ti,ab,id. (16694)
3 (manual adj act$1).ti,ab,id. (5)
4 (index adj finger$1).ti,ab,id. (1647)
5 gesturing.ti,ab,id. (424)
6 or/1–5 (18664)
7 exp language development/ or bilingualism/ (30830)
8 ((develop$ or emerg$) adj5 language).ti,ab,id. (21218)
9 (vocali$ or vocal referenc$).ti,ab,id. (8573)
10 (verbal adj5 interchange$1).ti,ab,id. (57)
11 vocabulary.ti,ab,id. (18889)
12 speech.ti,ab,id. (66289)
13 utterance$1.ti,ab,id. (8086)
14 vocabulary/ (6824)
15 oral communication/ (14159)
16 verbal communication/ (13239)
17 ((language or lexical) adj5 (acquisit$ or acquir$)).ti,ab,id. (10869)
18 (declare or declaring or request or requesting).ti,ab,id. (10317)
19 (talk or talking or (spoken adj language)).ti,ab,id. (31148)
20 (word or words or (lexical adj abilit$) or bilingual$ or second language).ti,ab,id. (135513)
21 language/ (35517)
Annotation: Not exploded - because to explode this would pick up topics that are not relevant
22 (oral$ adj4 communicat$).ti,ab,id. (1173)
23 (linguistic adj ability).ti,ab,id. (318)
24 or/7–23 (292110)
25 6 and 24 (5177)
26 limit 25 to english language (4729)
27 exp *communication disorders/ (45857)
28 (deaf or deafness or deafblind or hearing impair$).ti. (10471)
29 cochlear implant$.ti. (1731)
30 speech disorder$.ti. (364)
31 developmental disorder$.ti. (1386)
32 exp *developmental disabilities/ (10608)
33 developmental disabilit$.ti. (3005)
34 ((preterm or premature) and (infant$ or baby or babies or child or children or toddler$)).ti. (2643)
35 *premature birth/ (4062)
36 exp “primates (nonhuman)”/ (29857)
37 *autism spectrum disorders/ (28711)
38 (autism or autistic).ti. (25065)
39 *dyslexia/ (5309)
40 (dyslexia or dyslexic).ti. (4270)
41 ((language or speech) and impair$).ti. (3152)
42 sli.ti. (294)
43 ((speech or language) and disorder$).ti. (2534)
44 exp *attention deficit disorder/ (19704)
45 (asd or adhd).ti. (8365)
46 attention deficit.ti. (9456)
47 exp *Fragile X Syndrome/ (1252)
48 fragile x.ti. (1277)
49 exp *intellectual development disorder/ (26234)
50 autism.jw. (7665)
51 very low birth$.ti. (515)
52 (downs or down syndrome).ti. (4176)
53 *trisomy/ or *epilepsy/ (19208)
54 trisomy 18.ti. (16)
55 case report/ (22627)
56 exp *syndromes/ (123078)
57 rett syndrome.ti. (581)
58 *cerebral palsy/ or cerebral palsy.ti. (4258)
59 *brain damage/ or brain damag$.ti. (15114)
60 *traumatic brain injury/ or (brain and injur$).ti. (15607)
61 (“20″ not (”10″ and “20″)).po. (315745)
62 or/27-61 (598672)
63 26 not 62 (3571)
64 (infant or infants or infancy or infantile).ti,ab,id,hw. (84260)
65 exp early childhood development/ (28292)
66 (preschool$ or pre school$ or nursery or kindergarten).ti,ab,id. (53707)
67 (toddler$ or baby or babies or early childhood).ti,ab,id,hw. (44098)
68 (Pre linguistic or prelinguistic).ti,ab,id. (726)
69 (“120″ or ”140″ or “160″).ag. (146045)
70 or/64-69 (228894)
71 63 and 70 (796)

AND.
Language acquisition/development (sets 7 to 23).
AND.
Infants (sets 64 to 69).
NOT.

