Educational Psychology Review (2024) 36:111
https://doi.org/10.1007/s10648-024-09948-0

INTERVENTION STUDY

From Hands to Mind: How Gesture, Emotional Valence, and Individual Differences Impact Narrative Recall

Kavya Thakore[1] · Trisha Das[1] · Shamma Jahan[1] · Naomi Sweller[1]

Accepted: 4 September 2024

Abstract

Narrative recall and comprehension are important lifelong skills. While gesturing may improve recall by alleviating cognitive load, it may be differentially beneficial, depending on task and individual characteristics. While research on gesture's effects on a variety of task modalities is burgeoning, effects on recall of narratives read aloud are under-examined. Further, emotional valence and individual differences in verbal memory may affect recall, through effects on task difficulty. If gesturing lightens cognitive load, it may be more beneficial for harder tasks, namely for narratives lacking emotional content and for individuals with poorer verbal memory. Across two studies, impacts of gesture production, emotional valence, and individual differences on narrative recall were evaluated. In Study 1, participants ($N=100$) read aloud three emotive narratives (positive, negative, neutral) while either instructed to gesture or receiving no gesture instructions. Gesture production hindered recall, particularly for those with higher verbal memory. Emotion benefited recall, with enhanced recall of the negative narrative and impaired recall for the neutral narrative. In Study 2, following a measure of individual propensity to gesture, participants ($N=98$) similarly read aloud three emotive narratives. Instructions to gesture hindered recall for participants with a lower propensity to gesture, and emotional narratives again saw enhanced recall relative to the neutral narrative. Propensity to gesture and verbal memory were positively associated with narrative comprehension. Results suggest instructions to produce gestures may for some individuals hinder recall for self-guided learners when studying written texts, while emotional content benefits recall.

Keywords Gesture · Emotion valence · Narrative recall · Verbal memory · Propensity to gesture

✉ Naomi Sweller
naomi.sweller@mq.edu.au

[1] School of Psychological Sciences, Macquarie University, Sydney, NSW 2109, Australia

Published online: 24 September 2024

Narrative recall and comprehension involve remembering and extracting meaning from a text (Clarke et al., 2010). Beyond meaning extraction, narrative recall requires the recaller to organize the information in a cohesive manner to allow for an accurate and understandable recount of events (Koskinen et al., 1988). Current theories of text comprehension note the inferential processes required for narrative recall, including incorporating the reader's prior knowledge and strategies to construct meaning (see Kendeou et al., 2016, for a review).

Comprehension is positively linked to academic achievement as well as life skills such as critical thinking, social skills, and perspective taking (Bellaera et al., 2016; García-Madruga et al., 2014; Kidd & Castano, 2013; Sparapani et al., 2018). Examination of the factors that influence recall and comprehension will, therefore, allow for such factors to be targeted when seeking to improve learning.

Several factors have been studied to determine their impact on recall and comprehension. Firstly, research has shown gestures can have a positive effect on recall (Cook et al., 2010; Dargue et al., 2019). Gestures are hand movements that accompany speech and allow for communication of information in addition to, or in exchange of, verbal information (McNeill, 1992; Saryazdi & Chambers, 2022). Secondly, emotion valence affects recall (Buchannan & Lovallo, 2001). Emotional valence indicates the pleasantness of a stimulus, with positive and negative stimuli being more memorable than neutral (Van Bergen et al., 2015). While many studies find these factors have been beneficial for recall, some studies have shown mixed results (Bierta et al., 2021; Guilbert et al., 2021). Moreover, there is little research investigating interactions between gesture and emotion on narrative recall (Guilbert et al., 2021), and none examining this interaction on recall of written texts read aloud.

