Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin.* http://dx.doi.org/10.1037/bul0000202

Corresponding author: Naomi Sweller, Centre for Children's Learning in a Social World and Department of Psychology, Macquarie University, Australia.

Email: naomi.sweller@mq.edu.au

Abstract

Speech and gesture are two vital components of communication. Gesture itself provides an external support to speech, potentially promoting comprehension of a spoken message. The question of whether gesture promotes comprehension is not new, with research dating back to the 1970s. However, *when* gestures are most beneficial to comprehension is poorly understood. This meta-analysis explored two questions: whether and when gestures benefit comprehension of verbal information. We examined the effect sizes of 83 independent samples. Within each sample, a learner's comprehension was measured when gestures accompanied speech, compared to speech alone. Across all samples, gesture had a moderate, beneficial effect on comprehension when either produced or observed by a learner. Further stratified tests revealed that gestures significantly benefitted comprehension under a variety of circumstances, dependent on the type of gesture used, the information provided by gesture, the function of the gesture, the age of the learner, and the way comprehension was measured. The function of the gesture moderated the magnitude of the effect, with studies investigating the effect of producing gestures on comprehension yielding significantly larger effect sizes on average than studies investigating the effect of observing gestures on comprehension. The results from the current meta-analysis have theoretical and practical implications for gesture-related research and highlight new avenues for future studies.

Keywords: gesture, comprehension, learning, communication

Public Significance Statement

This meta-analysis reveals that gestures, when combined with speech, have a moderate effect on a learner's language comprehension. These effects are stronger when the learner produces the gestures themselves, as opposed to observing another individual's gestures.

When our hands help us understand: A meta-analysis into the effects of gesture on comprehension

Gestures, or movements made by the hands or arms, routinely accompany speech (McNeill, 1992). It is becoming increasingly clear that gestures can benefit comprehension of spoken messages. Indeed, a meta-analysis by Hostetter (2011) provided evidence across studies that observing gestures accompanying speech benefits comprehension to a greater extent than observing no gestures at all. Several studies have also found beneficial effects of gesture production on comprehension of spoken messages (e.g., Cook, Yip, & Goldin-Meadow, 2010), though such studies are yet to be analyzed through a meta-analysis. Despite these studies indicating beneficial effects of observing or producing gestures on speech comprehension, some studies still report non-significant findings that suggest observing or producing gestures, when combined with speech, are no more beneficial than neither observing nor producing gestures. Such variability in results suggests there may be factors that moderate the beneficial effects of gesture. The kind of gesture used, the information provided by gesture, the function of the gesture, the age of participants, and differences in the way "comprehension" is measured all have the potential to moderate gesture's effects on comprehension. For the purpose of the current meta-analysis the term comprehension refers to an individual's understanding of a presented verbal message, such as a narrative or set of verbal instructions.

Although several studies have looked at many of these possible moderators individually, it is currently unclear which of these moderators are the most influential in determining whether gesture benefits comprehension. A meta-analysis analyzing the effects of these potential moderators across studies will aid greatly in untangling their relative effects on the benefits of gesture on comprehension. The current meta-analysis focuses on investigating possible moderators of the effect of gesture on comprehension, in addition to furthering the results obtained by Hostetter (2011), to better establish *whether* and *when* gestures are beneficial to comprehension.

The question of whether gestures benefit comprehension is not new, with literature dating back to the 1970's. Berger and Popelka (1971) conducted one of the first experiments investigating whether the observation of gesture benefits comprehension. Indeed, Berger and Popelka found that when adults observed gesture, they comprehended sentences significantly better than when they did not observe gesture in conjunction with speech. However, they noted that they did not differentiate between different kinds of gestures observed and hypothesized that future research would likely find differential effects of different kinds of gestures. McNeill (1992) later suggested a classification system for gestures comprised of four distinct categories, based on Kendon's continuum (see Kendon, 1988 for a review). It is these categories that researchers primarily use to understand when and why gestures are beneficial to comprehension.

The Effect of Observing Different Kinds of Gestures on Comprehension

McNeill (1992) classified gestures in four ways: as iconic gestures, metaphoric gestures, deictic gestures, or beat gestures (McNeill, 1992). Iconic gestures represent a concrete action, event, or object, and the form they take is typically semantically related to the content of accompanying speech (Dargue & Sweller, 2018b). For example, a person making a fist with one hand and raising it to shoulder height while saying "the boy picked up the bucket." Gestures can also be used to represent an abstract metaphor, and in these circumstances a gesture would be categorized as metaphoric (McNeill, 1992). A person saying that their grades have improved while gesturing in an upward motion provides an example of a metaphoric gesture, as the upward motion symbolizes an improvement.

In contrast, deictic or pointing gestures function to indicate an event, direction, or object (McNeill, 1992). For example, a toddler pointing to an apple while stating “apple.” Finally, beat gestures are rhythmic, flicking movements of the hands that accompany speech (McNeill, 2000). Such gestures have no semantic relation to the content of accompanying speech, for example, a woman making a flicking motion with her hands while saying “book.” In this instance, the flicking motion bears no semantic relation to a book. Such gestures function to emphasize the content of accompanying speech through acting as a highlighter to focus a learner’s attention on important information (Biau & Soto-Faraco, 2013; Holle et al., 2012; Krahmer & Swerts, 2007).These four categories have been the subject of much research concerning gesture and comprehension, and the results of such research have led to a variety of theories surrounding why gesture may benefit comprehension.

It has long been suggested that observing gestures may benefit comprehension simply through drawing attention to the spoken message (Kendon, 1994). However, recent studies have suggested otherwise. Recent findings suggest that some gestures are more beneficial than others, and one mechanism underlying this might be due to the semantic relatedness that a gesture has with the accompanying speech (Dargue & Sweller, 2018b). That is, the more semantically related a gesture is with speech, the more beneficial it may be to comprehension, due to the semantic integration (i.e., binding) of speech and gestural information (Straube, Green, Weis, Chatterjee, & Kircher, 2008). If gestures are more beneficial when they are semantically related to the content of speech, then iconic or metaphoric gestures, which by definition are semantically related to speech, may be more beneficial to comprehension than other forms of gestures.

Indeed, several studies have found that iconic gestures are beneficial to comprehension when observed. An early experiment by Beattie and Shovelton (1999) found that adults who observed iconic gestures comprehended a verbal narrative significantly better than adults who did not see gestures. However, they suggested that not all iconic gestures are necessarily beneficial, and recent research has indeed found this to be the case, with some iconic gestures found to be more beneficial to comprehension than others (Dargue & Sweller, 2018b). For example, Dargue and Sweller (2018b) found that observing typical iconic gestures (i.e., gestures produced frequently by an individual) was significantly more beneficial to narrative comprehension than watching no gestures. However, observing atypical iconic gestures (e.g., gestures produced infrequently by an individual) was no more beneficial than observing no gestures at all. One reason for this difference could be due to the typical iconic gestures being more semantically related to the content of accompanying speech than the atypical iconic gestures (Dargue & Sweller, 2018b). Thus, although some studies investigating the effect of observing iconic gestures have found non-significant results (e.g., Dahl & Ludvigsen, 2014; Rowe, Silverman, & Mullan, 2013), it is possible that perhaps the iconic gestures used in these studies were not sufficiently semantically related to the content of speech to be beneficial.

Despite the large number of experiments examining iconic gestures, research on metaphoric gestures is limited. Only two experiments identified for the current meta-analysis found significant benefits of observing metaphoric gestures in combination with speech, with the metaphoric gestures promoting second language comprehension (Repetto, Pedroli, & Macedonia, 2017) and comprehension of abstract sentences (Straube et al., 2008) in adults.

If the underlying reason behind why gestures are beneficial to comprehension is as simple as whether the gestures are semantically related to the accompanying speech, it would be expected that deictic and beat gestures, which are not semantically related to the content of speech, would not benefit comprehension. However, this does not appear to be the case. Although some studies have failed to find a significant benefit of deictic gestures on comprehension (e.g., Kelly, 2001; Ouwehand, van Gog, & Paas, 2015; Sekine & Kita, 2017), many studies have found significant benefits. Cook, Duffy, and Fenn (2013) showed children in primary school deictic gestures when being taught how to complete mathematical problems. The children who viewed the deictic gestures in combination with speech performed significantly better on the mathematical problems than children who were provided with instruction through speech only (i.e., saw no gestures). Similar findings were obtained by Pi, Hong, and Yang (2017) using adults, with deictic gestures leading to significantly better comprehension of a visual lecture compared to when no gestures were observed. This same study by Pi et al. also found that the adults who observed deictic gestures attended to the lecture significantly more than adults who saw no gestures, suggesting that perhaps the deictic gestures benefit comprehension through capturing attention.

If gestures can benefit comprehension simply through capturing attention, it would be expected that beat gestures would also benefit comprehension. Indeed, Igualada, Esteve-Gibert, and Prieto (2017) found that children performed significantly better on a word comprehension task when a word was accompanied by a beat gesture. However, the results obtained by Gluhareva and Prieto (2017) suggested that the story is more complicated than beat gestures simply capturing attention. In their study, beat gestures only significantly benefitted adult second language comprehension when the words taught were difficult. When the words were easy, the observation of beat gestures was no more beneficial than seeing no gestures at all.

Given the variable results obtained across different kinds of gestures, it is of interest to determine whether, across studies, the different kinds of gestures benefit comprehension, and whether some gestures are more beneficial than others. Perhaps iconic and metaphoric gestures are most beneficial to comprehension, given the semantic integration of speech and gesture. Hostetter (2011) could not answer this question, as at the time of the 2011 analysis, limited research on different kinds of gestures had been conducted. In the intervening years, there has been much growth in the publication of gesture research, enabling such comparisons to now be made. The current meta-analysis therefore aims to explore whether certain kinds of gestures are more beneficial than others, to better understand when gestures are beneficial. However, perhaps gestures are also more beneficial when they function as a means of disambiguating accompanying speech through providing additional information.

The Effect of Information Provided Through Gesture on Comprehension

While many studies have explored the effect of gestures that simply match the content of accompanying speech (i.e., gestures that are redundant with speech), the number of studies exploring the effect of gestures that provide additional information above and beyond speech is increasing. Furthermore, Hostetter (2011) provided evidence across studies that gestures that provide additional information to speech are more beneficial to comprehension than gestures that are redundant with speech. This finding suggests that gestures may add information to the accompanying speech that informs the intention of the speaker, and in this way, the gestures may combine with speech to clarify the meaning of an ambiguous spoken message (Kelly, 2001; Kelly & Barr, 1999). Such findings are consistent with neuroscience research examining the disambiguation of spoken messages through the use of gestures that provide additional information to speech (Gunter & Weinbrenner, 2017; Holle et al., 2012; Holle & Gunter, 2007; Holle, Obleser, Rueschemeyer, & Gunter, 2010). If gestures indeed interact with and subsequently clarify the meaning of the speaker's message, it would be expected that gestures that provide additional information above and beyond speech would benefit comprehension.

Several studies also report that the observation of gestures that are redundant with speech significantly benefit comprehension compared to observing no gestures at all. For example, Dargue and Sweller (2018b) found that the observation of redundant iconic gestures benefitted adult comprehension to a greater extent than observing no gestures. A similar result was found in a study with preschool children, whereby preschool children who observed redundant gestures that reinforced the accompanying speech comprehended a verbal narrative significantly better than preschool children who observed no gestures (Dargue & Sweller, 2018a). It therefore appears that redundant gestures can be beneficial, and as a result this meta-analysis aims to not only replicate the findings of Hostetter (2011), but also extend them through distinguishing whether redundant gestures across studies indeed have a significant, beneficial effect on comprehension. While gestures may differ in terms of whether they provide additional information to speech, gestures also differ in terms of the function that they serve.

The Effect of Gesture Observation Compared to Gesture Production on Comprehension

Although the meta-analysis by Hostetter (2011) found that, across studies, gestures were beneficial to comprehension when observed by a learner, gestures can also have a self-oriented function (Kita, 2000). That is, the production of gesture can be beneficial not only to the observer, but also to the speaker themselves (Kita, 2000). The effects of gesture production as compared with gesture observation are yet to be investigated through a meta-analysis. It has been claimed that the production of gesture by a learner benefits comprehension more than the observation of gesture (Goldin-Meadow et al., 2012). It is thought that the production of gesture reduces cognitive load, allowing for more resources to be allocated to the task at hand, subsequently benefitting comprehension (Cook et al., 2010).

