**Supplementary Materials**

**Single-study analysis using a selection hierarchy**

We replicated the analysis in Anderson et al. (2021) using a single reported effect size per study. We devised a hierarchy for determining which measures would be included in our final analyses similar in many respects to the Anderson meta-analysis (Table S1). The goal of the hierarchy was to create a single correlation statistic from each study that could be included in each of our analyses. First, we considered when language input and outcomes were collected. We preferred longitudinal correlations (i.e., input collected at T1, outcome collected at T2) over cross-sectional correlations (i.e., input and outcome collected at T1). If there were multiple longitudinal correlations, we selected the one with the longest time between T1 and T2, preferring correlations from older ages over younger ages. Second, we considered how language outcomes were assessed. We preferred direct assessments over parent reports, and parent reports over observation. We also prioritized expressive language assessments over receptive language assessments when both were available. Third, we considered how input was coded. We preferred measures that were more commonly used over measures that were less commonly used. For example, we preferred correlations with word types over those with type-token ratio or VOCD. Finally, we considered additional factors as were necessary to select a single measure (e.g., correlations with input from mother over input from father, child-directed speech over speech to others, etc.).

**Table S1:** Study selection hierarchy

|  |  |  |
| --- | --- | --- |
| Condition | Selection | Rationale |
| If there are multiple timepoints | Longitudinal > Cross-Sectional  (+ Longest time between)  Older age > Younger age | Longitudinal correlations more likely to be causal  Assessments of older children less noisy; Older children have more overall speech exposure |
| If there are multiple language outcome measures | Direct Assessment > Parent Report > Observed (Types > MLU > Tokens)  Expressive > Receptive | Prefer measures less likely to be affected by caregivers; Prefer more common observational measures  Expressive measures less noisy for younger children |
| If there are multiple language input measures | Types > TTR > VOCD > other  MLU (morphemes) > MLU (words) | Prefer more commonly used input measures in each case |
| If there are other factors | Child-directed speech > Overheard  Longer duration > Shorter duration  Mother > Father > Adults > Other  Naturalistic > Free Play > Structured Activity > Other | Child-directed speech is the focus of our analysis  Longer sessions better represent everyday speech  Prefer speech from primary caregivers  Natural conditions better represent everyday speech |

**Word Tokens**

We examined 41 studies that measured word tokens (n=2007 participants). We found a medium sized effect across studies (*r*=0.23, *p*<0.001; CI [0.17; 0.29]) suggesting that word tokens reliably predict language outcomes (Figure S1). *Q*-statistics revealed marginally significant evidence for between-study heterogeneity (*Q*(40) = 58.98, *p* = 0.03), which motivated an analysis of possible moderators.However, we did not find any evidence that differences in subject characteristics, language assessments, or study design features moderated the effect of word tokens on children’s outcomes (Table S2).

To assess publication bias, we compared three groups of studies: effect sizes from studies that were not peer-reviewed such as dissertations and book chapters, effects from peer-reviewed studies where data was obtained by contacting authors which we defined as “unreported statistics”, and a baseline group of effects from peer-reviewed studies which compose the remainder of our meta-analysis sample. To check for moderating effects of publication type, we compared the non-peer-reviewed and studies with unreported statistics to our baseline. Neither of these variables were found to moderate the effect of word tokens on outcomes. Additionally, we did not find any evidence of asymmetry in our funnel plot using Egger’s test (*t*(39) = 0.90, *p*=0.37) (Figure S2). In sum, there was no evidence of publication bias in the word token studies.

**Figure S1:** Forest plot of token study correlations

A table of statistics with numbers and text

Description automatically generated with medium confidence

**Table S2:** Results from the analysis of variables moderating the relationship between parent word tokens and language outcomes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Moderator | k | *β* | CI95 | zval | pval | R2 |
| Subject Characteristics |  |  |  |  |  |  |
| Child Gender (% Female) | 40 | 0.00 | [-0.08, 0.08] | -0.01 | 0.99 | 0.00 |
| Household SES | 41 |  |  |  |  | 0.00 |
| Diverse vs. MU |  | 0.07 | [-0.08, 0.23] | 0.97 | 0.34 |  |
| Low vs. MU |  | -0.02 | [-0.18, 0.13] | -0.31 | 0.76 |  |
| Language | 41 |  |  |  |  | 0.00 |
| English vs. Non-English |  | -0.08 | [-0.25, 0.08] | -1.01 | 0.32 |  |
| Region | 41 |  |  |  |  | 0.00 |
| US vs. Non-U.S. |  | 0.008 | [-0.16, 0.17] | 0.1 | 0.92 |  |
| Child Age (Input) | 40 | 0.02 | [-0.05, 0.08] | 0.48 | 0.64 | 0.00 |
| Child Age (Assessment) | 40 | 0.05 | [-0.02, 0.11] | 1.34 | 0.19 | 11.66 |
| Assessment Characteristics |  |  |  |  |  |  |
| Assessment Type | 41 |  |  |  |  | 0.28 |
| Report vs. Direct |  | -0.11 | [-0.25, 0.03] | -1.64 | 0.11 |  |
| Observed vs. Direct |  | -0.12 | [-0.28, 0.04] | -1.53 | 0.13 |  |
| Assessment Measure | 41 |  |  |  |  | 0.03 |
| Receptive vs. Expressive |  | 0.14 | [-0.03, 0.32] | 1.63 | 0.11 |  |
| Both vs. Expressive |  | 0.02 | [-0.19, 0.23] | 0.2 | 0.84 |  |
| Vocabulary Measure | 41 |  |  |  |  | 0.00 |
| Vocab vs. Non-Vocab |  | -0.04 | [-0.18, 0.1] | -0.62 | 0.54 |  |
| Study Design |  |  |  |  |  |  |
| Observation Duration | 39 | -0.001 | [-0.06, 0.06] | -0.05 | 0.96 | 0.00 |
| Observation Context | 41 |  |  |  |  | 0.00 |
| Nat. Play vs. Naturalistic |  | -0.1 | [-0.31, 0.11] | -1.01 | 0.32 |  |
| Struc. Play vs. Naturalistic |  | -0.07 | [-0.22, 0.07] | -0.98 | 0.33 |  |
| Other vs. Naturalistic |  | -0.1 | [-0.4, 0.2] | -0.69 | 0.5 |  |
| Observation Location | 41 |  |  |  |  | 0.00 |
| Lab vs. Home |  | 0.01 | [-0.15, 0.17] | 0.09 | 0.93 |  |
| Other vs. Home |  | -0.14 | [-0.54, 0.27] | -0.69 | 0.5 |  |
| Study Timeline | 41 |  |  |  |  | 0.31 |
| Cross-Lagged vs. Concurrent |  | -0.1 | [-0.23, 0.02] | -1.67 | 0.1 |  |
| Publication Bias |  |  |  |  |  |  |
| Peer-Review Status | 41 |  |  |  |  | 0.02 |
| Peer-Reviewed vs. Non-PR |  | 0.07 | [-0.1, 0.23] | 0.84 | 0.41 |  |
| Reported in PR Pubs. | 41 |  |  |  |  | 0.00 |
| Reported vs. Not Reported |  | 0.02 | [-0.1, 0.15] | 0.39 | 0.7 |  |