Individuals' verbal memory (VM) ability may similarly affect narrative recall. While links between recall and each of VM and gesture (Gillespie et al., 2014; Ward et al., 2016), as well as emotion (Buchannan & Lovallo, 2001), have independently been demonstrated, the interaction of gesture, VM, and emotional valence has not been examined. Investigating how these factors influence narrative recall will further our understanding of how to improve such recall, as well as provide theoretical insights into the underlying mechanisms of the effects of task demands and individual differences on recall. Taking an individual differences perspective (see Özer & Göksun, 2020), the current studies examine how gesture, emotional valence, and individual differences including VM impact the recall of English narratives.

Gesture

The field of embodied cognition has long noted that our understanding of meaning is grounded in multisensory experiences (see Sadoski, 2018, for a review). Indeed, even reading comprehension, which may be considered an activity notably lacking in physical activity, is embodied (see Glenberg, 2011, for a review). The field of embodied cognition is closely linked to gestures (see Cook, 2018, for a discussion of embodied and situated mechanisms relating to gesture). Gestures can be observed (a learner views someone else performing a gesture) or produced (a learner/speaker produces gesture themselves). Here, we focus on gesture production.

The effects of producing gestures on learning have been mixed. For example, producing gestures during encoding has been shown to improve recall of route directions (So et al., 2014) and recall of animated vignettes (Cook et al., 2010), albeit with mixed findings across studies regarding vignettes. A more recent study focusing on verb learning found better recall for the production of meaningful gestures (gesture related to a verb) when compared to unrelated gestures and no gestures (Sivashankar & Fernandes, 2022). Tasks such as learning route directions, recalling animated vignettes, and verb learning are qualitatively different tasks however from encoding of narratives. It is to this point unclear whether similar mechanisms are at play in the effects of gestures on these tasks and therefore whether gestures may be used similarly across tasks. Dargue and Sweller (2018) found gesture production was not beneficial to recall of a video narrative, though again, the mechanisms involved in the effects of gesturing on video narrative recall may differ to learning a written narrative such as that used in the present studies. Further, participants in Dargue and Sweller (2018) gestured during recall and were not required to gesture during the encoding process. Gesturing during encoding (rather than recall) may be most beneficial for learning, due to one mechanism by which gesturing has been proposed to aid recall: through alleviating cognitive load.

Gesture and cognitive load

Gesture and cognitive load. In their seminal study, Goldin-Meadow et al. (2001) examined the effect of gesture on working memory. Children and adults were instructed to solve a mathematics problem and were then given a list of items to remember. Next, participants explained how they solved the problems in either a gesture-permitted or gesture-restricted condition. Participants were then asked to recall the list of items, to measure the impact of gesturing during the mathematics explanation on cognitive load. Findings showed that producing gestures while explaining mathematics strategies allowed for better recall of long word lists. This finding indicated that gesturing during the mathematics explanation reduced the burden on cognitive resources, leaving more resources for recall. Goldin-Meadow et al. (2001) proposed that gestures alleviate cognitive load through creating visuospatial representations, facilitating encoding, and processing of information, thereby reducing the cognitive effort required to complete a task. In this way, an individual's capacity to remember information is enhanced.

Evidence for a reduction in cognitive load is further demonstrated in a study exploring the impact of gesturing when in a tip-of-tongue state, that is, when struggling to retrieve a word (Pyers et al., 2021). Once participants were in a tip-of-tongue state, their cognitive load was increased, and gesturing during this state was associated with more success in retrieving the correct word. Such research suggests gesture production reduces cognitive load, thereby making it easier to recall information. If gesture production can alleviate the load on cognitive resources, its benefits may differ depending on other task, or learner, characteristics. Not all learners will find all tasks equivalently challenging, and individual differences in VM may affect how difficult a learner finds a verbally based task.