Communication disorders/developmental disorders/premature babies/nonhuman primates/animals/case reports (sets 27 to 61)

The searches are limited to studies reported in English (set 26). To minimise the adverse impact of using the NOT operator we have only excluded terms that appear in the title (e.g. brain damage$.ti.) and PsycINFO subject headings which have been designated as major subject headings (e.g. *autism spectrum disorders/). These approaches have been chosen to minimise the impact of missing studies which mention children with and without impairments in the same record.

Animal studies are removed safely by the options in sets 36 and 61. Set 36 removes studies indexed as involving nonhuman primates and set 61 finds studies involving animals, but not also involving humans using the population coding field in PsycINFO.

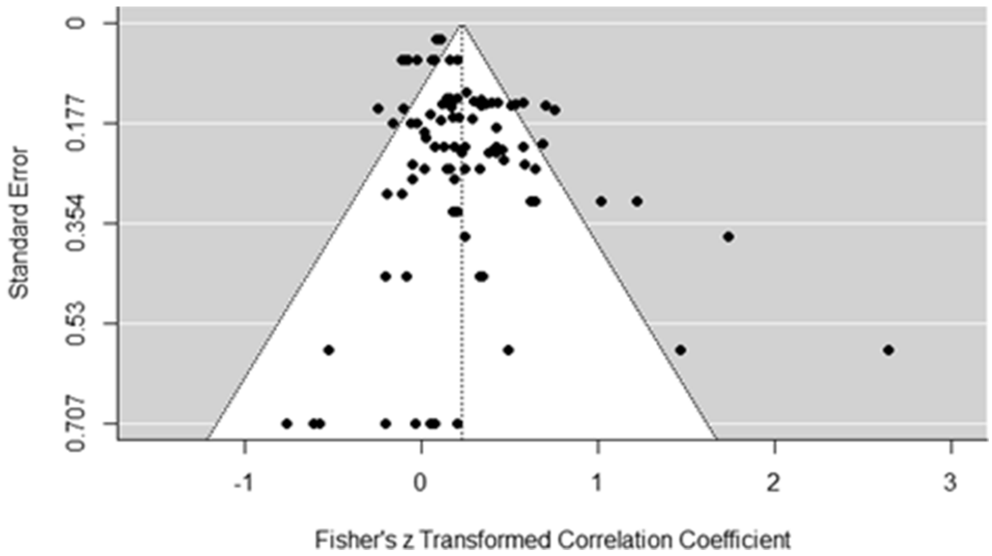
Appendix B. . Methodological quality assessment categories

Calculated using G*power with alpha 0.05. For medium effect (0.3) and strong power (0.8) n = 64, moderate power (0.6) = 38 and low power (0.4) = 22.

Appendix C. . Justification for removal of effect sizes based on small sample sizes