Numerous studies have found that the production of gesture can indeed benefit comprehension, for example in the completion of mental rotation (Chu & Kita, 2011) and mathematical problem solving (Goldin-Meadow, Cook, & Mitchell, 2009) tasks. However, other studies have found no benefit of producing gesture in conjunction with speech on comprehension (e.g., Alibali, Spencer, Knox, & Kita, 2011; Lajevardi, Narang, Marcus, & Ayres, 2017). Given this variation across studies, a consensus is yet to be reached surrounding whether the production of gesture indeed benefits comprehension. Another aim of the current meta-analysis is to ascertain whether this beneficial effect of gesture production is found across studies, in addition to determining whether producing and observing gestures have differential effects on comprehension. While the potential moderators covered so far focus mainly on the gestures themselves used in experiments, it has been argued throughout the literature that regardless of gesture type, perhaps gestures are more beneficial to children than adults.

The Effect of Gesture on Comprehension Across Different Age Groups

Verbal skills are not yet fully developed in young children (Hostetter, 2011), and as a result it is possible that young children may benefit more from the visual information provided by gestures compared to older children or adults. That is, the gesture may help to disambiguate the meaning of difficult speech. Indeed, McNeil, Alibali, and Evans (2000) found gestures were beneficial to comprehension in preschool children, but not children in their first year of primary school. It was suggested that perhaps these findings result from the speech being difficult for the preschool children to understand, and thus they benefitted from the visual information provided by the gestures. Although task difficulty is a plausible explanation for this difference, past behavioral research has suggested that the beneficial effect that gesture has on comprehension may follow a U-shaped curve across development (Church, Kelly, & Lynch, 2000). Perhaps a learner's age therefore moderates the beneficial effect of gesture.

Other studies have also found significant benefits of gesture on comprehension in preschool children but not adults (Austin & Sweller, 2014), further suggesting that perhaps gestures are more beneficial to young children. However, there are several studies in the literature that do find significant benefits of gesture for children in primary school (e.g., Beaudoin-Ryan & Goldin-Meadow, 2014; Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Church, Ayman-Nolley, & Mahootian, 2004; Cook et al., 2013; Kirk & Lewis, 2017), adolescents (e.g., Dahl & Ludvigsen, 2014), and adults (e.g., Beattie & Shovelton, 1999; Berger & Popelka, 1971; Chu & Kita, 2011; Dargue & Sweller, 2018b; Driskell & Radtke, 2003). While Hostetter

(2011) found evidence for age being a possible moderator of the effects of gesture on comprehension, with children benefitting more from gesture than adults, the extent to which gesture benefits comprehension is yet to be determined in the different age groups (i.e., preschool aged children, primary school aged children, adolescents, and adults) across studies. Thus, the current meta-analysis also aims to extend Hostetter's findings to determine which age groups gesture significantly benefits, and to determine whether preschool children indeed benefit from producing or observing gesture significantly more than primary school aged children, adolescents, and adults.

The potential moderators discussed to this point have involved characteristics either of the gestures themselves, or of the participants involved in the studies. A further overarching variable must be considered, however. The way in which comprehension is *measured* could also affect how beneficial gestures are to comprehension.

The Effect of Gesture on Different Measures of Comprehension

Across the studies identified for inclusion in the current meta-analysis, a variety of measures of comprehension were used. The majority of studies used recall to gauge comprehension. Some studies used free recall (e.g., "Tell me everything you can remember about the story you saw earlier"), while others relied on open-ended questions that included some form of semantic prompt (e.g., "How did Donald Duck feel when the water was not going into the bucket?"). Other studies used multiple-choice or forced-choice questions as a means of measuring comprehension (e.g., "Did Donald duck feel sad or frustrated?"), or a mixture of the methods mentioned above. Although various studies have shown significant benefits of gesture across the methods mentioned above, some studies that have used free recall have not (e.g., Dargue & Sweller, 2018a; Macoun & Sweller, 2016), and have argued that perhaps gestures are only beneficial to the specific spoken content that they accompany (Dargue & Sweller, 2018a). If gestures primarily benefit the speech that they directly accompany (i.e., the gestures do not improve comprehension of an entire spoken message or story), then it would be expected that studies using free recall would find smaller effects than studies that use other methods such as open-ended questions with semantic prompting, forced-choice questions, or multiple-choice questions.

The Current Study

In summary, the current paper aimed to summarize the existing work on the effects of observing and producing gestures on comprehension to better understand when gestures are beneficial. The following three questions were addressed:

1. Across studies, does the presence of gesture (either observed or produced) benefit comprehension of a spoken message to a greater extent than when the spoken message is not accompanied by gesture?
2. Across studies, does gesture benefit comprehension when:
    - Observing iconic, metaphoric, deictic, beat, and a mixture of gestures?
    - Observing gestures that provide additional information and observing gestures that are redundant?
    - A learner observes or produces gesture?
    - Preschool aged children, primary school aged children, adolescents, and adults observe or produce gesture?
    - Tested through free recall, open-ended/specific questions, multiple-choice/forced-choice questions, and a mixture of the abovementioned methods?
3. Do the factors listed above in question two significantly moderate the effect that gesture has on learning?

Method

Selection Criteria

There were six criteria each study had to meet for inclusion in the meta-analysis. First, the study had to use a human sample. Although the effectiveness of observing gesture on comprehension has been investigated in non-human primates (e.g., Bohn, Call, & Tomasello, 2016), such issues fall outside the scope of the research questions of interest in the current study.

Second, the study had to use a sample free of developmental abnormalities with normal or corrected to normal hearing and vision, free from developmental disorders (e.g., Autism Spectrum Disorder, Down Syndrome) or acquired disorders (e.g., Alzheimer's Dementia, Apraxia). Several studies have investigated the production or observation of gestures with respect to comprehension in individuals with developmental or acquired disorders (e.g., Pashek & DiVenere, 2006; Rothi, Heilman, & Watson, 1985; Viher et al., 2018; Wang, Bernas, & Eberhard, 2004; Wong & So, 2018). However, including individuals with developmental or acquired disorders within the overall analysis could confound results, due to the cognitive differences that commonly exist between typically developing individuals and individuals with developmental or acquired disorders, particularly with respect to comprehension.

Third, the participants in the study had to be at least 3 years old. Although prior research has investigated the production and observation of gesture with respect to comprehension in infants (e.g., Iverson, Capirci, Volterra, & Goldin-Meadow, 2008; Namy, Vallas, & Knight-Schwarz, 2008; Rowe, Özçalişkan, & Goldin-Meadow, 2008), the ability to relate an observed gesture to a given referent in speech does not develop until during the second year of life (Namy, 2008). Children are successfully able to grasp the meaning of co-speech gestures by 3 years of age (Stanfield, Williamson, & Özçalişkan, 2014). Thus, using children younger than 3 years old could confound results, given any null effect of gesture could simply be due to the inability of the children to interpret co-speech gestures.

Fourth, only studies that experimentally manipulated whether a learner produced or observed gestures were included in the analysis. For studies concerning whether the production of gesture by a learner aids comprehension, it was a requirement that there was one condition whereby participants were given explicit instruction to gesture, and one condition where no instructions were given surrounding the production of gesture. For studies that investigated whether the observation of gesture by a learner benefits comprehension, there had to be one condition whereby participants observed a speaker produce gestures and one condition where no gestures were observed. That is, a condition in which the speaker held their hands still was required. As a result, studies concerning the observation of spontaneously produced gesture were not included in the analysis, as there was no deliberate manipulation of gesture (e.g., Francaviglia & Servidio, 2011). Furthermore, studies that compared the observation of one kind of gesture in one condition to the observation of another kind of gesture in another condition (e.g., iconic gesture vs. deictic gestures) in the absence of a speech only control condition were also excluded (e.g., Alibali, Flevares, & Goldin-Meadow, 1997; Kang & Tversky, 2016).

Fifth, some studies included conditions whereby participants observed gestures during a comprehension task (e.g., while being told a story), and were then asked to produce gestures during recall (e.g., Macedonia & Knösche, 2011). As it is unclear whether any beneficial effect on comprehension arises from gesture production or observation in these situations, analyses were only conducted when the observation of gesture only was compared to a no gesture control, or when the instructed production of gesture during recall was compared to no gesture production during recall.

Sixth, studies had to include a behavioral measure of comprehension as the dependent variable. Studies that were concerned with neurophysiological measures (e.g., electroencephalography or functional magnetic resonance imaging) in the absence of a behavioral measure were excluded (e.g., Kelly, Kravitz, & Hopkins, 2004; Wu & Coulson, 2005). Case studies were also excluded (e.g., McCafferty, 2002), as were studies concerned with the comprehension of an observed gesture rather than comprehension of a verbal task (e.g., Colletta, Pellenq, & Guidetti, 2010).

Databases were searched between October 2017 and April 2018 for relevant articles using a variety of key terms, as described in detail below, to reduce the risk of bias associated with missing relevant articles. The period over which the databases were searched ranged from January 1970 to March 2018. Databases searched included Education Resources Information Center (ERIC), PsycINFO, and Scopus. For each database, the keyword *gesture* was searched in conjunction with the terms *comprehension*, *memory*, *problem solving*, or *recall* separately. This is such that for ERIC, PsycINFO, and Scopus the following search strings were used: gesture and comprehension, gesture and memory, gesture and problem solving, gesture and recall. In August 2018, a second independent coder familiar with the field of gesture and learning searched the terms “gesture and comprehension” in ERIC, PsycINFO, and Scopus to check the consistency of the articles that were retrieved. Both the first and second coders retrieved the same papers. Further studies were identified through Google Scholar and through examination of the reference list of the prior meta-analysis concerned with gesture and comprehension (Hostetter, 2011).

All searches were limited to peer-reviewed papers. For each study retrieved, the title and abstract were considered alongside the abovementioned six criteria. If a study appeared to meet the six criteria after consideration of the title and abstract, the full text was then read in depth to determine if the study met all of the inclusion criteria. We considered only peer-reviewed, published manuscripts in order to reduce bias due to poor study quality that is associated with the inclusion of non-peer-reviewed data (e.g., conference proceedings or dissertations).

The primary author performed the initial critical appraisal and eligibility assessment of each paper. In the event of ambiguity, papers were assessed by the full panel of authors to determine whether a given study was eligible for inclusion or not. The use of the abovementioned strictly defined eligibility criteria reduced the possible impact of poor-quality empirical studies on results (see Figure 1 for PRISMA flowchart).

Quality Assessment

Each study that was deemed to have met the six-abovementioned inclusion criteria was evaluated for study quality. Given that all studies included were empirical in nature, the quality of each study was evaluated against relevant CASP checklist criteria for experimental studies (Singh, 2013), and relevant additional criteria for within-subjects studies. Five factors were assessed including random allocation (between-subjects studies only), counterbalancing to prevent order effects (within-subjects studies only), whether all participants who initially participated were accounted for at the conclusion, whether study personnel were blind to the condition that participants were allocated to, whether groups were treated equally aside from the experimental manipulation (e.g., all non-manipulated instructions were identical), and whether the dependent variable was clearly specified. The number of criteria that each study met was summed to yield a total quality score out of five, such that a higher score represented a higher quality study. All studies received a score of either 4 or 5.

All between-subject studies randomly allocated participants to conditions, and all within-subjects studies counterbalanced appropriately. All studies accounted for all participants who participated initially and gave clear reasons for any participants being excluded from analysis (e.g., technological failure, experimenter error, participant did not pay adequate attention). Groups were treated equally in all studies, and the dependent variable of interest was clearly defined by all studies. However, included studies seldom ensured that study personnel were blind to the condition each participant was in, with only one study out of the 83 studies meeting this criterion. This result highlights the need for further research in this area to consider ensuring, if possible, that study personnel are blind to a participant's allocated condition.