Verbal Memory

VM is memory for spoken information and is part of the larger working memory model (Baddeley, 2000; 2021; Tatsumi & Watanabe, 2009). Working memory is the temporary storage and manipulation of information when performing complex tasks (Baddeley, 2000). It has four main components: the central executive system, the phonological loop, the visuospatial sketchpad, and the episodic buffer, the latter of which involves all sensory modalities, including haptics (Baddeley, 2021). Just and Carpenter (1992) proposed the capacity of working memory to remember information is limited when engaging in complex tasks and working memory capacity positively predicts comprehension. Indeed, working memory capacity has been shown to be negatively correlated with unrelated thoughts (mind wandering) but positively correlated with comprehension (McVay & Kane, 2012; Robison & Unsworth, 2015; Unsworth & McMillan, 2013). These findings suggest individuals with low working memory capacity are more prone to losing focus on tasks, leading to poorer comprehension.

A meta-analysis of 77 studies found VM was a consistent predictor of comprehension (Daneman & Merikle, 1996). However, while previous studies have measured single sentence comprehension, there is limited research examining more complex texts. Ward et al. (2016) investigated the relationship between VM and narrative recall. Participants heard 12 short stories and after each story were asked to recall what they heard, before listening to the next story. VM positively predicted recall accuracy across both younger and older adult participants. Overall, research shows a strong link between VM and comprehension; however, there is limited research examining more complex texts.

VM and Gesture

Findings are mixed regarding the link between VM and gesture. One study examined the impact of different types of gesture on route recall for children (van Wermeskerken et al., 2016). No correlation was found between VM and gesture and no enhancing effect of gesture on recall, for those with lower VM scores. However, the study was limited as an adult VM task was used for children. Additionally, only gesture observation was examined, not production. A study with adolescent and young adult participants found similar results, with VM not correlating with gesture frequency (Chu et al., 2014). However, their gesture elicitation tasks, defining phrases or answering social dilemmas, were not tasks that imposed a high cognitive load, meaning cognitive load may not have been strained enough to use gestures.

In studies of appropriate task complexity, namely recalling narrative videos, results suggest that VM ability was negatively associated with gesture frequency (Gillespie et al., 2014; Smithson & Nicoladis, 2013). This negative association suggests those with lower VM capacity may use gestures to compensate for their weaker cognitive capacity. Supporting this idea, Wagner et al. (2004) asked participants to solve a mathematics problem and then remember a set of items (visual or verbal) while explaining their mathematics strategy. For the explanation of ten problems, gesturing was allowed, while for ten problems, gesturing was restricted. After each explanation, participants were asked to recall the items. Verbal items were better recalled than visual, and gesturing enhanced recall in both mediums, suggesting gesturing benefits VM. These findings suggest gesturing reduces cognitive load, making more resources available for VM, thereby enhancing recall. As a result, those with lower VM abilities may gesture more to reduce their cognitive load when performing complex verbal tasks. Considering this relationship between gesture and VM, we propose that VM may play a moderating role between factors that influence recall, namely gesture and emotion valence.

Emotion

The dominant theory explaining the link between emotion and recall is that of emotion-enhanced memory (Buchanan & Lovallo, 2001). This theory states that recall is superior for emotionally valenced items, in contrast to neutral items. Emotion is suggested to enhance memory through attention. It has been posited that emotional stimuli capture attention, thereby making them salient in the individual's mind, and making non-emotional details peripheral and hence more forgettable (Kensinger, 2009).

This attention capture can be seen in studies where emotional information was recalled well, whereas peripheral, non-emotional details were not well remembered (Davidson & Vanegas, 2015; Kensinger et al., 2007). Further support was found by Long et al. (2015), who showed positive and negative words were better recalled than neutral words. There was no difference, however, between recall of positive and negative words. Similarly, Bireta et al. (2021) found no difference between recall of emotionally valenced words. Interestingly, they may not have found significant results as they attempted to equate positive and negative words on multiple dimensions including semantic relatedness, meaning there may not have remained sufficient differentiating factors between them. However, similar results were found in studies of word recognition, emotional actions, and word recall, which found emotional items were better recalled than neutral items, but there was no difference in recall of positive and negative items (Adelman & Estes, 2013; Earles et al., 2016; Siddiqui & Unsworth, 2011).