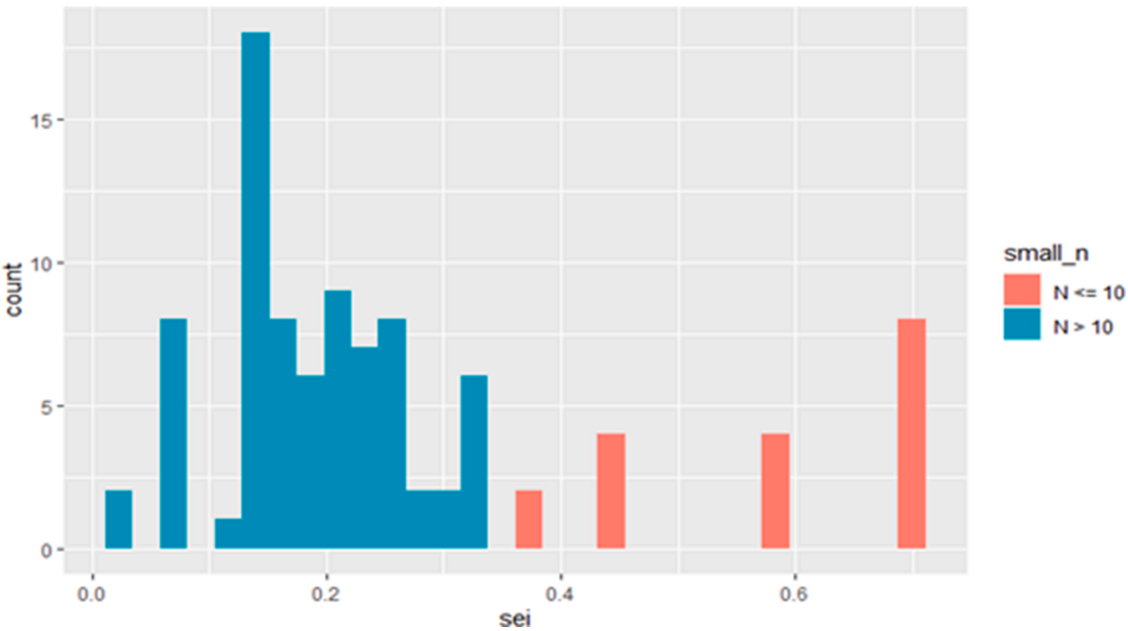
Several studies extracted during the literature search yielded multiple effect sizes with unusually small sample sizes. These effect sizes are problematic for publication bias corrections applied in our analyses because they necessarily produce extremely large effect size variances, and could act as leverage points having a disproportionate influence on the slope for the effect size variance. If this slope is biased, then the intercept the model-implied effect size when the sampling error is 0 will not be interpretable. That there are several large effect size variances can be seen in the funnel plot below (note that the y axis refers to the effect size standard error, which is the square root of the effect size variance):.

Strength of Evidence	Explanation	Sample Size*	Gesture measurement	Outcome measures	Confounds
High	Findings are highly secure. Bias has been controlled.	A good number of cases and well powered study (*N* = 64)	Measurement of infant gestures thorough and reliable with good inter-rater reliability.	Dependent variables clearly defined, valid, reliable, and implemented consistently Standardised measure of language.	Efforts made to address potential sources of bias, key potential confounding variables measured and adjusted statistically
Medium	Findings are moderately secure. Some potential sources of bias.	A medium number of cases (*N* = 38)	Valid measure of gesture, no measure of inter-rater reliability reported	Robust, valid outcome measure	Some attempt to address potential sources of bias.
Low	Findings are insecure. High risk of bias.	A small study or analysis of interest conducted on a small subset of infants. Low powered (*N* = 22)	Concerns about the validity and/or reliability of gesture measure	Concerns about the validity of the outcome measure	Potential sources of bias are not addressed



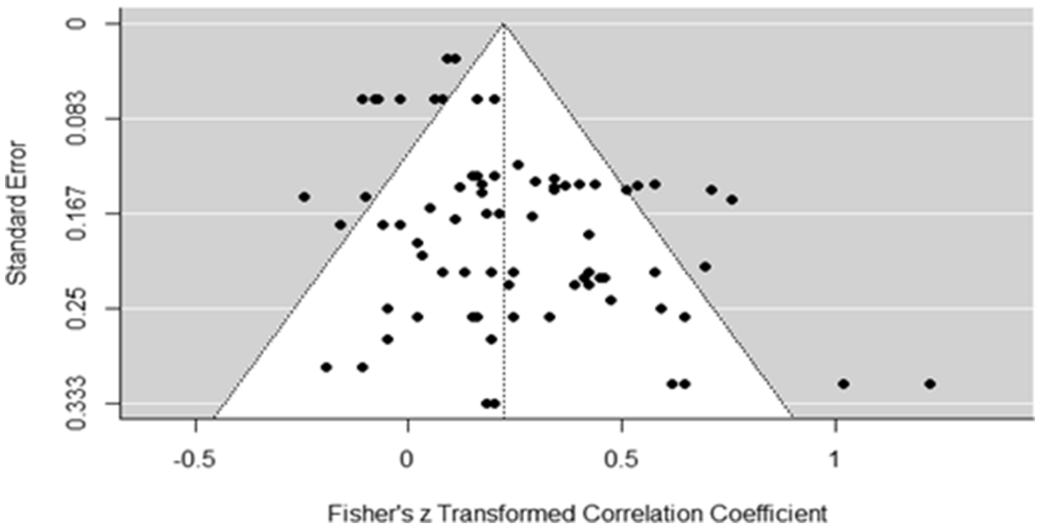
.