**Figure S2:** Funnel plot of token study correlations

**A graph of a funnel plot

Description automatically generated**

**Utterances**

*Summary statistics*

We examined 19 studies that measured the number of utterances (n=968 participants). We found a modest effect across studies (*r*=0.19, *p*<0.001; CI [0.08; 0.29]) (Figure S3). *Q*-statistics revealed no significant evidence of between-study heterogeneity (*Q*(18) = 27.64, *p*=0.06). Therefore, we did not conduct an analysis of moderating variables. We found no effect of peer-review or whether the correlation coefficient was reported on the size of the pooled correlation. Additionally, we did not find any evidence of asymmetry in our funnel plot using Egger’s test (*t*(17) = 0.52, *p*=0.61) (Figure S4).

**Figure S3:** Forest plot of utterance study correlations

A screenshot of a graph

Description automatically generated

**Figure S4:** Funnel plot of utterance study effect sizes

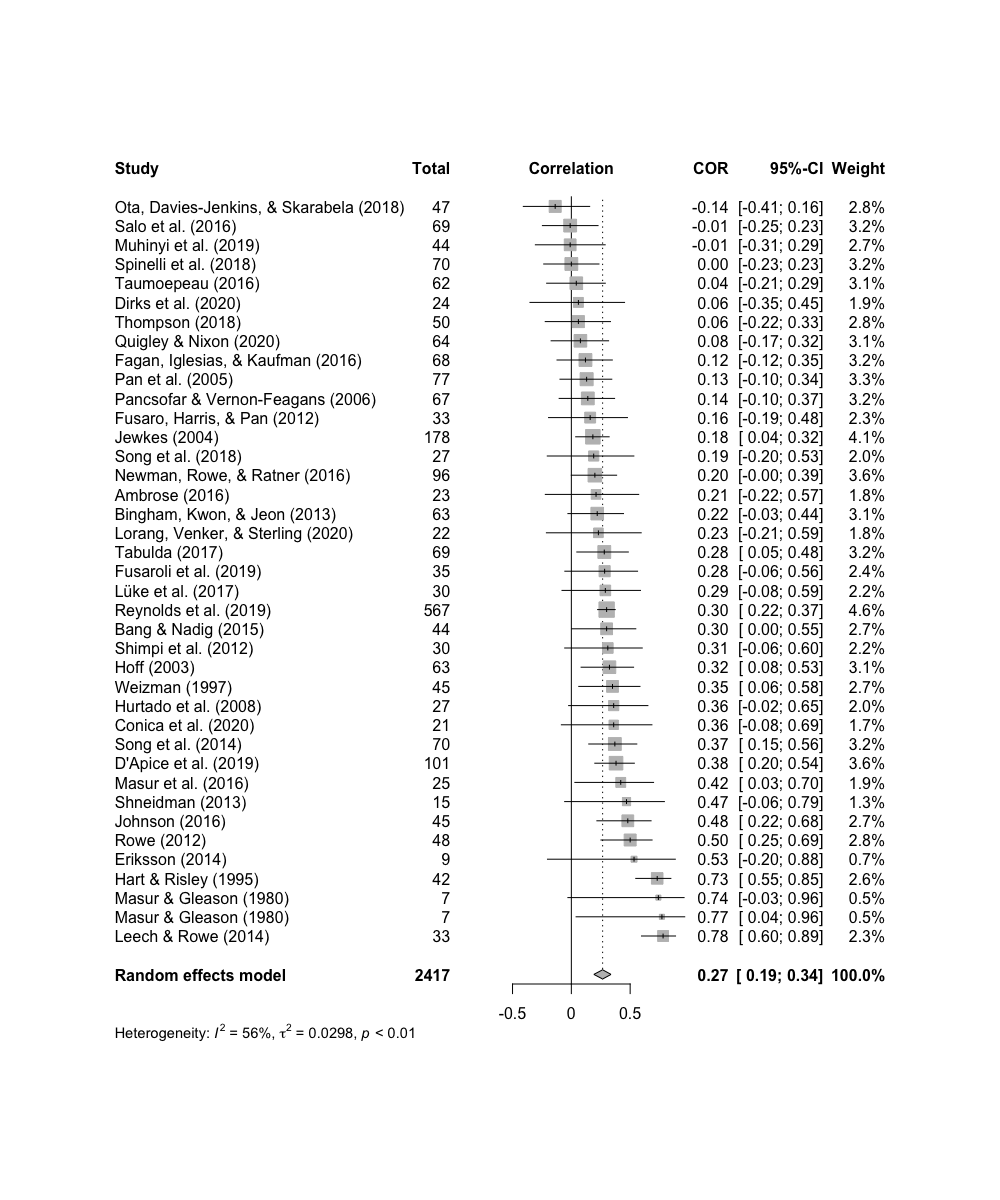
**A graph of a funnel plot

Description automatically generated**

**Word Types**

We examined 39 studies that measured word types (n=2417 participants). As with our word tokens, we found a medium-sized effect across studies (*r*=0.27, *p*<0.001; CI [0.19; 0.34]) (Figure S5). *Q*-statistics revealed significant evidence for between-study heterogeneity (*Q*(38) = 87.23, *p*<0.001).We found evidence for moderation in our analysis of outcome measures in our subject and study characteristic variables. Specifically, the age of participants moderated effect sizes, such that studies with children who were older at the time of input data collection had larger effects (*β*=0.005, *SE*=0.002, *p*=0.03) and studies with children who were older at the time of language outcome data collection also had marginally larger effect sizes (*β*=0.005, *SE*=0.002, *p*=0.06). Unsurprisingly, we found that these two variables were highly correlated with one another (*r*=0.87, *p*<0.001), and thus it is unclear if it is the age at input, outcome or both that predicts the bigger effect size. We also found significantly higher effect sizes in studies where input and outcome measures were matched (i.e., parent word types and child types produced during observation) (*β*=0.17, *SE*=0.08, *p*=0.04), suggesting that characterizing the input and the outcome in the same way resulted in higher correlations (Table S3). We revisit this finding in our Discussion section.In our analysis of publication bias, neither peer review nor whether the correlation was reported was found to moderate the effect of word types on outcomes. Additionally, we did not find any evidence of asymmetry in our funnel plot using Egger’s test (*t*(37) = 0.87, *p*=0.39) (Figure S6).

**Figure S5:** Forest plot of type study correlations

****

**Table S3:** Results from the analysis of variables moderating the relationship between parent word types and language outcomes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Moderator | k | *β* | CI95 | zval | pval | R2 |
| Subject Characteristics |  |  |  |  |  |  |