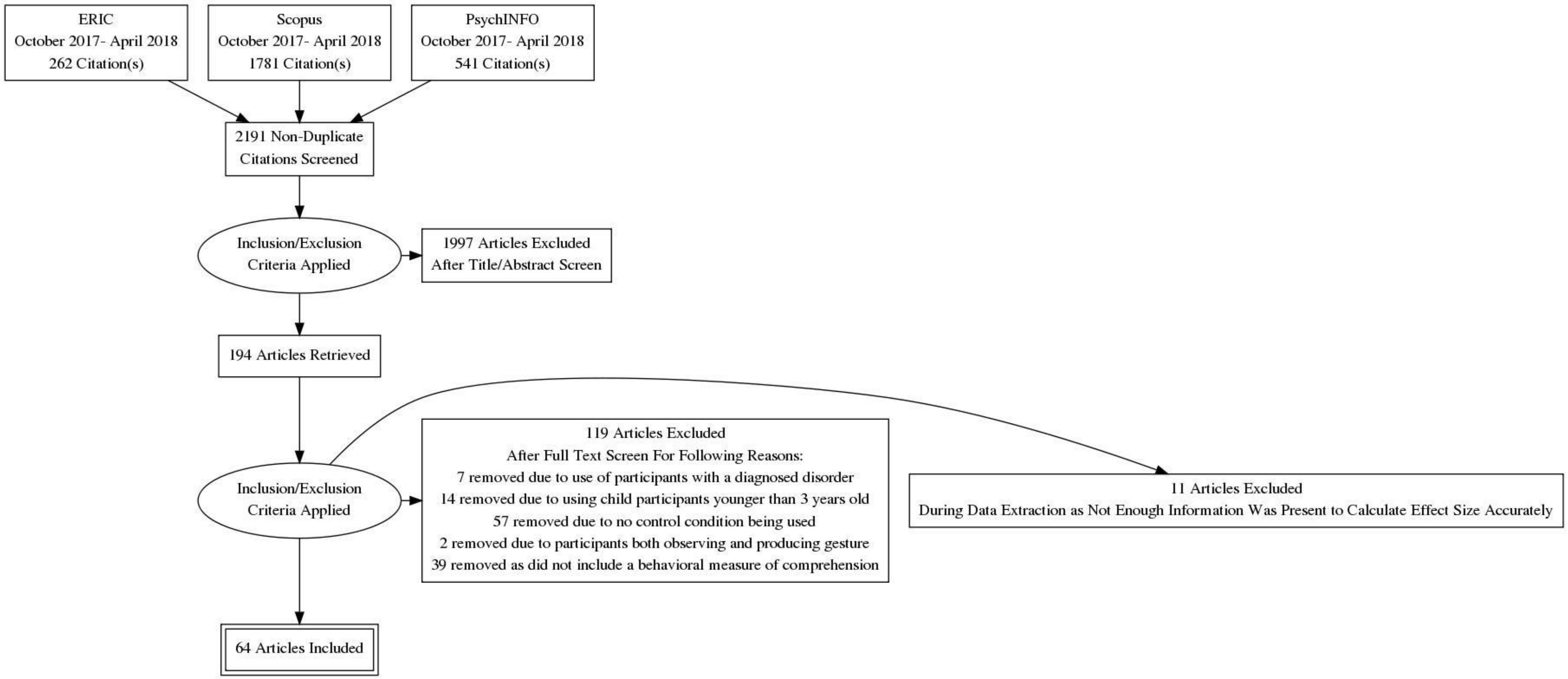

Figure 1. PRISMA flowchart of screening process for article selection (Moher, Liberati, Tetzlaff, & Altman, 2010).

Data Extraction

Data relating to the study (e.g., country, publication year), participants (e.g., age group, number of participants), methods used (e.g., between- or within-subjects design, type of gestures used), and measures of comprehension (e.g., means, standard deviations, exact *F*-, *t*-, or *p*-statistics) were extracted from each study. If any study reported having conducted analyses on relevant variables but did not publish the relevant data, the corresponding author was contacted, and the data were requested for inclusion. Of the seven authors contacted, two responded with the relevant data and were subsequently included in the meta-analysis.

Each study was classified according to the following characteristics to identify whether significant variation in effect sizes could be explained by these variables:

- The design of the study (between-subjects or within-subjects)
- The kind of gesture observed (iconic, metaphoric, deictic, beat, or mixed)
- Whether gestures observed provided additional information beyond accompanying speech
- Whether gestures were produced or observed
- The age of participants
- The method used to measure comprehension

Study designs were classified as between-subjects or within-subjects. Studies were further categorized according to the kinds of gestures observed in a given study into five categories: iconic, metaphoric, deictic, beat, or mixed. Studies that used gestures that provided information that was semantically related to the content of accompanying speech were classified as having used iconic gestures. If a study used gestures that presented an abstract or concrete metaphor for a concept, the study was categorized as having used metaphoric gestures. Studies that used pointing gestures that functioned as a way of indicating objects, events, or directions were categorized as having used deictic gestures. When studies used gestures that were simple rhythmic movements used to emphasize particular words without portraying semantic meaning, the study was classified as having used beat gestures. If a paper indicated that they used a combination of different kinds of gestures (e.g., a mix of iconic and deictic gestures), then the study was categorized as mixed. These categories of gestures reflected the gesture classification system provided by McNeill (1992).

The kinds of information portrayed by observed gesture was also classified into two categories: additional information or redundant information. Categorization was dependent on whether the gestures observed provided additional information beyond the content of associated speech or not. That is, gestures that portrayed information not presented in speech were categorized as providing additional information, whereas gestures that provided no additional information (i.e., the information provided through gesture matched the information provided through speech) were categorized as providing redundant information. Studies were further categorized dependent on whether the effect of interest concerned the production of gesture by a learner, or the observation of gesture by a learner, resulting in a further two categories: Gesture production or gesture observation.

Studies were also classified into five categories according to age: Preschool children, primary (elementary) school children, adolescents (students attending high school), adults (university students and community members over 18 years of age), or older adults (over 60 years of age). Finally, studies were categorized according to the way that comprehension was measured, into one of four classifications: free recall, open-ended/specific questions, multiple/forced-choice, or mixed. If participants were asked to recall information freely in the absence of any specific prompts or cues, the study was categorized as using a free recall measure. If participants were asked a series of open-ended or specific questions, then the study was categorized as using open-ended/specific questions. In comparison, if participants were given multiple-choice or forced-choice questions then the study was categorized as using a multiple/forced-choice measure. Lastly, if the study presented a mix of the above-mentioned measures to participants, such as a mix of specific and forced-choice questions, then the study was coded as using a mixed measure. See Table 1 for the full list of studies retrieved, along with each article's classification according to the criteria listed above.

Table 1

Summary of Studies Included in the Current Meta-analysis

Author(s)	Year	Country	Study	Age Group	Design	N	Gesture Type	Production or Observation	Additional Information	Comprehension Measure	Benefit	*d*
Alibali & Dirusso	1999	USA	1	Preschool	Within	20	Production	Production	N/A	Specific/Open-ended Questions	Y	1.60
Alibali et al.	2011	USA	1	Adults	Between	85	Production	Production	N/A	Specific/Open-ended Questions	N	0.00
Alibali, et al.	2011	USA	2	Adults	Between	109	Production	Production	N/A	Specific/Open-ended Questions	N	0.18
Austin & Sweller	2014	Australia	1	Preschool	Between	91	Mixed	Observation	N	Free Recall	Y	0.68
Austin & Sweller	2014	Australia	1	Adults	Between	94	Mixed	Observation	N	Free Recall	N	0.02
Austin & Sweller	2017	Australia	1	Preschool	Between	172	Mixed	Observation	N	Specific/Open-ended Questions	Y	0.50
Austin, Sweller, & Van Bergen	2018	Australia	1	Adults	Between	57	Mixed	Observation	N	Free Recall	N	0.17
Austin et al.	2018	Australia	2	Adults	Between	125	Mixed	Observation	N	Free Recall	N	0.20
Beattie & Shovelton	1999	UK	1	Adults	Between	60	Iconic	Observation	N	Multiple/Forced-choice	Y	1.48
Beaudoin-Ryan & Goldin-Meadow	2014	USA	1	Primary School	Between	26	Production	Production	N/A	Specific/Open-ended Questions	Y	0.84
Berger & Popelka	1971	USA	1	Adults	Within	32	Mixed	Observation	N	Specific/Open-ended Questions	Y	0.62
Broaders et al.	2007	USA	2	Primary School	Between	70	Production	Production	N/A	Specific/Open-ended Questions	Y	0.49
Bucciarelli, Mackiewicz,	2016	Italy	3	Primary School	Within	32	Production	Production	N/A	Specific/Open-ended Questions	Y	0.93

Khemlani, & Johnson-Laird												
Cameron & Xu	2011	USA	2	Preschool	Between	30	Production	Production	N/A	Free Recall	Y	2.16
Carlson, Jacobs, Perry, & Breckinridge Church	2014	USA	1	Adults	Between	56	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.59
Chu & Kita	2011	UK	3	Adults	Between	32	Production	Production	N/A	Multiple/Forced-choice	Y	0.70
Church et al.	2004	USA	1	Primary School (English Speaking)	Between	26	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.87
Church et al.	2004	USA	1	Primary School (Spanish Speaking)	Between	25	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.62
Church, Garber, & Rogalski	2007	USA	1	Adults	Between	45	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.65
Congdon et al.	2017	USA	1	Primary School	Between	48	Deictic	Observation	N	Specific/Open-ended Questions	N	0.18
Cook et al.	2013	USA	1	Primary School	Between	184	Deictic	Observation	N	Specific/Open-ended Questions	Y	0.49
Cook, Friedman, Duggan, Cui, & Popescu	2017	USA	1	Primary School	Between	38	Mixed	Observation	N	Specific/Open-ended Questions	Y	1.04
Cook et al.	2010	USA	4	Adults	Between	15	Production	Production	N/A	Free Recall	Y	1.70
Dahl & Ludvigsen	2014	USA /Norway	1	Adolescents (Non-fluent)	Between	46	Iconic	Observation	N	Free Recall	Y	1.02
Dahl & Ludvigsen	2014	USA /Norway	1	Adolescents (Fluent)	Between	28	Iconic	Observation	N	Free Recall	N	0.26
Dargue & Sweller	2018a	Australia	1	Preschool	Between	42	Iconic	Observation	N	Mixed	Y	0.79

Dargue & Sweller	2018b	Australia	2	Adults	Between	86	Iconic	Observation	N	Mixed	Y	0.84
De Nooijer, Van Gog, Paas, & Zwaan	2013	Netherlands	1	Primary School	Between	115	N/A	Production	N	Specific/Open-ended Questions	Y	0.61
De Nooijer, Van Gog, Paas, & Zwaan	2014	Netherlands	1	Primary School	Within	49	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.30
Driskell & Radtke	2003	USA	1	Adults	Between	40	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.76
Feyereisen	2006	Belgium	1	Adults	Within	52	Mixed	Observation	N	Free Recall	Y	0.52
Gluhareva & Prieto	2017	Spain	1	Adults	Within	20	Beat	Observation	N	Specific/Open-ended Questions	N	0.25
Gunderson, Spaepen, Gibson, Goldin-Meadow, & Levine	2015	USA	1	Preschool	Within	73	Production	Production	N/A	Specific/Open-ended Questions	Y	0.53
Heidari	2015	Iran	1	Preschool	Between	50	Mixed	Observation	N	Specific/Open-ended Questions	Y	0.54
Hupp & Gingras	2016	USA	1	Adults	Between	67	Iconic	Observation	N	Multiple/Forced-choice	Y	1.14
Hupp & Gingras	2016	USA	2	Adults	Between	41	Iconic	Observation	N	Multiple/Forced-choice	Y	2.31
Ianì & Bucciarelli	2017	Italy	1	Adults	Within	25	Iconic	Observation	N	Free Recall	Y	0.59
Igualada et al.	2017	Spain	1	Preschool	Within	106	Beat	Observation	N	Free Recall	Y	0.19
Kang, Hallman, Son, & Black	2013	USA	1	Adults	Between	33	Iconic	Observation	N	Mixed	Y	0.47
Kelly	2001	USA	1	Preschool (younger)	Within	14	Deictic	Observation	Y	Specific/Open-ended Questions	N	0.08
Kelly	2001	USA	1	Preschool (older)	Within	15	Deictic	Observation	Y	Specific/Open-ended Questions	Y	0.67
Kelly	2001	USA	2	Preschool (younger)	Within	13	Deictic	Observation	Y	Specific/Open-ended Questions	Y	0.81