If we look at a histogram of the effect size standard errors, we see that all of the outlying standard errors come from sample sizes of 10 or less:.



.

These reflect five studies, which produced a total of 18 effect sizes. However, given that these effect sizes are (a) likely to disproportionately affect publication bias corrections, and thereby average effect size estimates and that (b) effect sizes based on such small sample sizes are likely not reliable to begin with, we removed these studies from our analyses. When these studies are removed, we see much less evidence of potentially overly influential effect size variances in the funnel plot. We also see, that the publication bias is not an artefact of those particularly small effect sizes, which is difficult to discern from the figure above.



.

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dr.2022.101023.

Reference

Abels, M. (2020). Triadic interaction and gestural communication: Hierarchical and child-centered interactions of rural and urban Gujarati (Indian) caregivers and 9-month-old infants. *Developmental Psychology*. https://doi.org/10.1037/dev0001094

Aureli, T., Perucchini, P., & Palazzo, A. (2008). The development of pointing from 9 to 15 months of age in imperative and declarative contexts: relationships with actions and language. *Poster presented at the XVIth biennial international conference on infant studies*.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development, 21*(3), 205–226.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology, 40*(3), 293–310. https://doi.org/10.1002/dev.10034

Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., … Ukoumunne, O. C. (2008). The early language in Victoria study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language, 35*(3), 687–701.

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development, 11*(4), 369–380.

* Bidgood, A., Cameron-Faulkner, T., Durrant, S., Peter, M. & Rowland, C.F. (2016). The role of gestures and babble in the vocabulary growth of children aged 8-18 months. Paper presented at the 1st Lancaster Conference on Infant and Child Development, Lancaster, UK.

Blake, J., Osborne, P., Cabral, M., & Gluck, P. (2003). The development of communicative gestures in Japanese infants. *First Language, 23*(1), 3–20. https://doi.org/10.1177/01427237030230001001

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language, 35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

Brown, P. (1998). Conversational structure and language acquisition: The role of repetition in Tzeltal. *Journal of Linguistic Anthropology, 8*(2), 197–221. https://doi.org/10.1525/jlin.1998.8.2.197

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy, 5*(3), 291–308. https://doi.org/10.1207/s15327078in0503_3

Cameron-Faulkner, T., Malik, N., Steele, C., Coretta, S., Serratrice, L., & Lieven, E. (2020). A cross-cultural analysis of early prelinguistic gesture development and its relationship to language development. *Child Development, cdev.13406*. https://doi.org/10.1111/cdev.13406

Capone, N. C., & McGregor, K. K. (2004). Gesture development. *Journal of Speech, Language, and Hearing Research, 47*(1), 173–186. https://doi.org/10.1044/1092-4388(2004/015)

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i. https://doi.org/10.2307/1166214

* Cheong, A. (2015) Infant pointing gesture and vocalization combinations: precursors to first labels? [unpublished thesis] University of Iowa.

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 1–10.

Cochet, H., & Byrne, R. W. (2016). Communication in the second and third year of life: Relationships between nonverbal social skills and language. *Infant Behavior and Development, 44*, 189–198. https://doi.org/10.1016/j.infbeh.2016.07.003

Colonnesi, C., Rieffe, C., Koops, W., & Perucchini, P. (2008). Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology, 26*(4), 561–577. https://doi.org/10.1348/026151008x285660

Colonnesi, C., Stams, G., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Desrochers, S., Morissette, P., & Ricard, M. (1995). *Two perspectives on pointing in infancy* (pp. 85–101). Joint attention: Its origins and role in development.