| Child Gender (% Female) | 37 | 0.03 | [-0.05, 0.11] | 0.74 | 0.47 | 0.00 |
| Household SES | 39 |  |  |  |  | 0.00 |
| Diverse vs. MU |  | 0.05 | [-0.18, 0.27] | 0.42 | 0.67 |  |
| Low vs. MU |  | -0.05 | [-0.26, 0.16] | -0.48 | 0.63 |  |
| Language | 39 |  |  |  |  | 0.00 |
| English vs. Non-English |  | 0.02 | [-0.2, 0.24] | 0.21 | 0.83 |  |
| Region | 39 |  |  |  |  | 0.04 |
| US vs. Non-U.S. |  | 0.141 | [-0.05, 0.33] | 1.47 | 0.149 |  |
| Child Age (Input) | 39 | 0.08 | [0, 0.01] | 2.23 | 0.03\* | 0.1 |
| Child Age (Assessment) | 38 | 0.08 | [0, 0.01] | 1.96 | 0.04\* | 0.12 |
| Assessment Characteristics |  |  |  |  |  |  |
| Assessment Type | 39 |  |  |  |  | 0.00 |
| Report vs. Direct |  | -0.08 | [-0.28, 0.11] | -0.86 | 0.4 |  |
| Observed vs. Direct |  | 0.09 | [-0.09, 0.27] | 1.03 | 0.31 |  |
| Assessment Measure | 39 |  |  |  |  | 0.00 |
| Receptive vs. Expressive |  | 0.05 | [-0.16, 0.26] | 0.49 | 0.63 |  |
| Both vs. Expressive |  | -0.05 | [-0.39, 0.28] | -0.32 | 0.75 |  |
| Vocabulary Measure | 39 |  |  |  |  | 0.03 |
| Vocab vs. Non-Vocab |  | 0.13 | [-0.06, 0.32] | 1.35 | 0.18 |  |
| Matched Input/Outcome | 39 |  |  |  |  | 0.1 |
| Matched vs. Non-Matched |  | 0.17 | [0.01, 0.33] | 2.16 | 0.04\* |  |
| Study Design |  |  |  |  |  |  |
| Observation Duration | 38 | 0.07 | [-0.01, 0.15] | 1.73 | 0.09’ | 0.00 |
| Observation Context | 39 |  |  |  |  | 0.00 |
| Nat. Play vs. Naturalistic |  | -0.1 | [-0.4, 0.21] | -0.65 | 0.52 |  |
| Struc. Play vs. Naturalistic |  | -0.08 | [-0.28, 0.12] | -0.8 | 0.43 |  |
| Other vs. Naturalistic |  | -0.02 | [-0.28, 0.24] | -0.16 | 0.87 |  |
| Observation Location | 39 |  |  |  |  | 0.00 |
| Lab vs. Home |  | -0.07 | [-0.24, 0.09] | -0.89 | 0.38 |  |
| Other vs. Home |  | 0.01 | [-0.49, 0.51] | 0.04 | 0.97 |  |
| Study Timeline | 39 |  |  |  |  | 0.00 |
| Cross-Lagged vs. Concurrent |  | 0.00 | [-0.16, 0.16] | -0.03 | 0.98 |  |
| Publication Bias |  |  |  |  |  |  |
| Peer-Review Status | 39 |  |  |  |  | 0.02 |
| Peer-Reviewed vs. Non-PR |  | -0.13 | [-0.33, 0.07] | -1.3 | 0.2 |  |
| Reported in PR Pubs. | 39 |  |  |  |  | 0.00 |
| Reported vs. Not Reported |  | -0.02 | [-0.18, 0.15] | -0.2 | 0.84 |  |

**Figure S6:** Funnel plot of type study correlations

**A graph of funnel plot

Description automatically generated**

**MLU**

We examined 29 studies that measured MLU (n=2300 participants). We found a medium-sized effect of MLU on language outcomes overall (*r*=0.21, *p*<0.001; CI [0.13; 0.30]) (Figure S7). Q-statistics revealed significant evidence for between-study heterogeneity (*Q*(28) = 68.39, *p*<0.001).We found a significant and positive correlation between the length of the observation session and effect size. (*β*=0.01, *SE*=0.003, *p*=0.001), with longer studies producing larger effect sizes. We might expect to see such an effect if MLU measures were more stable when the sample of utterances is larger. Neither of these variables were found to moderate the effect of MLU on outcomes (Table S4). Additionally, we did not find any evidence of asymmetry in our funnel plot using Egger’s test (*t*(27)=0.82, *p*=0.42) (Figure S8).