Kelly	2001	USA	2	Preschool (older)	Within	14	Deictic	Observation	Y	Specific/Open-ended Questions	N	0.13
Kelly, Barr, Church, & Lynch	1999	USA	1	Adults	Within	16	Deictic	Observation	Y	Specific/Open-ended Questions	Y	0.72
Kelly et al.	1999	USA	2	Adults	Within	18	Deictic	Observation	Y	Specific/Open-ended Questions	Y	0.48
Kelly et al.	1999	USA	4	Adults	Within	15	Iconic	Observation	Y	Specific/Open-ended Questions	Y	0.48
Kelly & Church	1998	USA	1	Adults	Within	18	Iconic	Observation	N	Free Recall	N	0.13
Kelly & Lee	2011	USA	1	Adults	Within	42	Iconic	Observation	N	Multiple/Forced-choice	Y	0.21
Kelly, McDevitt, & Esch	2009	USA	1	Adults	Within	27	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.50
Kelly et al.	2009	USA	2	Adults	Within	24	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.63
Kirk & Lewis	2017	UK	2	Primary School	Between	54	Production	Production	N/A	Specific/Open-ended Questions	Y	0.51
Koumoutsakis, Church, Alibali, Singer, & Ayman-Nolley	2016	USA	1	Primary School	Between	63	Deictic	Observation	N	Specific/Open-ended Questions	Y	0.36
Krauss, Dushay, Chen, & Rauscher	1995	Colombia	1	Adults	Between	86	Mixed	Observation	N	Multiple/Forced-choice	N	0.12
Krauss et al.	1995	Colombia	2	Adults	Between	98	Mixed	Observation	N	Multiple/Forced-choice	N	0.11
Krauss et al.	1995	Colombia	3	Adults	Between	43	Mixed	Observation	N	Multiple/Forced-choice	N	0.05
Krauss, Morrel-Samuels, & Colasante	1991	Colombia	3	Adults	Between	72	Mixed	Observation	N	Multiple/Forced-choice	N	0.09

Lajevardi et al.	2017	Australia	1	Adults	Between	24	Production	Production	N/A	Specific/Open-ended Questions	N	0.51
Lajevardi et al.	2017	Australia	2	Primary School	Between	22	Production	Production	N/A	Specific/Open-ended Questions	Y	2.50
Lickiss & Wellens	1978	USA	1	Adults	Between	20	Mixed	Observation	N	Specific/Open-ended Questions	N	-1.14
Macken & Ginns	2014	Australia	1	Adults	Between	42	Production	Production	N/A	Multiple/Forced-choice	Y	0.62
Macoun & Sweller	2016	Australia	1	Preschool	Between	50	Mixed	Observation	N	Free Recall	Y	0.89
McNeil et al.	2000	USA	2	Preschool	Within	13	Iconic	Observation	N	Specific/Open-ended Questions	N	0.00
Ouwehand et al.	2015	Netherlands	1	Primary School	Between	61	Deictic	Observation	N	Multiple/Forced-choice	N	-0.17
Ouwehand et al.	2015	Netherlands	1	Adults	Between	39	Deictic	Observation	N	Multiple/Forced-choice	N	0.09
Ouwehand et al.	2015	Netherlands	1	Older Adults	Between	58	Deictic	Observation	N	Multiple/Forced-choice	N	0.29
Perry, Berch, & Singleton	1995	USA	1	Primary School	Between	38	Deictic	Observation	N	Specific/Open-ended Questions	Y	0.92
Pi et al.	2017	China	1	Adults	Between	56	Deictic	Observation	N	Mixed	Y	1.27
Pine, Knott, & Fletcher	2010	UK	1	Preschool	Between	44	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.71
Ping & Goldin-Meadow	2008	USA	1	Primary School	Between	52	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.73
Ping & Goldin-Meadow	2008	USA	2	Primary School	Between	45	Iconic	Observation	N	Specific/Open-ended Questions	Y	0.78
Repetto et al.	2017	Italy	1	Adults	Within	20	Metaphoric	Observation	N	Free Recall	Y	2.05
Riseborough	1981	UK	2	Adults	Between	12	Iconic	Observation	N	Specific/Open-ended Questions	Y	1.33
Rogers	1978	USA	1	Adults	Between	40	Mixed	Observation	N	Multiple/Forced-choice	Y	0.51
Rowbotham, Holler, Wearden, & Lloyd	2016	UK	1	Adults	Between	67	Mixed	Observation	Y	Free Recall	Y	2.64

Rowe et al.	2013	USA	1	Preschool	Within	62	Iconic	Observation	N	Specific/Open-ended Questions	N	0.02
Sekine & Kita	2017	Netherlands	3	Adults	Within	30	Deictic	Observation	N	Specific/Open-ended Questions	N	0.15
Stieff, Lira, & Scopelitis	2016	USA	1	Adults	Between	44	Iconic	Observation	N	Specific/Open-ended Questions	N	0.20
Straube et al.	2008	Germany	1	Adults	Within	12	Metaphoric	Observation	N	Forced-choice	Y	0.78
Sueyoshi & Hardison	2005	USA	1	Adults (low-proficiency)	Between	14	Mixed	Observation	N	Multiple/Forced-choice	Y	0.92
Sueyoshi & Hardison	2005	USA	1	Adults (high-proficiency)	Between	14	Mixed	Observation	N	Multiple/Forced-choice	N	-1.07
Valenzeno, Alibali, & Klatzky	2003	USA	1	Preschool	Between	25	Deictic	Observation	N	Specific/Open-ended Questions	Y	0.70
van Wermeskerken, Fijan, Eielts, & Pouw	2016	Netherlands	1	Primary School	Between	51	Iconic	Observation	N	Free Recall	N	0.51
Wakefield & James	2015	USA	1	Primary School	Between	60	Production	Production	N/A	Specific/Open-ended Questions	Y	0.57

Note. *d* represents unbiased Cohen’s d (effect size).

Reliability

A second coder familiar with the literature surrounding gesture and learning independently coded 17 randomly selected samples (20% of the data included within the meta-analysis) to assess inter-rater reliability. Cohen's kappa was used to evaluate agreement of categorical data. For study design, κ= 1.00, $p < .0005$, for gesture type κ= 1.00, $p < .0005$, for additional vs. redundant information $\kappa = 1.00$, $p < .0005$, for gesture observation vs. gesture production $\kappa = 1.00$, $p < .0005$, for age group $\kappa = 1.00$, for comprehension measure $\kappa = 1.00$, $p < .0005$, and for significant effect reported $\kappa = .85$, $p < .0005$. In all cases, the codes of the original coder were used.

Intra-class correlations (ICC) were obtained to evaluate reliability of continuous data using an absolute agreement model. As only the first coder's scores were used in the final analyses, the single measures ICC is reported. For number of participants reported in the analysis, ICC = .94, $p < .0005$.

Estimation of Effect Sizes

The effect size used in the current analysis was Cohen's *d*. For each sample included, Cohen's *d* was calculated in one of a variety of ways, dependent on the information available. In all instances, Cohen's *d* was calculated in such a way that positive effect sizes corresponded to beneficial effects of gesture. Where possible, Cohen's *d* was calculated directly from t, F, β, or χ^2 values using formulae suggested by Cumming (2012), Lipsey and Wilson (2001), and Rosenthal (1984), using separate formulae for between- and within-subjects designs. In the event that these statistics were not available but relevant means and standard deviations were, Cohen's *d* was calculated by taking the difference between means and dividing this difference by the pooled standard deviation. In some instances, means and standard errors were given. In these cases, standard errors were transformed into standard deviations so that the pooled standard deviation could be calculated and applied to the abovementioned formula.

In the event that a between-subjects study provided the proportion of participants who improved on a measure of comprehension for one condition compared to another (e.g., the proportion of individuals who improved on a measure of comprehension after observing gestures compared to the proportion improved after observing no gestures), Cohen's *d* was calculated through the use of the arcsine transformation suggested by Lipsey and Wilson (2001). This was done by calculating the difference between the arcsine of the proportion for one condition and the arcsine of the proportion for a second condition. The arcsine method was chosen as it is more conservative than other methods, such as the probit transformation (Lipsey & Wilson, 2001), preventing an over-estimate of Cohen's *d*.

After Cohen's *d* had been successfully calculated for all samples, an adjustment recommended by Hedges (1981) was applied to calculate unbiased estimates of Cohen's *d*. Unbiased estimates of Cohen's *d* on average will not underestimate or overestimate the parameter, and such an adjustment is particularly important when sample sizes are small given the increased risk of overestimation (Cumming, 2012). Finally, estimates of standard error around unbiased Cohen's *d* were calculated separately for between- and within-subjects studies, using the formulae recommended by Lipsey and Wilson (2001).

Results

Analyses of Effect Sizes

Data were analyzed using Stata v. 15 (StataCorp, 2015). Cohen's *d* was successfully calculated from 83 unique samples from 64 studies, as some of the papers contained more than one study. In the event that a paper contained more than one study, samples were only included in the meta-analysis if they used participants that differed from those used in other reported studies. Thus, we regard each sample as independent of others from that study and the unit of analysis is therefore independent cohorts of participants rather than independent studies. The size of the effects was interpreted using the guidelines suggested by Cohen (1988), with $d = .2$ indicating a small effect size, $d = .5$ indicating a medium effect size, and $d = .8$ indicating a large effect size. Sample sizes ranged from 12 to 184 participants (M = 46.78, SD = 32.50), with a total number of 3883 participants represented in the overall meta-analysis. The effect sizes ranged from -1.14 to 2.90, and 80 of the 83 effect sizes were positive.

A random-effects model was used for all analyses given past reports of significant heterogeneity of effect sizes in the literature (Hostetter, 2011). Such models take systematic heterogeneity into consideration in the calculation of weights and hence yield more appropriate pooled estimates and variances of the pooled estimate than fixed-effects models, which assume only random variation between studies. A standard test of heterogeneity (Cochrane's test; DerSimonian & Laird, 1986) was also used to test for the presence of heterogeneity in our sample. In the current sample, Cochran's $Q = 265.60$, $p < .0005$. This statistically significant finding suggests that the variance across studies is due to more than just sampling error. Furthermore, I^2 was used as an indicator of the percentage of heterogeneity present between studies, with $I^2 = 25\%$ indicating a small amount of heterogeneity, $I^2 = 50\%$ indicating a medium amount of heterogeneity, and $I^2 = 75\%$ indicating a large amount of heterogeneity (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003).

A medium amount of heterogeneity remained in the current meta-analysis, with the analysis explaining 30.9% of the total between study variation (I^2 = 69.1%). That is, the unexplained between-study variance in effect size was greater than the explained variance in the included studies, which is unsurprising given the hypotheses of the current meta-analysis relate to different methodologies (e.g., observation vs. production of gesture; additional vs. redundant gesture etc.).

Using the random-effects model, the weighted mean effect size was .61 (95% CI = .50, .72)[1], and was found to be significantly greater than zero (z = 11.04, p < .0005). Such a result suggests that across studies, gestures have a beneficial, medium effect on comprehension. However, given the presence of between-study variance, stratified meta-analyses were undertaken to investigate when, across studies, gestures were beneficial.

The second research question, which investigated whether gesture is beneficial to comprehension under certain circumstances, was explored by undertaking five stratified meta-analyses. The first stratification explored whether observing different kinds of gestures benefits comprehension, including iconic gestures, metaphoric gestures, deictic gestures, beat gestures, or a mix of different kinds of gestures. Studies investigating the effect of gesture production were not included in this specific analysis, given participants are seldom told what kind of gestures to produce. That is, participants are typically told to produce any kind of gestures, not necessarily just iconic or deictic gestures for example. Thus, it is difficult to ascertain the benefit of producing specific kinds of gestures on comprehension from these studies.

Of the 67 samples included, 28 investigated the effect of observing iconic gestures on comprehension, 2 investigated the effect of observing metaphoric gestures

[1]Two outliers were identified (d = -1.07 and -1.14). To determine whether the outliers had a significant impact on the results, analyses were repeated excluding these two studies. As the results were unchanged in any substantive way (overall weighted mean = .63, overall I^2 = 67.6), we opted to report the original results. All results excluding outliers are available on request.

on comprehension, 16 investigated the effect of observing deictic gestures on comprehension, 2 investigated the effect of observing beat gestures on comprehension and 19 investigated the effect of observing a mixture of gestures on comprehension. A random-effects model was used for iconic gesture, metaphoric gesture, deictic gesture, beat gesture, and mixed gesture strata (see Table 2; note that overall pooled estimates will differ across the stratifications dependent on the studies included in each analysis).