It has long been demonstrated that when reading, we may be more likely to remember content that is subjectively important to us, including affective information. A program of research by Sadoski and colleagues (e.g., Sadoski & Quast, 1990; Sadoski et al., 1988) examined links between affect and recall in both fiction and nonfiction narratives. Interestingly, some studies of narrative recall show a distinction between recall of positive and negative stories. Van Bergen et al. (2015) explored the impact of emotional valence on children's recall. Children listened to emotional narratives and were then tested on free and prompted recall. Their results show negative stories were best remembered, followed by positive, and then neutral. Tournier et al. (2016) extended this finding to adult participants, who alternated between hearing an emotive story and completing a cognitive task. In immediate recall, the negative story was best recalled, followed by positive, then neutral. Similarly, in a recent study involving both child and adult emotional narrative recall, both age groups remembered negative stories significantly better than positive or neutral stories (Guilbert et al., 2021).

Overall, research suggests negative narratives are better recalled than positive narratives, with neutral narratives being the least memorable. The superior recall of negative narratives can be explained by the attention-narrowing hypothesis (Eastbrook, 1959). This theory states that negative emotional stimuli draw more attention and may elicit higher arousal levels, making them more salient and memorable. A review of the attentional processing of emotional information supports this view; negative information attracts attention over other information, thus attentional resources focus on negative information, making it more salient and hence easier to remember (Yiend, 2010). If negative stimuli are remembered best, the beneficial impacts of gesture and VM capacity should be greatest for neutral stimuli, followed by positive stimuli. As negative stimuli are already well recalled, suggesting perhaps a lower load on cognitive resources, the benefits of gesture and high VM capacity should be the least impactful for such stimuli.

Gesture, VM, and Emotional Valence

As outlined above, there are independent links between gesture, VM, and emotion, on recall. However, there is scant research showing how these mechanisms may interact to enhance recall (see Kelly & Tran, 2023, for a review of the emotional functions of gestures, and a call for further research into expanding our understanding of the links between the cognitive and emotional functions of gesture, and De Stefani & De Marco, 2019, for a review including discussion of the role of the motor system in perception of emotion).

Guilbert et al. (2021) investigated the impact of emotional valence and gesture on narrative recall. Participants watched video narratives containing a series of stories with either positive, negative, or neutral content. While observing the videos, they were allocated to one of three gesture conditions: the narrator produced no gesture, gestured during emotional parts of the narrative, or gestured when an event (non-emotional) occurred in the narrative. Results showed children had enhanced recall for information that was presented while the narrator was gesturing. Further, negative stories were recalled better than positive or neutral stories. Contrastingly, for adults, gesturing did not impact recall, and while emotional valence aided event recall, it did not aid recall of emotion itself. Their findings for adults are inconsistent with the literature and may have been due to the recall activity being too simple for adults, rendering gestures redundant.

Similarly, Levy and Kelly (2020) examined the integration of gesture and speech, and how this integration might change depending on the emotional valence of speech. Observed gestures were found to boost memory for positive, negative, and neutral narratives, but were least likely to be integrated into negatively valenced speech. For both Guilbert et al. (2021) and Levy and Kelly (2020) studies, the focus was on gesture observation, not gesture production. Given a meta-analysis found gesture production to have a larger effect size than gesture observation on comprehension (Dargue et al., 2019), the impact of gesture production and emotional valence on adults' recall perhaps needs to be studied with an adequately complex task for effects to be seen.

Recent research has examined the effects of gesture production on bilinguals' retelling of emotional narratives (Özder et al., 2023). L1-Turkish and L2-English speakers' gesture use in emotional contexts was examined, with results indicating that producing gestures might enhance the encoding of emotional content. All gesture production was spontaneous, however, and it is as yet unclear whether instructing participants to gesture, thereby presumably increasing number of gestures produced, might similarly enhance the encoding of emotional narratives.