**Figure S7:** Forest plot of MLU study correlations

A graph of a number of people

Description automatically generated with medium confidence

**Table S4:** Results from the analysis of variables moderating the relationship between parent MLU and language outcomes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Moderator | k | *β* | CI95 | zval | pval | R2 |
| Subject Characteristics |  |  |  |  |  |  |
| Child Gender (% Female) | 27 | -0.01 | [-0.11, 0.10] | -0.17 | 0.87 | 0.00 |
| Household SES | 29 |  |  |  |  | 0.00 |
| Diverse vs. MU |  | 0.00 | [-0.21, 0.21] | 0 | 1 |  |
| Low vs. MU |  | -0.08 | [-0.38, 0.23] | -0.5 | 0.62 |  |
| Language | 29 |  |  |  |  | 0.00 |
| English vs. Non-English |  | 0.07 | [-0.17, 0.31] | 0.58 | 0.56 |  |
| Region | 29 |  |  |  |  | 0.19 |
| US vs. Non-U.S. |  | 0.167 | [-0.03, 0.37] | 1.7 | 0.1 |  |
| Child Age (Input) | 29 | 0.009 | [-0.08, 0.10] | -0.24 | 0.81 | 0.00 |
| Child Age (Assessment) | 29 | -0.02 | [-0.11, 0.07] | -0.45 | 0.65 | 0.00 |
| Assessment Characteristics |  |  |  |  |  |  |
| Assessment Type | 29 |  |  |  |  | 0.00 |
| Report vs. Direct |  | -0.03 | [-0.3, 0.24] | -0.24 | 0.81 |  |
| Observed vs. Direct |  | 0.02 | [-0.2, 0.23] | 0.15 | 0.88 |  |
| Assessment Measure | 29 |  |  |  |  | 0.00 |
| Receptive vs. Expressive |  | -0.14 | [-0.38, 0.11] | -1.17 | 0.25 |  |
| Both vs. Expressive |  | -0.01 | [-0.25, 0.23] | -0.06 | 0.95 |  |
| Vocabulary Measure | 29 |  |  |  |  | 0.00 |
| Vocab vs. Non-Vocab |  | -0.06 | [-0.25, 0.13] | -0.66 | 0.51 |  |
| Matched Input/Outcome | 29 |  |  |  |  |  |
| Matched vs. Non-Matched |  | -0.21 | [-0.55, 0.12] | -1.3 | 0.21 |  |
| Study Design |  |  |  |  |  |  |
| Observation Duration | 27 | 0.1 | [0.02, 0.17] | 2.81 | 0.01\*\* | 0.38 |
| Observation Context | 29 |  |  |  |  | 0.00 |
| Nat. Play vs. Naturalistic |  | -0.17 | [-0.52, 0.17] | -1.02 | 0.32 |  |
| Struc. Play vs. Naturalistic |  | -0.14 | [-0.44, 0.16] | -0.94 | 0.35 |  |
| Other vs. Naturalistic |  | -0.26 | [-0.59, 0.07] | -1.61 | 0.12 |  |
| Observation Location | 29 |  |  |  |  | 0.13 |
| Lab vs. Home |  | -0.15 | [-0.34, 0.03] | -1.7 | 0.10 |  |
| Other vs. Home |  | 0.17 | [-0.3, 0.65] | 0.75 | 0.46 |  |
| Study Timeline | 29 |  |  |  |  | 0.00 |
| Cross-Lagged vs. Concurrent |  | 0.07 | [-0.12, 0.26] | 0.78 | 0.44 |  |
| Publication Bias |  |  |  |  |  |  |
| Peer-Review Status | 29 |  |  |  |  | 0.00 |
| Peer-Reviewed vs. Non-PR |  | -0.04 | [-0.28, 0.2] | -0.34 | 0.74 |  |
| Reported in PR Pubs. | 29 |  |  |  |  | 0.003 |
| Reported vs. Not Reported |  | -0.08 | [-0.27, 0.12] | -0.8 | 0.43 |  |

**Figure S8:** Funnel plot of MLU study correlations

**A graph of a funnel plot

Description automatically generated**

**Comparison of input measures**

We also directly compared the effect sizes of different inputs using the method in Anderson et al. (2021), where we checked the confidence intervals for each of our input coefficient estimates for overlap. We opted to use an 85% confidence interval for comparison as is common practice in other meta-analyses (e.g., Anderson et al., 2021). Using this method on our estimates for tokens (*r*=0.23; 85% CI = [0.19; 0.27]), utterances (*r*=0.19; 85% CI [0.11; 0.26]), types (*r*=0.27; 85% CI = [0.21; 0.32]), and MLU (*r*=0.22; 85% CI = [0.15; 0.27]), we did not find any difference in pooled effect size between any of the four input measures.

**Using study duration as selection criteria**

A possible concern with our analysis comes from our inclusion of studies that have low observation durations. Studies that draw from brief observations introduce a substantial amount of noise in their estimates of input, which risks introducing unusually large effect sizes into our studies. In addition, these noisy estimates cannot be detected by examining funnel plot asymmetry, where standard errors are only calculated on the basis of sample size: a large sample where estimates are drawn from brief observations would not appear noisy in a traditional analysis of bias. As a result, we set a minimum duration threshold at 5 minutes, which is commonly used to briefly sample from structured free play interactions. Input samples collected from structured free play activities of this duration have been found to correlate with input collected from longer naturalistic observations (Tamis-LeMonda et al., 2017). This resulted in the total number of studies being reduced by five (*N=*66) and the total number of effect sizes by 39 (*k*=289). We then conducted our three analyses (pooled effect size, moderators, and publication bias) on each of the four input measures.

We find roughly similar effect sizes across our input measures, with no significant differences between them. We find no difference in our analysis of moderators for studies of word tokens (*r*=0.23, *p*<0.001; CI [0.18; 0.29]) or word types (*r*=0.29, *p*<0.001; CI [0.22; 0.36]). Likewise, we find no difference in our analysis of publication bias: only word types showed evidence of asymmetry. As an additional check, we also removed studies that used LENA outputs as input variables, as these estimates are produced at a much larger timescale (8 hours+) than any other kind of study in our sample. Because these studies use an algorithmic estimation of word counts, these studies are all counted as word token studies, which brings the total token studies to *N=27* and the total effect sizes to *k*=70. We did not find any difference in the calculated pooled effect size (*r*=0.23, *p*<0.001; CI [0.10; 0.35]). We found that studies of receptive vocabulary produced significantly larger correlations than studies of expressive vocabulary (*β*=0.14, *SE*=0.05, *p*<0.05), but inclusion of this variable did not improve overall model fit (*χ*2(3)=5.47, *p*=0.06).