For samples that investigated iconic gestures, the weighted mean effect size was significantly greater than zero, suggesting that observing iconic gestures has a medium beneficial effect on comprehension. For samples that investigated metaphoric gestures, the weighted mean effect size was significantly greater than zero, suggesting that observing metaphoric gestures has a large beneficial effect on comprehension. For samples that investigated deictic gestures, the weighted mean effect size was significantly greater than zero, suggesting that observing deictic gestures has a small beneficial effect on comprehension. However, for samples that investigated beat gestures, the weighted mean effect size was not significantly different from zero, suggesting that observing beat gestures has a non-significant effect on comprehension. For samples that investigated mixed gestures, the weighted mean effect size was significantly greater than zero, suggesting that observing a mix of different kinds of gestures has a small beneficial effect on comprehension.

While in the overall meta-analysis of the total between study variation 69% was systematic, stratifying by gesture type indicated 54.8% systematic variance within iconic gestures and 38.2% systematic variance within deictic gestures, indicating less heterogeneity within gesture types. Note that I^2 is influenced by the number of studies included within the analysis (Hippel, 2015), with I^2 often over or underestimating the amount of systematic variance when less than seven studies are included. Only two studies were included in the current meta-analysis that investigated the effect of beat gestures on comprehension, and only two studies investigated the effect of metaphoric gestures on comprehension. Consequently, the reported I^2 values for beat gestures and metaphoric gestures cannot be meaningfully interpreted.

Table 2

Cochran's Q and stratification test results for gesture type

	Heterogeneity			Stratified Pooled Estimates				
	Q	*p*-value	I^2	*M*	*CI* (95%)		*z*-score	*p*-value
Iconic	59.79	<.0005	54.8%	.66	.50	.82	8.05	<.0005
Metaphoric	6.48	.011	84.6%	1.41	.16	2.67	2.21	.027
Deictic	24.28	.061	38.2%	.43	.25	.61	4.76	<.0005
Beat	.04	.846	0.0%	.20	-.03	.43	1.68	.093
Mixed	82.66	<.0005	78.2%	.42	.17	.67	3.32	.001
Overall	199.02	<.0005	66.8%	.54	.43	.66	9.33	<.0005

Note. *M* refers to the pooled value of Cohen's *d* (effect size).

The second stratification investigated whether the observation of gestures that provide additional information or redundant information affected the degree to which gesture benefitted comprehension. While one study investigated whether a mix of additional and redundant gestures benefitted comprehension, the results are not reported here given less than two studies could be identified (Valentine, Pigott, & Rothstein, 2010) that investigated a mix of additional and redundant gestures. Of the 66 samples included, 8 investigated the effect of observing gestures that provided additional information beyond speech on comprehension, and 58 investigated the effect of observing redundant gestures on comprehension. A random-effects model was used for both additional and redundant information strata (see Table 3).

For additional information samples, the weighted mean effect size was significantly greater than zero, suggesting that observing gestures that provide additional information beyond speech has a medium beneficial effect on comprehension. For samples that investigated the effect of observing redundant gestures on comprehension, the weighted mean effect size was significantly greater than zero, suggesting that observing redundant gestures has a medium beneficial effect on comprehension. Stratifying by the information provided by an observed gesture indicated that within studies of gestures that provide additional information, there is 82.6% systematic variance whereas within redundant gestures there is 62.2% systematic variance.

Table 3

Cochran's Q and stratification test results for information provided through gesture

	Heterogeneity			Stratified Pooled Estimates				
	Q	*p*-value	I^2	*M*	*CI* (95%)		*z*-score	*p*-value
Additional	40.23	<.0005	82.6%	.75	.22	1.28	2.78	.006
Redundant	150.79	<.0005	62.2%	.52	.41	.64	8.95	<.0005
Overall	196.44	<.0005	66.9%	.55	.44	.67	9.29	<.0005

Note. *M* refers to the pooled value of Cohen's *d* (effect size).

The third stratification investigated whether the production and observation of gesture affected the degree to which gesture benefitted comprehension. Of the 83 samples included, 17 investigated the effect of producing gesture on comprehension and 66 investigated the effect of observing gesture on comprehension. A random-effects model was used for both gesture production and gesture observation strata (see

Table 4). For gesture production samples, the weighted mean effect size was significantly greater than zero, suggesting that the production of gestures has a large beneficial effect on comprehension. For gesture observation samples, the weighted mean effect size was significantly greater than zero, suggesting that the observation of gesture has a medium beneficial effect on comprehension. Stratifying by whether gesture was produced or observed suggested that within gesture production there was 72.1% systematic variance and within gesture observation there was 67.3% systematic variance.

Table 4

Cochran's Q and stratification test results for gesture production or observation

	Heterogeneity			Stratified Pooled Estimates				
	Q	*p*-value	I^2	*M*	*CI* (95%)		*z*-score	*p*-value
Production	57.44	<.0005	72.1%	.91	.64	1.19	6.53	<.0005
Observation	198.48	<.0005	67.3%	.54	.43	.66	9.18	<.0005
Overall	265.60	<.0005	69.1%	.61	.50	.72	11.04	<.0005

Note. *M* refers to the pooled value of Cohen's *d* (effect size).

The fourth stratification investigated whether gestures benefit comprehension in preschool children, primary school children, adolescents, and adults. Studies investigating the effect of gestures on comprehension in older adults were included in the analysis, but the results are not reported here as less than two studies were identified (Valentine, Pigott, & Rothstein, 2010). Of the 82 samples included, 18 investigated the effect of gesture on preschool children, 19 investigated the effect of gesture on primary school children, 2 investigated the effect of gesture on adolescents, and 43 investigated the effect of gesture on adults. A random-effects model was used for preschool children, primary school children, adolescent, and adult strata (see Table 5).

For samples that included preschool children, the weighted mean effect size was significantly greater than zero, suggesting that gestures had a medium beneficial effect on comprehension for preschool children. For samples that included primary school children, the weighted mean effect size was significantly greater than zero, suggesting that gestures had a medium beneficial effect on comprehension in primary school children. For samples that included adolescents, the weighted mean effect size was not significantly greater than zero, suggesting that gestures had a non-significant effect on comprehension in adolescents. Finally, for samples that included adults, the weighted mean effect size was significantly greater than zero, suggesting that gestures had a medium beneficial effect on comprehension in adults. Stratifying by age group suggested that within the preschool stratum there was 74.8% systematic variance, while within the primary school stratum there was 41.6% systematic variance, and 74.2% systematic variance within the adult stratum. Only two studies were included in the current meta-analysis that investigated the effect of gestures on adolescent comprehension and consequently the reported I^2 value for adolescents cannot be meaningfully interpreted.

Table 5

Cochran's Q and stratification test results for age group

	Heterogeneity			Stratified Pooled Estimates				
	Q	*p*-value	I^2	*M*	*CI* (95%)		*z*-score	*p*-value
Preschool	67.41	<.0005	74.8%	.73	.49	.97	5.95	<.0005
Primary	30.82	.030	41.6%	.58	.42	.75	6.91	<.0005
Adolescent	2.31	.128	56.8%	.68	-.07	1.41	1.79	.074
Adult	163.01	<.0005	74.2%	.57	.40	.74	6.61	<.0005
Overall	261.84	<.0005	69.4%	.61	.50	.72	10.83	<.0005

Note. *M* refers to the pooled value of Cohen's *d* (effect size).

The fifth and final stratification investigated whether gestures benefit comprehension when different kinds of measures of comprehension are used including free recall, use of open-ended or specific questions, use of multiple or forced-choice questions, or a mix of different kinds of measurement techniques. Of the 83 samples included, 15 used a free recall question to measure comprehension, 47 used open-ended specific questions, 17 used forced or multiple-choice questions, and 4 used a mixture of the abovementioned measurement techniques. A random-effects model was used for free recall, specific/open-ended, multiple/forced-choice, and mixed strata (see Table 6).

For samples that used free recall as a measure of comprehension, the weighted mean effect size was significantly greater than zero, suggesting that the gestures themselves had a large beneficial effect on comprehension when measured through free recall. For samples that measured comprehension using open-ended or specific questions, the weighted mean effect size was significantly greater than zero, suggesting that the gestures themselves had a medium beneficial effect on comprehension when measured through open-ended or specific questions. For samples that measured comprehension through using multiple or forced-choice questions, the weighted mean effect size was significantly greater than zero, suggesting that the gestures themselves had a small beneficial effect on comprehension when measured through multiple or forced-choice questions. For samples that measured comprehension using a mixture of the abovementioned methods, the weighted mean effect size was significantly greater than zero, suggesting that the gestures themselves had a large beneficial effect on comprehension when measured through a mixture of the abovementioned measures. Stratifying by the way that comprehension was measured indicated that there was 87.5% systematic variance within free recall, 39% systematic variance within specific questions, and 75.6% systematic variance within multiple or forced-choice. Given only four studies used a mixture of methods to measure comprehension, the reported I^2 value cannot be meaningfully interpreted.

Table 6

Cochran's Q and stratification test results for comprehension measure type

	Heterogeneity			Stratified Pooled Estimates				
	Q	*p*-value	I^2	*M*	*CI* (95%)		*z*-score	*p*-value
Free	112.38	<.0005	87.5%	.91	.54	1.28	4.83	<.0005
Specific	75.39	.004	39.0%	.55	.45	.65	10.53	<.0005
Forced	65.60	<.0005	75.6%	.47	.19	.76	3.26	.001
Mixture	3.27	.351	8.4%	.86	.56	1.16	5.67	<.0005
Overall	265.60	<.0005	69.1%	.61	.50	.72	11.04	<.0005

Note. *M* refers to the pooled value of Cohen's *d* (effect size).

Across the stratified meta-analyses reported above, the pooled effect sizes appear to vary substantially. Furthermore, heterogeneity was considerable. As a result, meta-regression analyses were undertaken to determine whether the differences in variation were significant, and any impact on the level of heterogeneity.

Moderator Analyses

Meta-regression analyses were run to investigate the third and final research question: which of the abovementioned factors moderate how beneficial gesture is to comprehension? The outcome variable was the unbiased effect size of each study, and the meta-regression model used random-effect weights. A total of six predictors were analyzed. Two predictors had two levels: group (observed vs. produced), and gesture content (additional vs. redundant). Two predictors had four levels: Age group (preschool, primary school, adolescent, adult), and comprehension measure type (free recall, open-ended/specific, multiple/forced-choice, mixed). The remaining predictor, gesture type, had five levels (iconic, metaphoric, deictic, beat, mixed). These variables were dummy coded with the reference category as the category of studies with the smallest effect size. For gesture type, the reference category was beat gestures. For age group, the reference category was adults. For comprehension measure type, the reference category was multiple/forced-choice. All results reported (see Table 7) involving multiple comparisons are Bonferroni adjusted for the number of comparisons.

Whether the kind of gesture observed was iconic, metaphoric, deictic, beat, or mixed did not significantly predict the size of the effect that gesture had on comprehension against a Bonferroni adjusted alpha of .013. Furthermore, whether samples used gestures that provided additional information above and beyond the content of speech, or gestures that provided redundant information, did not significantly predict the size of the effect that gesture had on comprehension. Samples that investigated the effect of producing gestures on comprehension found a larger effect of gesture than samples that investigated the effect of gesture observation on comprehension. However, whether a sample was composed of preschool children, primary school children, adolescents, or adults did not significantly predict the size of the effect that gesture had on comprehension against a Bonferroni adjusted alpha of .017. Whether comprehension was measured using free recall, open-ended or specific questions, multiple or forced-choice questions, or a mixture of questions did not significantly predict the effect that gesture had on comprehension against a Bonferroni adjusted alpha of .017. A medium amount of heterogeneity remained for all moderator variables, with the kind of gesture observed explaining 38.10% (I^2 = 61.90) of the total between study variation, information provided by gesture explaining 35.60% (I^2 = 64.40%), the observation or production of gesture explaining 31.49% (I^2= 66.40%), the age of participants explaining 31.49% (I^2 = 68.51%), and the way comprehension was measured explaining 32.48% (I^2 = 67.52%).