Here, we take an individual differences perspective and propose that if gestures benefit task performance through impacts on cognitive load, the role of VM as a moderating factor warrants attention. Individuals with lower cognitive resources, in this case lower VM, may benefit more from using gestures, particularly when performing difficult tasks such as encoding narratives lacking in emotional content, than would individuals with higher cognitive resources (see Özer & Göksun, 2020, for a review).

Current Studies

Across two studies, we examined the impact of emotion, gesture, and individual differences on narrative recall. Previous research on the effects of gesture production to date has not focused on tasks involving reading aloud narratives, but rather on tasks such as mental rotation (Chu & Kita, 2011), mathematical problem solving (Goldin-Meadow et al., 2009), mechanical problem solving (Alibali et al., 2011), and narrative tasks involving recall of previously heard, spoken narratives (Dargue & Sweller, 2020). As noted above, the mechanisms involved in the effects of gesturing during the recall phase of video narrative recall may differ to the effects of gesturing during encoding of a narrative that is read aloud. In the current studies, we attempt to generalize from these studies to a narrative that is read aloud by the learner, to better establish the boundaries of gesture's beneficial effects on cognitively challenging tasks.

In Study 1, participants read aloud three separate narratives (positive, negative, neutral) in either an instructed gesture or spontaneous gesture condition. Participants completed a VM task and were then asked to recall the stories. Recall was tested through both free recall and specific recall questions, hereafter referred to as specific recall. Study 2 followed a similar protocol, with the addition of a measure of individual propensity to gesture, measured prior to the main narrative learning and recall task. Across these two studies, we asked (1) whether emotion valence, instructions to gesture, and individual differences in VM interact to affect narrative recall and (2) whether there is a relationship between individual propensity to gesture and VM, and whether the propensity to gesture affects the relationship between emotion valence, instructions to gesture, and narrative recall. We first conducted a Preliminary Study to determine procedural details, followed by Studies 1 and 2, outlined below.

Preliminary Study

A Preliminary Study testing (1) the efficacy of manipulating emotional valence through a single narrative task and (2) the necessity of running three conditions including instructed, spontaneous, and constrained conditions (Ethics approval reference Code: 52021954424524) was conducted with 105 participants. Participants completed one of the three gesture conditions, and verbal memory was measured as a continuous predictor. The dependent variable was a narrative recall. The primary stimulus was an adaptation of the narrative "The Dinner Party" (Gardner, 1941). Participants read the narrative and then completed the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1941) of verbal memory (see Strauss et al., 2006 for procedure). Finally, participants recalled the narrative.

Findings showed gesture production did not facilitate recall. We found no evidence of emotion-enhanced memory, with neutral content remembered best. These results for emotion content may be attributed to how adults process emotion-orientated texts. According to text-processing theory, information perceived to have consequences for the protagonist is regarded as more important to the reader, leading to greater recall (Trabasso & van den Broek, 1985). The neutral narrative information may have been perceived as highly relevant to the protagonist's goals. For example, neutral information such as the "bowl of milk" could be viewed as important as it signaled the consequences of "luring a snake." The related emotions, however, did not directly affect the protagonist's goal of luring the snake. Neutral content may therefore have received greater attention than emotional states, promoting in-depth processing and better memory.

Study 1

The Preliminary Study established that including all emotional valences in one narrative was not a successful method to examine emotion-enhanced memory. In Study 1 therefore, the emotional content was separated between narratives, with three separate narratives created with one emotional valence each. Recall across each narrative was assessed as a whole. Additionally, the constrained gesture condition was excluded, as constraining gesture is perhaps the least ecologically valid of the gesture conditions. In naturalistic instructional settings, constraining gesture is unlikely to occur. Finally, parallel results were seen in the Preliminary Study for the immediate and delayed RAVLT trials, so only immediate recall was tested in Study 1. Immediate verbal memory is therefore referred to from this point on as VM.