However, we did find some differences between our original analysis and the current analysis for MLU (*r*=0.24, *p*<0.001; CI [0.16; 0.32]) and utterances (*r*=0.23, *p*<0.001; CI [0.10; 0.35]). Namely, for MLU, in addition to a significant effect of study duration (*β*=0.94, *SE*=0.3, *p*=0.02), we now also find a significant effect of observation activity (*χ*2(3)=12.83, *p*<0.01), such that naturalistic studies produce significantly larger input-outcome correlations than studies using naturalistic free play (*β*=0.48, *SE*=0.11, *p*<0.05), structured free play (*β*=0.45, *SE*=0.08, *p*<0.05), and other activities (*β*=0.53, *SE*=0.08, *p*<0.05). This moderator was also found to be significant in Anderson et al. (2021), albeit with studies of input quantity (i.e., tokens and utterances). But in our study, we believe that this effect is related to duration. Naturalistic studies were found to have a longer duration than other studies in our sample, even after removing longer LENA studies (none of which were found in our MLU sample). As a result, the two variables are confounded, and adding observation activity to our duration-only model fails to improve fit (*χ*2(3)=3.64, *p*=0.30). Thus, we believe this effect is driven by the effect of observation duration found previously.

For utterances, we found an effect of child gender on correlations (*χ*2(3)=5.35, *p*<0.05)., such that studies with more female subjects produced significantly larger effects sizes than studies with more male subjects (*β*=0.16, *SE*=0.03, *p*<0.05). There is no prediction we would make a priori that would have resulted in larger correlations for female children. Given that there are no other significant moderating variables, it is also unclear whether this effect results from some other feature of these studies.

Overall, we find few differences between the results of this analysis and our original analysis. Omitting studies with durations that were unusually small (< 5 minutes) as well as those that were unusually large (> 8 hours) did not substantially alter our findings with regard to pooled effect size or publication bias. For MLU, it resulted in an additional moderating effect of observation activity, one which was confounded with the effect of study duration. For utterances, this resulted in a moderating effect of child gender on input-outcome correlations.

**Table S1:** Boolean search query

|  |  |
| --- | --- |
| **Criteria** | **Search Terms** |
| Parent | (parent\* OR mother\* OR caregiver OR father) AND … |
| Child | (child\* OR infant OR toddler) AND … |
| Input | (input OR interaction OR conversation OR "directed speech") AND… |
| Outcome | ("language development" OR "word learning" OR vocabulary OR "language acquisition") |

**Table S2:** Databases queried

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Database** | **Search By** | **Total** | **Duplicates Removed** | **Date Searched** | **Abstracts Approved** |
| EBSCO (ERIC, PsycInfo, Academic Search Premier) | Title, Abstract, Subject Headers | 3632 | 2623 | 01-14-2021 |  |
| PubMed | Title, Abstract | 460 | 130 | 01-14-2021 |  |
| Web of Science | Topic (Title, Abstract, Keywords) | 1824 | 881 | 01-14-2021 |  |
| Proquest | All but Full Text | 847 | 564 | 01-14-2021 |  |
|  |  | 6763 | 4198 |  | 188 |

**Table S3:** Study inclusion criteria

|  |  |
| --- | --- |
| **Question** | How do the linguistic properties of caregiver speech affect language development? |
| **Criteria** | **Description** |
| 1. What documents do we include? | **Included:** English language Journal articles, book chapters, dissertations, conference proceedings  **Excluded:** Reviews, meta-analyses, non-English language  **Notes:** Both published and unpublished acceptable. Interventions with pre-test/control measures acceptable. |
| 2. What participants are we considering? | **Included:** Typically developing, monolingual children between the ages of 1 and 8 years  **Excluded:** Atypically developing children (i.e., ASD, SLI), children with disabilities (i.e., deaf or hard-of-hearing), preterm infants, bilingual/multilingual children  **Notes:** Studies with typically developing monolingual control groups acceptable |
| 3. What kinds of input are we considering? | **Included:** Quality (word types, lexical diversity, MLU) and quantity (word tokens, words per minute, utterances) of speech directed to children by caregivers in naturalistic/semi-naturalistic settings (i.e., home or structured free play)  **Excluded:** Other linguistic properties (i.e., questions, decontextualized speech), interactive characteristics (i.e., warmth, responsiveness), or observer judgements (i.e., sensitivity ratings); scripted speech (i.e., only words read to children) |
| 4. What kinds of language outcomes are we considering? | **Included:** Measures of child vocabulary (i.e., CDI, PPVT) or equivalent (i.e., receptive/expressive language); samples of observed child word usage (i.e., word types produced)  **Excluded:** Other measures of language use (i.e., syntax, pragmatics, use of certain lexical forms, novel word learning) |