Table 7

Meta-regression test results for predicted moderator variables

Comparison	β	SE_{β}	*p*-value
Iconic vs. beat	.45	.33	.179
Metaphoric vs. beat	1.19	.50	.020
Deictic vs. beat	.22	.34	.516
Mixed vs. beat	.21	.34	.538
Additional vs. redundant	.22	.20	.279
Observation vs. Production	-.36	.16	.024*
Preschool vs. adults	.17	.17	.303
Primary school vs. adults	.06	.16	.730
Adolescents vs. adults	.10	.09	.817
Free recall vs. multiple/forced-choice	.38	.20	.068
Open-ended/specific vs. multiple/forced-choice	.11	.16	.493
Mixed vs. multiple/forced-choice	.42	.43	.334

Note. * denotes significant moderator at $p < .05$, or after Bonferroni adjustment

Tests for Publication Bias and Related Small Study Effects

Lastly, the data were examined for the potential for publication bias (i.e., the over-representation of positive or negative results) or related small study effects (i.e., the phenomenon whereby smaller studies tend to have larger effect sizes than larger studies). Given publication bias cannot adequately be measured through a single method, multiple approaches were employed (Rücker, Carpenter, & Schwarzer, 2011). First, a funnel plot of the effect size of each sample against its standard error was inspected (see Figure 2 for a funnel plot). In the absence of publication bias or related small study effects, the funnel plot should appear symmetrical, approximating the shape of a funnel. However, inspection of the funnel plot revealed a pattern of effects that were asymmetrically distributed, suggesting possible positive overestimation of the overall effect size.

Given the visual asymmetry of the funnel plot, the trim-and-fill method (Duval & Tweedie, 2000) was used to estimate the number of missing studies, and to impute the effect sizes of any identified missing studies to adjust the mean effect size as though there was no presence of publication bias (i.e., as though the funnel plot were symmetrical). The trim-and-fill method did not identify any studies that needed to be filled in the data, and thus no correction to the mean effect size was performed.

Additionally, Egger's test was conducted both on the overall and stratified analyses to further assess whether publication bias was present in the current meta-analysis. Egger's test was significant for some analyses, which may indicate the potential for publication bias or related small study effects. However, we do note that 15 Egger tests were performed, and the statistically significant findings should be interpreted in that context. Table 8 gives details of analyses with significant Egger's test results. A full list of Egger's test results including non-significant analyses is available from the authors on request.

Table 8

Significant Egger's Test Results

	β	$SE_β$	*p*-value
Iconic gestures	3.13	.98	.004
Redundant gestures	1.58	.67	.021
Observation of gesture	1.79	.68	.001
Production of gesture	3.51	.90	.001
Preschool children	3.30	.93	.003
Primary school children	2.25	.89	.022
Free recall	5.26	1.36	.002
Open-ended/specific	1.25	.58	.035
Overall	2.32	.57	<.0005

There was no evidence of publication bias on Egger's tests for strata that investigated additional gestures, deictic gestures, mixed gestures, adults, comprehension using multiple/forced-choice questions, or comprehension using a mixture of question types as a measure of comprehension (*p*'s > .05). As there were only two studies that investigated metaphoric gestures, beat gestures, and adolescents, estimates of publication bias could not be calculated. Overall, the evidence for publication bias or related small study effects in the current results is inconsistent but may represent a limitation of our findings given the possibility that our effects sizes are overestimated in some cases.

Although multiple databases were searched using multiple terms for the current meta-analysis, there is a small likelihood that not every paper addressing whether the observation or production of gesture benefits comprehension was found, leading to the possibility that some non-significant findings were missed. To establish the likelihood of this, a fail-safe N was calculated to determine the number of studies with non-significant results that would have to be located to reduce the mean effect size across all strata from medium to small. Using the formula outlined by Orwin (1983), a fail-safe N of 170 was calculated. That is, 170 papers with non-significant findings would have to be located to reduce the current medium mean effect size of .61 to a small effect size of .20. Therefore, it is unlikely that the results presented are the consequence of including only samples with significant findings.

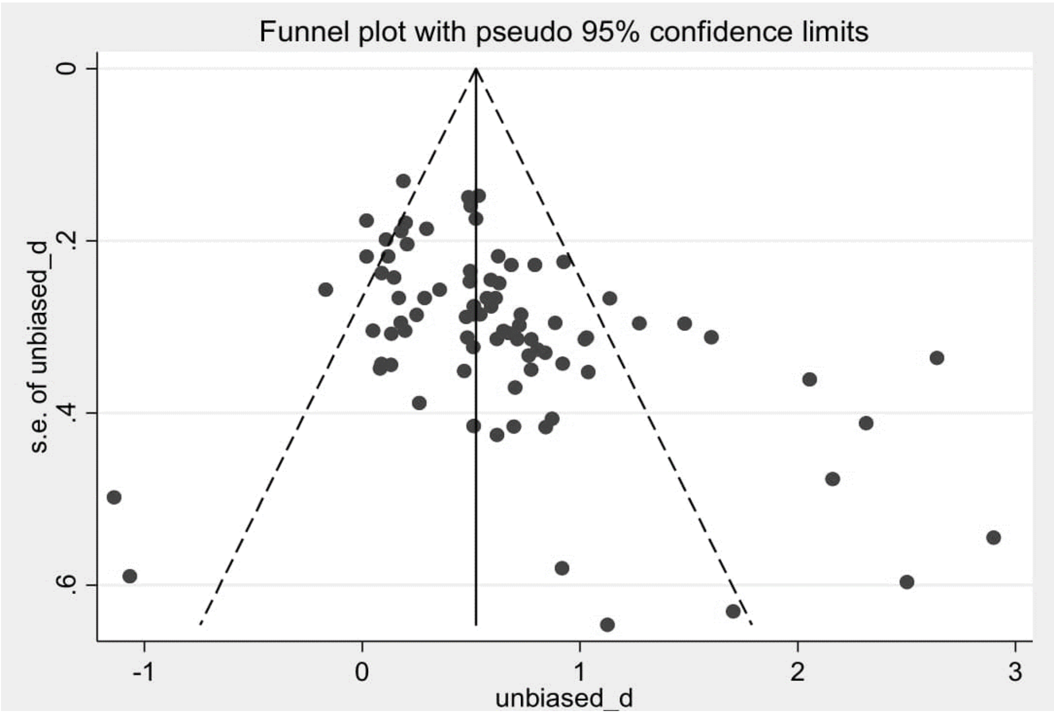

Figure 2. Funnel plot of the samples included in the meta-analysis

Discussion

When we speak, we gesture, and the gestures that accompany speech have the potential to aid comprehension (Littlejohn & Foss, 2010). In the current meta-analysis, we aimed to determine *whether* gestures benefit comprehension across studies, and if so, *when* gestures are the most beneficial through examining potential moderators not previously investigated through meta-analysis. Factors of interest included the kinds of gestures used, whether the gesture provides additional or redundant information to the accompanying speech, whether gesture is produced or observed by a learner, the age of the participants, and finally, the way that studies measured comprehension. The results have implications for gesture research, through highlighting when gestures are most beneficial to comprehension.

The Effect of Gesture on Comprehension Across Studies

As expected, across all studies included in the current meta-analysis, there is evidence that gestures do indeed significantly benefit comprehension of a spoken message. The mean unbiased effect size itself, .61, was significantly different from zero and can be interpreted as a medium effect size. Such a result is in line with a prior meta-analysis conducted by Hostetter (2011), who obtained a mean unbiased effect size of .61 across 63 samples. Across the 83 samples included in the current meta-analysis, 96% of the samples reported a positive effect size, with only 4% of samples reporting a negative effect size. Despite the overwhelming majority reporting positive effect sizes, only 68% of samples reported a significant difference between gesture and no gesture conditions. Regardless of this variability, it appears that across studies gestures indeed have a significant, beneficial effect on comprehension.

The remaining questions investigated by the current meta-analysis explored a variety of factors that could potentially moderate the beneficial effect that gesture has on comprehension. The first discussed here is whether different kinds of gestures are beneficial to comprehension when observed.

The Effect of Observing Different Kinds of Gestures on Comprehension

The results obtained indicated that iconic gestures, metaphoric gestures, deictic gestures, and the observation of a mixture of gestures benefit comprehension. Such findings are in line with past research that has shown that iconic gestures

(Macoun & Sweller, 2016), metaphoric gestures (Repetto et al., 2017), and deictic gestures (Pi et al., 2017) benefit comprehension. Contrary to the findings of Igualada et al. (2017) in which observing beat gestures significantly benefitted comprehension the current meta-analysis found that observing beat gestures was no more beneficial than observing no gestures at all. Such a result is, however, in line with the findings of Gluhareva and Prieto (2017), which suggested that beat gestures are beneficial when the accompanying task is difficult, but not when it is simple. While it may be that beat gestures are primarily beneficial when a task is difficult to understand, such as in the study by Gluhareva and Prieto, given the current meta-analysis only had data available for two studies investigating beat gestures and comprehension, it is possible that the analysis was underpowered. It therefore remains unclear across studies whether or not observing beat gestures benefits comprehension.

There is a small amount of additional research regarding beat gestures not included in the current meta-analysis. For example, research conducted by the neuroscience community has shown observing beat gestures can benefit processing of syntactic structure (Holle et al., 2012), and suggested that such observation can direct attention to important components of a spoken message (Biau & Soto-Faraco, 2013). Furthermore, recent behavioral research has suggested that observing beat gestures can enhance narrative structure (Vilà-Giménez, Igualada, & Prieto, 2019) and second language learning (Kushch, Igualada, & Prieto, 2018). However, neither study met inclusion criteria for the current meta-analysis. The studies were conducted and published after the current meta-analysis was completed, and manipulations affecting the prominence of speech and gesture in Kushch et al.'s (2018) study could confound results. Nonetheless, it is therefore likely that with additional future behavioral studies conducted with beat gestures, a beneficial effect of observing beat gestures across studies may be found.

When iconic or metaphoric gestures are observed, it is thought that the content of accompanying speech binds with the semantic information provided by the gesture (Straube et al., 2008). It has been argued that such gestures benefit comprehension as a result of this process (Straube et al., 2008), and as a result it has been suggested that perhaps the more semantically related a gesture is to the accompanying speech, the more beneficial it is (Dargue & Sweller, 2018b). However, it has also been argued that gestures can also benefit comprehension through capturing attention. The fact that deictic gestures were found in the current meta-analysis to be beneficial to comprehension supports the idea that gestures do not have to be semantically related to the content of speech to benefit learning. However, it may be that the semantic content of iconic and metaphoric gestures makes them more beneficial to learning than deictic gestures or beat gestures, which by definition are not semantically related to the content of speech. However, the results from the current meta-regression suggest that iconic and metaphoric gestures are no more beneficial to comprehension than deictic or beat gestures. It is notable that despite beat gestures being shown to be no more beneficial than no gestures in the stratified analysis, the pooled effect sizes for iconic, metaphoric, and deictic gestures did not differ significantly from that of beat gestures as shown through the meta-regression analysis. It is therefore possible that the reason beat gestures were not shown to be beneficial to comprehension compared to no gestures overall was due to the analysis being underpowered.

Although these findings might suggest that gestures may benefit learning through capturing one’s attention, past research has found that some iconic gestures benefit comprehension while others do not (Dargue & Sweller, 2018b). Therefore, it is possible that there are a variety of mechanisms underlying why particular gestures are beneficial, an area to be further explored with future research.

The Effect of Information Provided Through Gesture on Comprehension

It has previously been suggested that gestures may benefit comprehension to a greater extent when the gestures provide additional information that is not present in speech (Hostetter, 2011). In line with this idea, the results of the current meta-analysis found that gestures which provide additional information to speech benefit comprehension across studies, compared to observing no gestures. However, even redundant gestures were found to benefit comprehension significantly across studies.

In the current meta-analysis, samples that investigated the effect of gestures that provided additional information to speech were no more beneficial than gestures that provided redundant information. Such a result contradicts the finding reported by Hostetter (2011) and suggests that perhaps there are different mechanisms underlying the beneficial effects of gestures that provide additional information to speech, and gestures that are redundant with speech. It might be, for example, that gestures that provide additional information are more beneficial only in certain circumstances, such as when the gestures serve to disambiguate a poor-quality spoken message. This idea is in line with a study by Holle et al. (2010) who investigated whether observing iconic gestures that provide additional information benefits speech comprehension, particularly when the speech itself is difficult to comprehend. The 2010 study demonstrated that although observing the gestures was more beneficial to speech comprehension than observing no gestures generally, the additional gestures were particularly beneficial when speech was accompanied by a moderate degree of multi-speaker “babble sounds.”