Dimitrova, N., Özçalışkan, Ş., & Adamson, L. B. (2016). Parents' translations of child gesture facilitate word learning in children with autism, down syndrome and typical development. *Journal of Autism and Developmental Disorders, 46*(1), 221–231. https://doi.org/10.1007/s10803-015-2566-7

Dobrich, W., & Scarborough, H. S. (1984). Form and function in early communication: language and pointing gestures. *Journal of Experimental Child Psychology, 38*(3), 475–490.

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science, 23*(1). https://doi.org/10.1111/desc.12843

Dunn, L. M., Dunn, L. M., Bulheller, S., & Häcker, H. (1965). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Esseily, R., Jacquet, A.-Y., & Fagard, J. (2011). Handedness for grasping objects and pointing and the development of language in 14-month-old infants. *Laterality, 16*(5), 565–585.

Fasolo, M., & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture, 12*(1), 1–15. https://doi.org/10.1075/gest.12.1.01fas

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59(5),* 1–173; discussion 174.

* Goldin-Meadow, S., Butcher, C., 2003. Pointing toward two-word speech in young children. *Pointing:* Where language, culture, and cognition meet, 85, 107.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10*(6), 778–785. https://doi.org/10.1111/j.1467-7687.2007.00636.x

Hall, S., Rumney, L., Holler, J., & Kidd, E. (2013). Associations among play, gesture and early spoken language acquisition. *First Language, 33*(3), 294–312. https://doi.org/10.1177/0142723713487618

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

Edwards, S., Letts, C. & Sinka, I. (2011). The new Reynell Developmental Language Scales. GL Assessment.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. Monographs of the Society for Research in Child Development, Serial No. 242, Vol. 59, No. 5.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior & Development, 39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Iverson, J. M., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language, 28*(2), 164–181. https://doi.org/10.1177/0142723707087736

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Lawrence Erlbaum.

Keller, H., Otto, H., Lamm, B., Yovsi, R. D., & Kärtner, J. (2008). The timing of verbal/vocal communications between mothers and their infants: A longitudinal cross-cultural comparison. *Infant Behavior and Development, 31*(2), 217–226. https://doi.org/10.1016/j.infbeh.2007.10.001

Kirk, E., Warmington, M., Furman, R., Glanville, J., Eggleston, A., & Donnelly, S. (2016). A systematic review of the relationship between pointing and language development in monolingual and bilingual infants. PROSPERO 2016 CRD42016047246. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016047246.

Kishimoto, T., Shizawa, Y., Yasuda, J., Hinobayashi, T., & Minami, T. (2007). Do pointing gestures by infants provoke comments from adults? *Infant Behavior & Development, 30*(4), 562–567. https://doi.org/10.1016/j.infbeh.2007.04.001

Kuhn, L.J., Willoughby, M.T., Wilbourn, M.P., Vernon-Feagans, L., Blair, C.B., & Family Life Project, K. (2014). Early Communicative Gestures Prospectively Predict Language Development and Executive Function in Early Childhood. *Child Development, 85(5),* 1898–1914. Doi: 10.1111/cdev.12249.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science, 36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U. & Ruether, J. (2021). Ontogenetic origins of infant pointing. In: Nathalie Gontier, Andy Lock, and Chris Sinha (Eds.). The Oxford Handbook of Human Symbolic Evolution.

Lüke, C., Ritterfeld, U., Grimminger, A., Liszkowski, U., & Rohlfing, K. J. (2017). Development of pointing gestures in children with typical and delayed language acquisition. *Journal of Speech, Language, and Hearing Research, 60*(11), 3185–3197.

Lüke, C., Leinweber, J., & Ritterfeld, U. (2019). Walking, pointing, talking–the predictive value of early walking and pointing behavior for later language skills. *Journal of child language, 46*(6), 1228–1237.