**Table S4:** Input and moderating variables

|  |  |  |
| --- | --- | --- |
| **Measure** | **Description** | **Categories** |
| **Input measures** |  |  |
| Word tokens | Total number of words produced |  |
| Utterances | Total continuous segments of speech |  |
| Word types | Total number of different words produced |  |
| Mean-length of utterance | Average number of words in utterances |  |
| **Outcome measures** |  |  |
| Assessment type | Type of assessment used to measure children’s language | 1. Direct assessment 2. Observation 3. Parent report |
| Assessment measure | Measure of language that was assessed in children | 1. Expressive 2. Receptive |
| Vocabulary | Was the assessment a measure of vocabulary (e.g., CDI, PPVT)? | 1. Vocabulary 2. Not vocabulary |
| Matching | Does the input capture the same construct as the outcome? | 1. Matched 2. Not matched |
| **Subject characteristics** |  |  |
| Gender | Proportion of female participants |  |
| Age | Age of child at the time input/outcome was collected |  |
| Sources of input | Speakers included in the input measure | 1. Mother 2. Father 3. Primary caregiver 4. Any adult 5. Other |
| Native language | Was the observation/assessment conducted in English? | 1. English 2. Not English |
| Region | Did the observation/assessment take place in the US? | 1. US 2. Not US |
| Household SES | Reported socioeconomic status of families in study | 1. Middle-upper 2. Diverse 3. Low |
| **Study characteristics** |  |  |
| Duration | Length of study observation |  |
| Location | Location of observation | 1. Home 2. Lab 3. Other |
| Activity | Activity caregivers were asked to engage in during observation | 1. Naturalistic 2. Natural play 3. Structured play 4. Other |
| Study timeline | Was input/outcome collection lagged or concurrent? | 1. Lagged 2. Concurrent |
| Publication type | Was the study peer-reviewed? | 1. Peer-reviewed 2. Non-PR |
| Reported | Were the stats reported in a peer-reviewed publication? | 1. Reported 2. Not reported |

**Table S6:** Description of participants in each study for all analyses of input

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | Language | Country | SES | Input |
| Adams et al. (2018) | English | US | MU | Token |
| Ambrose (2016) | English | US | MU | Type |
| Bang & Nadig (2015) | French | Canada | MU | Token, Utt, Type, MLU |
| Bingham, Kwon, & Jeon (2013) | English | US | MU | Type, MLU |
| Conica et al. (2020) | English | US | MU | Type |
| D’Apice et al. (2019) | English | UK | MU | Token, Type |
| Dave et al. (2018) | English | US | MU | Token, Utt |
| Dunham & Dunham (1992) | English | US | MU | Utt |
| Dwyer (2017) | English | Australia | D | Token |
| Eriksson (2014) | Swedish | Sweden | MU | Token, Utt, Type, MLU |
| Fagan, Iglesias, & Kaufman (2016) | English | US | L | Type, MLU |
| Fusaro, Harris, & Pan (2012) | English | US | MU | Token, Type |
| Fusaroli et al. (2019) | English | US | MU | Token, Type, MLU |
| Gilkerson et al. (2017) | English | US | D | Token |
| Hardy-Brown & Plomin (1985) | English | US | MU | Utt, MLU |
| Hart & Risley (1995) | English | US | D | Type |
| Himes (1978) | English | US | MU | Utt, MLU |
| Hirsch-Pasek et al. (2015) | English | US | L | Token |
| Hoff (2003) | English | US | D | Token, Utt, Type, MLU |
| Hoff-Ginsberg (1986) | English | US | D | MLU |
| Hurtado et al. (2008) | Spanish | US | L | Token, Utt, Type, MLU |
| Jewkes (2004) | English | US | MU | Type, MLU |
| Johnson (2016) | English | US | MU | Type |
| Kauffman (1985) | English | US | MU | Utt |
| Kyger (2013) | English | US | MU | MLU |
| Larson, Barrett, & McConnell (2019) | English | US | L | Token |
| Lecheile et al. (2020) | English | US | D | MLU |
| Leech & Rowe (2014) | English | US | MU | Type |
| Liu (2014) | Chinese | Taiwan | MU | Token |
| Lopez, Walle, Pretzer, & Warlaumont (2020) | English | US | MU | Utt |
| Lorang, Venker, & Sterling (2020) | English | US | MU | Type, MLU |
| Lu¨ke et al. (2017) | German | Germany | MU | Token, Utt, Type, MLU |
| Mahr & Edwards (2018) | English | US | MU | Token |
| Masur & Gleason (1980) | English | US | MU | Type, Type |
| Masur et al. (2016) | English | US | MU | Token, Type |
| Mimeau et al. (2020) | English | Canada | D | Utt |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | Language | Country | SES | Input |
| Moerk (1975) | English | US | MU | MLU |
| Muhinyi et al. (2019) | English | US | D | Token, Type, MLU |
| Newman, Rowe, & Ratner (2016) | English | US | MU | Token, Type |
| Ota, Davies-Jenkins, & Skarabela (2018) | English | UK | MU | Type |
| Paavola-Ruotsalainen et al. (2018) | Finnish | Finland | MU | Utt |
| Pan et al. (2005) | English | US | L | Token, Type |
| Pancsofar & Vernon-Feagans (2006) | English | US | MU | Type, MLU |
| Perlmann (1984) | English | US | MU | Token |
| Quigley & Nixon (2020) | English | US | MU | Type |
| Ramirez-Esparza et al. (2014) | English | US | MU | Token |
| Reynolds et al. (2019) | English | US | MU | Type, MLU |
| Romeo et al. (2018) | English | US | D | Token |
| Rowe (2012) | English | US | MU | Token, Type |
| Rufsvold et al. (2018) | English | US | MU | Token |
| Salo et al. (2016) | English | US | L | Token, Type, MLU |
| Scarborough & Wyckoff (1986) | English | US | MU | MLU |
| Sheran (1999) | English | US | D | Token, MLU |
| Shimpi & Huttenlocher (2007) | English | US | MU | Token |
| Shimpi et al. (2012) | English | US | L | Type |
| Shneidman (2013) | English | US | D | Token, Type |
| Song et al. (2014) | Spanish | US | L | Token, Type |
| Song et al. (2018) | English | US | MU | Token, Type, MLU |
| Spinelli et al. (2018) | Italian | Italy | MU | Type, MLU |
| Stafford (1987) | English | US | MU | Utt |
| Sultana, Wong, & Purdy (2020) | English | New Zealand | D | Token |
| Tabulda (2017) | English | US | D | Token, Type, MLU |
| Tardif (1993) | Chinese | Beijing | D | Token |
| Taumoepeau (2016) | English | New Zealand | D | Utt, Type, MLU |
| Thompson (2018) | English | US | MU | Token, Type |
| Tomasello, Mannle, & Kruger (1986) | English | US | MU | Utt |
| Wade et al. (2018) | English | Canada | D | MLU |
| Walle & Campos (2014) | English | US | MU | Token, Utt |
| Weisleder & Fernald (2013) | Spanish | US | L | Token |
| Weizman (1997) | English | US | L | Token, Type |