The Effect of Gesture Observation Compared to Gesture Production on Comprehension

In line with previous research, the results of the current meta-analysis suggest that both observing gestures (Dargue & Sweller, 2018b) and producing gestures (Chu & Kita, 2011) benefits comprehension. However, past research suggests that the production of gesture is more beneficial than the observation of gesture (Goldin-Meadow et al., 2012). In line with this argument, samples that investigated the production of gesture on comprehension had significantly larger effect sizes than samples that investigated the observation of gesture on comprehension. Such a result is supported by the idea that the production of gesture by a learner might increase comprehension through reducing the cognitive load placed on working memory, allowing for additional cognitive resources to be allocated to the task at hand (Cook, Yip, & Goldin-Meadow, 2012).

Although the results of the current meta-analysis suggested that the observation of gesture by a learner still appeared to be more beneficial to comprehension than observing no gestures at all, the fact that the abovementioned results highlight an additional benefit of gesture production has implications for future research in this field. For example, there are numerous studies that have investigated the effect of observing different kinds of gesture on comprehension, such as iconic or deictic gestures. However, the effect of producing different kinds of gestures has seldom been investigated in the literature, and it may be that producing certain kinds of gestures benefits comprehension to a greater extent than others.

The Effect of Gesture on Comprehension Across Different Age Groups

It has been suggested that studies investigating the benefits of gesture in children found significantly greater effect sizes than studies investigating the benefits of gesture in adults, potentially because verbal skills are not yet fully developed in children (Hostetter, 2011). However, the current meta-analysis extended these findings, investigating whether gestures are beneficial separately for preschool children, primary school children, adolescents, and adults. Furthermore, this meta-analysis investigated whether age moderates the effect that gesture has on comprehension across these four age groups. Indeed, the current meta-analysis found evidence that gesture significantly benefitted preschool children, primary school children, and adults across studies. Unexpectedly, no beneficial effect of gesture was found for adolescents. However, given only two adolescent samples were identified, it may be that this particular analysis was underpowered. To better ascertain whether gesture benefits comprehension in adolescents, further research is required.

Regardless, no significant difference was found between the effect sizes obtained for preschool, primary school, and adolescent studies compared to adult studies in the current meta-analysis. This result suggests that studies yielded similar effect sizes across the different age categories, contrary to the findings of Hostetter (2011). Church et al. (2000) found behavioral evidence that the beneficial impact of observing gesture follows a U-shaped curve when comparing children aged 7 to 8 years, children aged 9 to 10 years, and adults. The current meta-analysis did not find evidence for such a pattern at this time, given the lack of a significant difference between age-groups. However, given the limited adolescent research available it is difficult to definitively conclude the presence or lack of a U-shape curve at this time. Based on available current research, it appears that gestures, on average, are equally beneficial across preschool aged children, primary school aged children, adolescents, and adults, but further research with adolescents is required.

The Effect of Gesture on Different Measures of Comprehension

Previous studies found that gestures only benefit comprehension for the speech that directly accompanies a gesture (Dargue & Sweller, 2018a). However, the current meta-analysis showed that gestures benefitted comprehension regardless of how it was measured. That is, through free recall, open-ended or specific questions, multiple or forced-choice questions, or a mixture of these methods. Given the observation of gesture has been found, averaged across studies, to benefit free recall of a verbal message, it appears that gestures do have the ability to benefit comprehension of speech that has not necessarily been directly accompanied by a gesture. However, the variability of findings in the literature suggests that gestures only benefit free recall of a message in some instances.

The results of the current meta-analysis also found no significant differences between the effect sizes obtained for studies that measured comprehension through free recall, open-ended or specific questions, multiple or forced-choice questions, or a mixture of these methods. Such a finding has implications for further research in gesture and comprehension, highlighting the different ways that comprehension can be successfully measured.

Limitations

One possible limitation of the current meta-analysis is that only published, peer-reviewed studies were considered for inclusion. Although this decision reduces the potential for bias associated with poor study quality (i.e., studies that have gone through peer review are more likely to be of higher quality to unpublished papers), it is possible that unpublished papers have smaller effect sizes on average or find no beneficial effects of gesture on comprehension. In addition, there was some evidence to indicate possible publication bias or related small study effects, potentially leading to some overestimation of effect size.

Moreover, despite the investigation of new possible moderating variables, there is still a lot of variance left unexplained surrounding when gestures are most beneficial. Future research is needed to investigate further potential moderators of the effect of gesture on comprehension. For example, it may be that within certain subtypes of gestures, such as iconic gestures, some gestures may be more beneficial to comprehension than others (Dargue & Sweller, 2018b).

In addition, it could be that more heterogeneity could be explained through investigating interactions between different combinations of the moderator variables described above. For example, perhaps the beneficial effect of certain gestures, such as those providing additional information vs. those providing redundant information, differs depending on how comprehension is measured. However, currently there are not enough studies to obtain enough statistical power be able to investigate interactions across each combination of two or more moderator variables.

Another possible moderating variable that might explain further heterogeneity could be the specific cognitive process being investigated by a given study. The effect of observing or producing gesture on comprehension of spoken messages has been investigated through a large variety of cognitive tasks, and as such it would be interesting to conduct a stratified meta-analysis on the different kinds of tasks used both in gesture observation and gesture production research. However, the tasks used often differ within a particular cognitive process. For example, studies that have investigated the benefit of gesture on problem solving differ from one another in terms of the specific problem-solving task used. For example, some studies use Tower of Hanoi (Garber & Goldin-Meadow, 2002), mathematical equations (Perry, Breckinridge Church, & Goldin-Meadow, 1988), counting (Alibali & DiRusso, 1999), or Piagetian conservation tasks (Ping & Goldin-Meadow, 2008) as a measure of the impact that gesture has on problem solving ability.

Furthermore, the tasks used tend to differ between studies that look at the effect of observing gesture on comprehension, and studies that look at the effect of producing gesture on comprehension. This is such that gesture observation studies tend to investigate a participant's understanding of aspects of a verbal message produced by the speaker making the gestures (Dargue & Sweller, 2018a), while gesture production studies tend to involve the participant producing gestures themselves while completing a task, such as a mathematical equation (Perry et al., 1988). Given the variation in the tasks or methodologies used between studies, it is currently difficult to group the studies available into the specific cognitive processes being investigated. More research is necessary using similar tasks when investigating a particular cognitive process, such as problem solving, to determine how much heterogeneity is explained by the specific kind of cognitive process being investigated.

Conclusions

In summary, the current meta-analysis aimed to better establish an answer to the following questions: are gestures beneficial to comprehension of speech, and if so, when? Our results suggest that on average, gestures are beneficial to comprehension whether they are observed or produced. However, the effects found for gesture production studies (whereby a learner produces gesture while completing a task) were significantly larger than those found for gesture observation studies (whereby a learner watches another individual gesturing while they listen). Such a finding has implications for further research in the field of gesture and learning – it appears that gestures are the most beneficial when a learner produces gesture themselves, and it is yet to be understood *which* gestures are the most beneficial to produce.

This meta-analysis investigated which kinds of gestures are the most beneficial to observe. Surprisingly, no one kind of gesture was significantly more beneficial to observe than another. Furthermore, gestures that provided additional information, although beneficial, were no more beneficial than gestures that were redundant with speech. While it was previously shown that children benefitted from gesture to a greater extent than adults, the current findings found gesture to be relatively equally beneficial across a variety of age groups, including in preschool aged children, primary school aged children, and adults. There also appeared to be no difference across studies as a result of how comprehension was measured.

In conclusion, it appears that gestures *can* benefit comprehension, and the current meta-analysis sheds light on the factors that surround *when* gestures are beneficial, extending previous findings. The current findings have theoretical and practical implications for gesture-related research and provide new avenues of research. Through better understanding when gestures benefit comprehension, gesture can be more effectively implemented as a tool for teaching strategies and interventions alike.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Alibali, M. W., & Dirusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*(1), 37–56. https://doi.org/10.1016/S0885-2014(99)80017-3

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, *89*(1), 183–193. https://doi.org/10.1037/0022-0663.89.1.183

*Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144. https://doi.org/10.1177/0956797611417722

*Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*(1), 92–103. https://doi.org/10.1016/j.jecp.2013.12.008

*Austin, E. E., & Sweller, N. (2017). Getting to the elephants: Gesture and preschoolers' comprehension of route direction information. *Journal of Experimental Child Psychology*, *163,* 1–14. https://doi.org/10.1016/j.jecp.2017.05.016

*Austin, E. E., Sweller, N., & Van Bergen, P. (2018). Pointing the way forward: Gesture and adults' recall of route direction information. *Journal of Experimental Psychology: Applied*, *24*(4), 490-508. http://dx.doi.org/10.1037/xap0000168

*Beattie, G., & Shovelton, H. (1999). Do iconic handgestures really contribute anything to the semantic information conveyed by speech?An experimental investigation. *Semiotica*, *123*, 1–30. https://doi.org/10.1515/semi.1999

*Beaudoin-Ryan, L., & Goldin-Meadow, S. (2014). Teaching moral reasoning through gesture. *Developmental Science*, *17*(6), 984–990. https://doi.org/10.1111/desc.12180

*Berger, K. W., & Popelka, G. R. (1971). Extra-facial gestures in relation to speechreading. *Journal of Communication Disorders*, *3*(4), 302–308. https://doi.org/10.1016/0021-9924(71)90036-0

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143–152. https://doi.org/10.1016/j.bandl.2012.10.008

Bohn, M., Call, J., & Tomasello, M. (2016). The role of past interactions in great apes' communication about absent entities. *Journal of Comparative Psychology*, *130*(4), 351–357. https://doi.org/10.1037/com0000042

Breckinridge Church, R., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, *7*(2), 137–158. https://doi.org/10.1075/gest.7.2.02bre

*Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539

*Bucciarelli, M., Mackiewicz, R., Khemlani, S. S., & Johnson-Laird, P. N. (2016). Children's creation of algorithms: Simulations and gestures. *Journal of Cognitive Psychology*, *28*(3), 297–318. https://doi.org/10.1080/20445911.2015.1134541

*Cameron, H., & Xu, X. (2011). Representational Gesture, Pointing Gesture, and Memory Recall of Preschool Children. *Journal of Nonverbal Behavior*, *35*(2), 155–171. https://doi.org/10.1007/s10919-010-0101-2

*Carlson, C., Jacobs, S. A., Perry, M., & Breckinridge Church, R. (2014). The effect of gestured instruction on the learning of physical causality problems. *Gesture*, *14*(1), 26–45. https://doi.org/10.1075/gest.14.1.02car

*Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

*Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism*, *7*(4), 303–319. https://doi.org/10.1080/13670050408667815

*Church, R. B., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, *7*(2), 137-158. doi:10.1075/gest.7.2.02bre

Church, R. B., Kelly, S. D., & Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *Journal of Nonverbal Behavior*, *24*(2), 151–174. https://doi.org/10.1023/A:1006610013873

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Colletta, J. M., Pellenq, C., & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication*, *52*, 565–576. https://doi.org/10.1016/j.specom.2010.02.009

*Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005

*Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097

*Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J., & Popescu, V. (2017). Hand gesture and mathematics learning: Lessons from an avatar. *Cognitive Science, 41*(2), 518–535. https://doi.org/10.1111/cogs.12344

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

*Cook, S. W., Yip, T. K. Y., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*(4), 465–475. https://doi.org/10.1016/j.jml.2010.07.002

Cumming, G. (2012). *Understanding the new statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York, NY: Routledge.