Mastin, J. D., & Vogt, P. (2016a). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language, 43*(2), 235–264. https://doi.org/10.1017/S0305000915000148

Masur, E. F. (1982). Mothers' responses to infants' object-related gestures: influences on lexical development. *Journal of Child Language, 9*(1), 23–30.

Mastin, J. D., & Vogt, P. (2016b). Infant engagement and early vocabulary development: a naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language, 43*, 235–264.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development, 88*(1), 156–166. https://doi.org/10.1111/cdev.12671

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters. *IEEE Transactions on Autonomous Mental Development, 5*(3), 240–248. https://doi.org/10.1109/TAMD.2013.2275949

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS edition). Circle Pines, MN: American Guidance Service.

Mumford, K. H., & Kita, S. (2016). At 10–12 months, pointing gesture handedness predicts the size of receptive vocabularies. *Infancy, 21*(6), 751–765. https://doi.org/10.1111/infa.12138

Murillo, E., Galera, N., & Casla, M. (2015). Gesture and speech combinations beyond two-word stage in an experimental task. *Language Cognition and Neuroscience, 30*(10), 1291–1305. https://doi.org/10.1080/23273798.2015.1066509

Murphy, V., & Unthiah, A. (2015). A systematic review of intervention research examining English language and literacy development in children with English as an additional language (EAL).

Nicoladis, E. (2002). Some gestures develop in conjunction with spoken language development and others don't: Evidence from bilingual preschoolers. *Journal of Nonverbal Behavior, 26*(4), 241–266. https://doi.org/10.1023/a:1022112201348

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology, 35*(2), 514–526. https://doi.org/10.1037/0012-1649.35.2.514

Olson, J., & Masur, E. F. (2015). Mothers' labeling responses to infants' gestures predict vocabulary outcomes. *Journal of Child Language, 42*(6), 1289–1311. https://doi.org/10.1017/s0305000914000828

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*(3), B101–B113.

Özçalışkan, Ş., & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and cognitive processes, 24*(2), 190–217.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics, 24*(2), 221–234.

Rowe, M. L. (2000). Pointing and talk by low-income mothers and their 14-month-old children. *First Language, 20*(60), 305–330. https://doi.org/10.1177/014272370002006005

Rowe, M. L., & Goldin-Meadow, S. (2009a). Early gesture selectively predicts later language learning. *Developmental Science, 12*(1), 182–187.

Rowe, M. L., & Goldin-Meadow, S. (2009b). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*(5916), 951–953.

Rowe, M. L., Özçalişkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Salo, V. C., Rowe, M. L., & Reeb-Sutherland, B. C. (2018). Exploring infant gesture and joint attention as related constructs and as predictors of later language. *Infancy, 23*(3), 432–452. https://doi.org/10.1111/infa.12229

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Development, 84*(4), 1296–1307. https://doi.org/10.1111/cdev.12026

Schieffelin, B. B., & Ochs, E. (1986). Language Socialization. *Annual Review of Anthropology, 15*(1), 163–191. https://doi.org/10.1146/annurev.an.15.100186.001115

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science, 15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language Typology and Syntactic Description, 3*, 57–149. https://digitalassets.lib.berkeley.edu/techreports/ucb/text/COGSCI-87-47.ps.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748–767.

Tamis-LeMonda, C. S., Song, L. L., Leavell, A. S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother-infant language and gestural communications are associated with specific skills in infants. *Developmental Science, 15*(3), 384–397.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*(3), 705–722.

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development, 57*(6), 1454. https://doi.org/10.2307/1130423

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*.

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research, 49*(2), 237–247. https://doi.org/10.1044/1092-4388(2006/020)

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72–90. https://doi.org/10.1177/0142723714521925

Zambrana, I. M., Ystrom, E., Schjølberg, S., & Pons, F. (2013). Action imitation at 11/2 years is better than pointing gesture in predicting late development of language production at 3 years of age. *Child Development, 84*(2), 560–573. https://doi.org/10.1111/j.1467-8624.2012.01872.x

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale* (4th ed.). San Antonio, TX: The Psychological Corporation.