|  |
| --- |
| **References**  Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., & Feldman, H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes in full term and preterm toddlers. *Child development, 89(5),* 1674-1690. |
| Ambrose, S. E. (2016). Gesture use in 14-month-old toddlers with hearing loss and their mothers' responses. *American journal of speech-language pathology*, *25(4),* 519-531. |
| Bang, J., & Nadig, A. (2015). Learning language in autism: Maternal linguistic input contributes to later vocabulary. *Autism Research, 8(2),* 214-223. |
| Bingham, G. E., Kwon, K. A., & Jeon, H. J. (2013). Examining relations among mothers', fathers', and children's language use in a dyadic and triadic context. *Early Child Development and Care, 183*, 394-414. |
| Clarke-Stewart, K. A., VanderStoep, L. P., & Killian, G. A. (1979). Analysis and replication of mother-child relations at two years of age. *Child Development,* 777-793. |
| Conica, M., Nixon, E., & Quigley, J. (2020). Fathers' but not Mothers' Repetition of Children's Utterances at Age Two is Associated with Child Vocabulary at Age Four. *Journal of experimental child psychology, 191*, 104738. |
| Cox Eriksson, C. (2014). Children's vocabulary development: the role of parental input, vocabulary composition and early communicative skills (Doctoral dissertation, Department of Special Education, Stockholm University). |
| d'Apice, K., Latham, R. M., & von Stumm, S. (2019). A Naturalistic Home Observational Approach to Children's Language, Cognition, and Behavior. *Developmental Psychology*, *55*(7), 1414–1427. |
| Dave, S., Mastergeorge, A.M., & Olswang, L.B. (2018). Motherese, affect, and vocabulary development: Dyadic communicative interactions in infants and toddlers. *Journal of Child Language, 45(4),* 917-938. |
| Dirks, E., Stevens, A., Sigrid, K. O. K., Frijns, J., & Rieffe, C. (2020). Talk with me! Parental linguistic input to toddlers with moderate hearing loss. *Journal of Child Language, 47(1),* 186-204. |
| Dunham, P., & Dunham, F. (1992). Lexical development during middle infancy: A mutually driven infant-caregiver process. *Developmental Psychology, 28(3),* 414. |
| Dwyer, A. R. (2017). Early language experience and later vocabulary among Australian infants from diverse socioeconomic backgrounds (Order No. 10742403). Available from ProQuest Dissertations & Theses Global; Social Science Premium Collection. (1983418761). |
| Fagan, J., Iglesias, A., & Kaufman, R. (2016). Associations among Head Start fathers‚Äô involvement with their preschoolers and child language skills. *Early Child Development and Care, 186(8*), 1342-1356. |
| Fusaro, M., Harris, P. L., & Pan, B. A. (2012). Head nodding and head shaking gestures in children’s early communication. *First Language, 32*, 439-458. |
| Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition, 183*, 1-18. |
| Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. |
| Hardy-Brown, K., & Plomin, R. (1985). Infant communicative development: Evidence from adoptive and biological families for genetic and environmental influences on rate differences. *Developmental Psychology, 21(2),* 378. |
| Hart, B., & Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Paul H Brookes Publishing. |
| Himes, M. C. (1978). Language development and patterns of mother-child interaction. Doctoral Dissertation, University of Toronto. |
| Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K., & Suma, K. (2015). The contribution of early communication quality to low-income children’s language success. *Psychological Science, 26(7),* 1071-1083. |
| Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology, 22(2),* 155. |
| Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74(5),* 1368-1378. |
| Hoffman, L., Hersey, A., Tucker, R., & Vohr, B. (2020). Randomised control language intervention for infants of adolescent mothers. *ACTA PAEDIATRICA*, *109*(12), 2604–2613 |
| Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11(6),* F31-F39. |
| Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. Cognitive Development, 14(1), 57-75. |
| Jewkes, A. M. (2004). The power of the family: A longitudinal investigation of how the home environment influences preschool language development. (Doctoral dissertation, University of Michigan) |
| Johnson, K. M. (2016). A Longitudinal Investigation of Infant Gesture Use and Parent Speech: Unique and Dynamic Influences on Infant Vocabulary Acquisition (Doctoral dissertation, Duke University). |
| Kaufman, C.M. (1985). Parental Control Speech In Relation To Child Characteristics And Parity (mother, Father, Discipline) (Order No. 8508912). Available from ProQuest Dissertations & Theses Global. (303389578). |
| Kyger, M.F. (2013). Parents provide children with social cues for word learning (Doctoral dissertation, UCLA). |
| Larson, A. L., Barrett, T. S., & McConnell, S. R. (2019). Exploring early childhood language environments: A comparison of language use, exposure, and interactions in the home and childcare settings. *Language, Speech, and Hearing Services in Schools*, 1-14. |
| Lasky, E. Z., & Klopp, K. (1982). Parent-child interactions in normal and language-disordered children. *Journal of Speech & Hearing Disorders, 47*, 7-18. |
| Lecheile, B. M., Spinrad, T. L., Xu, X., Lopez, J., & Eisenberg, N. (2020). Longitudinal relations among household chaos, SES, and effortful control in the prediction of language skills in early childhood. *Developmental Psychology, 56(4)*, 727-738. |
| Leech, K. A., & Rowe, M. L. (2014). A comparison of preschool children’s discussions with parents during picture book and chapter book reading. *First Language, 34*, 205-226. |
| Liu, H. M. (2014). Lexical and acoustic features of maternal utterances addressing preverbal infants in picture book reading link to 5-year-old children's language development. *Early Education and Development, 25(8),* 1103-1117. |
| Lopez, L. D., Walle, E. A., Pretzer, G. M., & Warlaumont, A. S. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *Plos one, 15(11),* e0242232. |
| Lorang, E., Venker, C. E., & Sterling, A. (2020). An investigation into maternal use of telegraphic input to children with Down syndrome. *Journal of Child Language, 47(1),* 225-249. |
| Lueke, 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. |
| Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science, 21(6),* E12685. |