*Dahl, T. I., & Ludvigsen, S. (2014). How I see what you’re saying: The role of gestures in native and foreign language listening comprehension. *Modern Language Journal*, *98*(3), 813–833. https://doi.org/10.1111/modl.12124

*Dargue, N., & Sweller, N. (2018a). Donald Duck’s garden: The effects of observing iconic reinforcing and contradictory gestures on narrative comprehension. *Journal of Experimental Child Psychology*, *175*, 96–107. https://doi.org/10.1016/j.jecp.2018.06.004

*Dargue, N., & Sweller, N. (2018b). Not all gestures are created equal: The effects of typical and atypical iconic gestures on narrative comprehension. *Journal of Nonverbal Behavior, 42*, 327-345. https://doi.org/10.1007/s10919-018-0278-3

*De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*(1), 173–179. https://doi.org/10.1016/j.actpsy.2013.05.013

*De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2014). Words in action: Using gestures to improve verb learning in primary school children. *Gesture, 14*(1), 46-69. doi:10.1075/gest.14.1.03noo

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

*Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 445–454. https://doi.org/10.1518/hfes.45.3.445.27258

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

*Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18*(2), 185–205. https://doi.org/10.1080/09541440540000158

Francaviglia, M., & Servidio, R. (2011). Gesture as a cognitive support to solve mathematical problems. *Psychology*, *2*(2), 91–97. https://doi.org/10.4236/psych.2011.22015

Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, *26*(6), 817–831. https://doi.org/10.1016/S0364-0213(02)00087-3

*Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609–631. https://doi.org/10.1177/1362168816651463

Goldin-Meadow, S., Cook, W. S., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about maths. *Psychological Science, 20*(3), 267–272. https://doi.org/10.1111/j.1467-9280.2009.02297.x

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, *15*(6), 876–884. https://doi.org/10.1111/j.1467-7687.2012.01185.x

*Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition*, *144*, 14–28. https://doi.org/10.1016/j.cognition.2015.07.008

Gunter, T. C., & Weinbrenner, J. E. (2017). When to take a gesture seriously: On how we use and prioritize communicative cues. *Journal of Cognitive Neuroscience*, *29*(8), 1355-1367. https://doi.org/10.1162/jocn_a_01125

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107

*Heidari, K. (2015). An investigation into the role of gesture in enhancing children's vocabulary command. *International Journal of Early Years Education*, *23*(4), 382–393. https://doi.org/10.1080/09669760.2015.1074556

Higgins, J.P.T. & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1). https://doi.org/10.1186/s12874-015-0024-z

Holle, H., & Gunter, T.C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175-1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*(MAR). https://doi.org/10.3389/fpsyg.2012.00074

Holle, H., Obleser, J., Rueschemeyer, S., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, *49*(1), 875-884. https://doi.org/10.1016/j.neuroimage.2009.08.058

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

*Hupp, J. M., & Gingras, M. C. (2016). The role of gesture meaningfulness in word learning. *Gesture, 15*(3), 340–356. https://doi.org/10.1075/gest.15.3.04hup

*Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121. https://doi.org/10.1016/j.jml.2017.05.004

*Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *Journal of Experimental Child Psychology*, *156*, 99–112. https://doi.org/10.1016/j.jecp.2016.11.017

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

*Kang, S., Hallman, G. L., Son, L. K., & Black, J. B. (2013). The different benefits from different gestures in understanding a concept. *Journal of Science Education and Technology*, *22*(6), 825–837. https://doi.org/10.1007/s10956-012-9433-5

Kang, S., & Tversky, B. (2016). From hands to minds: Gestures promote understanding. *Cognitive Research: Principles and Implications*, *1*(4), 4, 1-15. https://doi.org/10.1186/s41235-016-0004-9

*Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, *28*(2), 325–349. https://doi.org/10.1017/S0305000901004664

*Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*(4), 577–592. https://doi.org/10.1006/jmla.1999.2634

*Kelly, S. D., & Church, R. B. (1998). A comparison between children’s and adults’ ability to detect conceptual information conveyed through representational gestures. *Child Development*, *69*(1), 85–93. https://doi.org/10.1111/j.1467-8624.1998.tb06135.x

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*(1), 253–260. https://doi.org/10.1016/S0093-934X(03)00335-3

*Kelly, S. D., & Lee, A. L. (2011). Language and cognitive processes when actions speak too much louder than words : Hand gestures disrupt word learning when phonetic demands are high demands are high. *Language and Cognitive Processes*, *27*, 37–41. https://doi.org/10.1080/01690965.2011.581125org/10.1080/01690965.2011.581125

*Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes, 24*(2), 313–334. https://doi.org/10.1080/01690960802365567org/10.1080/01690960802365567

Kendon, A. (1988). How gestures can become like words. In F. Poyatos (Ed.), *Crosscultural perspectives in nonverbal communication* (pp. 131-141). Toronto, CA: Hogrefe.

Kendon, A. (1994). Research on language and social interaction communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, *27*, 223–237. https://doi.org/10.1207/s15327973rlsi2703

*Kirk, E., & Lewis, C. (2017). Gesture facilitates children's creative thinking. *Psychological Science*, *28*(2), 225–232. https://doi.org/10.1177/0956797616679183

Kita, S. (2000). How representational gestures help speaking. In David McNeill (Eds.), *Language and gesture* (pp. 162–185). Cambridge, UK: Cambridge University Press.

*Koumoutsakis, T., Church, R. B., Alibali, M. W., Singer, M., & Ayman-Nolley, S. (2016). Gesture in instruction: Evidence from live and video Lessons. *Journal of Nonverbal Behavior*, *40*(4), 301–315. https://doi.org/10.1007/s10919-016-0234-z

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005

*Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552. https://doi.org/10.1006/jesp.1995.1024

*Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*(5), 743–754. https://doi.org/10.1037/0022-3514.61.5.743

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, *33*(8), 992-1004. https://doi.org/10.1080/23273798.2018.1435894

*Lajevardi, N., Narang, N. S., Marcus, N., & Ayres, P. (2017). Can mimicking gestures facilitate learning from instructional animations and static graphics? *Computers and Education*, *110*, 64–76. https://doi.org/10.1016/j.compedu.2017.03.010

*Lickiss, K. P., & Wellens’, A. R. (1978). Effects of visual accessibility and hand restraint on fluency of gesticulator and effectiveness of message. *Perceptual and Motor Skills*, *46*(3), 925. https://doi.org/10.2466/pms.1978.46.3.925

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis (applied social research methods)*. Thousand Oaks, CA: SAGE Publications, Inc.

Littlejohn, S. W., & Foss, K. A. (2010). *Theories of human communication* (10th ed.). Long Grove, IL: Waveland Press Inc.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, *5*(4), 196–211. https://doi.org/10.1111/j.1751-228X.2011.01129.x

*Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical Teacher*, *36*(7), 596–601. https://doi.org/10.3109/0142159X.2014.899684

*Macoun, A., & Sweller, N. (2016). Listening and watching: The effects of observing gesture on preschoolers' narrative comprehension. *Cognitive Development*, *40,* 68–81. https://doi.org/10.1016/j.cogdev.2016.08.005

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal*, *86*(2), 192–203. https://doi.org/10.1111/1540-4781.00144

*McNeil, N., Alibali, M., & Evans, J. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*(2), 131–150. https://doi.org/10.1023/A:1006657929803

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: The University of Chicago Press.

McNeill, D. (2000). *Language and gesture*. Cambridge, UK: Cambridge University Press.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336-341. doi:10.1016/j.ijsu.2010.02.007

Namy, L. L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science*, *11*(6), 841–846. https://doi.org/10.1111/j.1467-7687.2008.00732.x

Namy, L. L., Vallas, R., & Knight-Schwarz, J. (2008). Linking parent input and child receptivity to symbolic gestures. *Gesture*, *8*(3), 302–324. https://doi.org/10.1075/gest.8.3.03nam

Orwin, R. G. (1983). A failsafe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157. https://doi.org/10.2307/1164923

*Ouwehand, K., van Gog, T., & Paas, F. (2015). Effects of gestures on older adults' learning from video-based models. *Applied Cognitive Psychology*, *29*(1), 115–128. https://doi.org/10.1002/acp.3097

Pashek, G. V, & DiVenere, E. (2006). Auditory comprehension in alzheimer disease: Influences of gesture and speech rate. *Journal of Medical Speech-Language Pathology*,*14*(3), 143-155. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc5&NEWS=N&AN=2006-11708-003

*Perry, M., Berch, D., & Singleton, J. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, *6*, 213–235. https://doi.org/10.3868/s050-004-015-0003-8

Perry, M., Breckinridge Church, R., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*(4), 359–400. https://doi.org/10.1016/0885-2014(88)90021-4

*Pi, Z., Hong, J., & Yang, J. (2017). Effects of the instructor's pointing gestures on learning performance in video lectures. *British Journal of Educational Technology*, *48*(4), 1020–1029. https://doi.org/10.1111/bjet.12471

*Pine, K. J., Knott, T., & Fletcher, B. C. (2010). Quand faire des gestes permet de mieux apprendre. *Enfance*, *2010*(3), 355–368. https://doi.org/10.4074/S0013754510003101

*Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, *44*(5), 1277–1287. https://doi.org/10.1037/0012-1649.44.5.1277

*Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02136

*Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, *5*(3), 172–183. https://doi.org/10.1007/BF00986134

*Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*, *5*(1), 54–62. https://doi.org/10.1111/j.1468-2958.1978.tb00622.x

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: SAGE Publications, Inc.

Rothi, L. J. G., Heilman, K. M., & Watson, R. T. (1985). Pantomime comprehension and ideomotor apraxia. *Journal of Neurology Neurosurgery and Psychiatry*, *48*(3), 207–210. https://doi.org/10.1136/jnnp.48.3.207

*Rowbotham, S. J., Holler, J., Wearden, A., & Lloyd, D. M. (2016). I see how you feel: Recipients obtain additional information from speakers' gestures about pain. *Patient Education and Counseling*, *99*(8), 1333–1342. https://doi.org/10.1016/j.pec.2016.03.007

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

*Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology*, *38*(2), 109–117. https://doi.org/10.1016/j.cedpsych.2012.12.001

Rücker, G., Carpenter, J., & Schwarzer, G. (2011). Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal*, *53*(2), 351-368. https://doi.org/10.1002/bimj.201000151

*Sekine, K., & Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychologica*, *179*, 89–95. https://doi.org/10.1016/j.actpsy.2017.07.009

Singh, J. (2013). Critical appraisal skills programme. *Journal of Pharmacology and Pharmacotherapeutics*, *4*(1), 76. https://doi.org/10.4103/0976-500X.107697

Stanfield, C., Williamson, R., & Özçalişkan, Ş. (2014). How early do children understand gesture-speech combinations with iconic gestures? *Journal of Child Language*, *41*(2), 462–471. https://doi.org/10.1017/S0305000913000019

*Stieff, M., Lira, M. E., & Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cognition and Instruction*, *34*(2), 80–99. https://doi.org/10.1080/07370008.2016.1145122

*Straube, B., Green, A., Weis, S., Chatterjee, A., & Kircher, T. (2008). Memory effects of speech and gesture binding: Cortical and hippocampal activation in relation to subsequent memory performance. *Journal of Cognitive Neuroscience*, *21*(4), 821–836. Retrieved from http://repository.upenn.edu/neuroethics_pubs/56

*Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need?: A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/1076998609346961

*Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187-204. https://doi.org/10.1016/S0361-476X(02)00007-3

*Van Wermeskerken, M., Fijan, N., Eielts, C., & Pouw, W. T. J. L. (2016). Observation of depictive versus tracing gestures selectively aids verbal versus visual–spatial learning in primary school children. *Applied Cognitive Psychology*, *30*(5), 806–814. https://doi.org/10.1002/acp.3256

Viher, P. V., Stegmayer, K., Kubicki, M., Karmacharya, S., Lyall, A. E., Federspiel, A., … Walther, S. (2018). The cortical signature of impaired gesturing: Findings from schizophrenia. *NeuroImage: Clinical*, *17*, 213–221. https://doi.org/10.1016/j.nicl.2017.10.017

Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology*, *55*(2), 250–262. https://doi.org/10.1037/dev0000604

*Wakefield, E. M., & James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Developmental Psychology*, *51*(8), 1105–1114. https://doi.org/10.1037/a0039471

Wang, X. L., Bernas, R., & Eberhard, P. (2004). Engaging ADHD students in tasks with hand gestures: A pedagogical possibility for teachers. *Educational Studies*, *30*(3), 217–229. https://doi.org/10.1080/0305569042000224189

Wong, M. K. Y., & So, W. C. (2018). Absence of delay in spontaneous use of gestures in spoken narratives among children with Autism Spectrum Disorders. *Research in Developmental Disabilities*, *72*, 128–139. https://doi.org/10.1016/j.ridd.2017.11.004

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*(6), 654–667. https://doi.org/10.1111/j.1469-8986.2005.00356.x