| Masur, E. F., Flynn, V., & Olson, J. (2016). Infants’ background television exposure during play: Negative relations to the quantity and quality of mothers’ speech and infants’ vocabulary acquisition. *First Language, 36(2),* 109-123. |
| Masur, E.F, & Gleason, J.B. (1980). Parent-child interaction and the acquisition of lexical information during play. *Developmental Psychology, 16(5),* 404-409. |
| Mimeau, C., Cantin, A., Tremblay, R. E., Boivin, M., & Dionne, G. (2020). The bidirectional association between maternal speech and child characteristics. *Journal of Child Language, 47(2)*, 435-456. |
| Moerk, E. L. (1975). Verbal interactions between children and their mothers during the preschool years. *Developmental Psychology, 11(6),* 788. |
| Muhinyi, A., & Rowe, M. L. (2019). Shared reading with preverbal infants and later language development. *Journal of Applied Developmental Psychology, 64*, 101053. |
| Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *Journal of Child Language, 43(5),* 1158-1173. |
| Ota, M., Davies‐Jenkins, N., & Skarabela, B. (2018). Why choo‐choo is better than train: The role of register‐specific words in early vocabulary growth. *Cognitive Science*, *42*(6), 1974-1999. |
| Paavola-Ruotsalainen, L., Lehtosaari, J., Palomäki, J., & Tervo, I. (2018). Maternal verbal responsiveness and directiveness: consistency, stability, and relations to child early linguistic development. *Journal of Child Language*, *45*(2), 319-339. |
| Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. Child Development, 76(4), 763-782. |
| Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology, 27(6),* 571-587. |
| Perlmann, R.Y. (1984). Variations In Socialization Styles: Family Talk At The Dinner Table (Order No. 8403811). Available from ProQuest Dissertations & Theses Global. (303298182). |
| Quigley, J., & Nixon, E. (2020). Infant language predicts fathers’ vocabulary in infant-directed speech. *Journal of Child Language, 47(1),* 146-158. |
| Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. |
| Reynolds, E., Vernon-Feagans, L., Bratsch-Hines, M., & Baker, C.E. (2019). Mothers’ and Fathers’ Language Input from 6 to 36 Months in Rural Two-Parent-Families: Relations to children’s kindergarten achievement. *Early Childhood Research Quarterly, 47*, 385-395. |
| Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children’s conversational exposure is associated with language-related brain function. *Psychological Science, 29(5),* 700-710. |
| Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83(5),* 1762-1774. |
| Rufsvold, R., Wang, Y., Hartman, M. C., Arora, S. B., & Smolen, E. R. (2018). The impact of language input on deaf and hard of hearing preschool children who use listening and spoken language. *American annals of the deaf, 163(1),* 35-60. |
| Salo, V. C., Rowe, M. L., Leech, K. A., & Cabrera, N. J. (2016). Low-income fathers’ speech to toddlers during book reading versus toy play. *Journal of Child Language, 43(6),* 1385-1399. |
| Scarborough, H., & Wyckoff, J. (1986). Mother, I'd still rather do it myself: some further non-effects of “motherese”. *Journal of Child Language, 13(2),* 431-437. |
| Sheran, C. P. (1999). Caregiver-child social communication: Effects of mother-child interactions on child development in the home. (Doctoral Dissertation, University of Minnesota). |
| Shimpi, P. M. & Huttenlocher, J. (2007). Redirective labels and early vocabulary development. *Journal of Child Language, 34*, 845-859. |
| Shimpi, P.M., Fedewa, A., & Hans, S. (2012). Social and linguistic input in low-income African American mother-child dyads from 1 month through 2 years: Relations to vocabulary development. *Applied Psycholinguistics, 33(4),* 781-798. |
| Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? Journal of Child Language, 40(3), 672-686. |
| Song, J. Y., Demuth, K., & Morgan, J. (2018). Input and processing factors affecting infants’ vocabulary size at 19 and 25 months*. Frontiers in psychology, 9*, 2398. |
| Song, L., Spier, E. T., & Tamis-Lemonda, C. S. (2014). Reciprocal influences between maternal language and children's language and cognitive development in low-income families. *Journal of Child Language, 41(2),* 305-326. |
| Spinelli, M., Fasolo, M., Shah, P. E., Genovese, G., & Aureli, T. (2018). The influence of early temperament on language development: the moderating role of maternal input. *Frontiers in psychology, 9*, 1527. |
| Stafford, L. (1987). Maternal input to twin and singleton children: Implications for language acquisition. *Human Communication Research, 13(4),* 429-462. |
| Sultana, N., Wong, L. L., & Purdy, S. C. (2020). Natural Language Input: Maternal Education, Socioeconomic Deprivation, and Language Outcomes in Typically Developing Children. *Language, Speech, and Hearing Services in Schools, 51(4),* 1049-1070. |
| Tabulda, G. A. (2017). *Role of parent oral language input in the development of child emergent literacy skills* (Doctoral dissertation, The Florida State University). |
| Tardif, T. Z. (1993). Adult-to-child speech and language acquisition in mandarin chinese (Order No. 9418555). Available from ProQuest Dissertations & Theses Global. (304053220). |
| Taumoepeau, M. (2016). Maternal expansions of child language relate to growth in children’s vocabulary. *Language Learning and Development, 12(4),* 429-446. |
| Thompson, A. (2018). Who’s Talking to Whom and Does It Matter? The Impact of Multiple Speakers, Overheard Speech, and Child-Directed Speech on Infants’ Language Development (Doctoral dissertation, University of Notre Dame). |
| Tomasello, M., Mannle, S., & Kruger, A. C. (1986). Linguistic environment of 1-to 2-year-old twins. *Developmental Psychology, 22(2),* 169. |
| Wade, M., Jenkins, J. M., Venkadasalam, V. P., Binnoon-Erez, N., & Ganea, P. A. (2018). The role of maternal responsiveness and linguistic input in pre-academic skill development: A longitudinal analysis of pathways. *Cognitive Development*, *45*, 125-140. |
| Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology, 50(2),* 336. |
| Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24(11),* 2143-2152. |
| Weizman, Z. O. (1997). Sophistication in maternal vocabulary input at home: Does it affect low-income children's vocabulary, literacy, and language success at school?  ProQuest Dissertations Publishing. |