It has been suggested that gesturing reduces cognitive load, thereby benefiting recall (Goldin-Meadow et al., 2001). An effect of the gesture condition was therefore expected, such that the instructed gesture condition should show enhanced recall beyond the spontaneous gesture condition. Research also suggests that emotional valence benefits recall, particularly negatively valenced stimuli (Buchanan & Lovallo, 2001). An effect of emotion valence was therefore expected, such that negative content should be best recalled, followed by positive and then neutral content. Finally, research shows that higher VM capacity benefits recall (Daneman & Merikle, 1996) and a positive relationship between VM and recall was therefore expected.

We propose interactions between these factors, allowing for additional benefits in recall. Firstly, if gesturing lightens cognitive load and non-emotional stimuli are difficult to remember, the benefits of gesturing (namely, the difference between the instructed and spontaneous gesture conditions) should be greatest for neutral stimuli, followed by positive stimuli, and have relatively less impact for negative stimuli that are already easier to remember. Further, again if gesturing lightens cognitive load, the benefits of gesturing should increase as VM capacity decreases, as difficulty in recall is perhaps alleviated by gesturing. Similarly, if emotionally valenced stimuli are easier to remember, the effects of emotion valence on recall should be greatest for those with lower VM, who may have more difficulty in recall. Finally, if the beneficial effects of gesture are strongest for non-emotional stimuli, this effect should in turn become larger as VM decreases.

Method

Pilot Study

As the narratives used were created specifically for the current study, a pilot study was conducted to check for the suitability of the narratives. Sixteen participants participated (3 males and 13 females), all recruited from the Undergraduate Psychology pool. Ages ranged from 17 to 59 years ($M=25.36$, $SD=13.31$). Participants were taken through the study procedure as stated below. After concluding the pilot study, changes were made to the narratives and experimental procedure to improve clarity and readability. Changes included alterations to the professions of the characters' parents and changes to follow-up questions that participants were confused by. The pilot participants did not participate in the final study, and their data are not included below.

Participants

Ethical approval was granted by the University Human Research Ethics Committee (Reference code: 52022954436697). First- and second-year participants were recruited from the University Undergraduate Psychology pool. Participants were required to be fluent in English and have normal or corrected to normal vision. A total of 101 students were recruited; however, one student was excluded due to a technology malfunction. The final sample included 100 participants, with 80 females, 19 males, and 1 participant who did not specify their gender. Participants were aged 17–54 ($M = 23.90$, $SD = 9.01$). All participants completed informed consent prior to commencing the study and received course credit on completion of the study.

Experimental Design

The study involved a 2 (gesture condition, between-subjects: instructed gesture, spontaneous gesture) × 3 (emotional valence, within-subjects: positive, negative, neutral) design, with VM measured as a continuous predictor. Participants were pseudo-randomized to a gesture condition such that 50 participants were allocated to each condition. To prevent order effects for emotional valence, the order of presentation of narratives was counterbalanced between participants. The dependent variable was recall, measured through free and specific recall.

Materials and Procedure

All participants attended the study online via Zoom. Participants completed an information and consent form via Qualtrics. Participants were directed to stand up from their desks and take three steps back from their screen, enough for their upper body to be visible. Following informed consent and initial instructions, participants were instructed they would be reading aloud three short narratives. The experimenter shared the narrative on their screen, for participants to read out loud. In the instructed gesture condition, participants were asked to show and tell using their hands while narrating the story. Participants in the spontaneous gesture condition were asked to narrate the story without further instructions.

Participants read three short narratives about someone awaiting a university acceptance letter (see Appendix 1). The narratives followed the main character through their morning until they opened their letter. The stories were differentiated by their emotional valence (positive, negative, neutral), but all followed a parallel structure. The narratives were of similar readability levels as measured by the Flesch Reading Ease metric: positive = 76.1, negative = 70.8, neutral = 79.9. All three narratives had Flesch-Kincaid grade levels in the elementary school range.

On completion of the narrative reading task, participants completed the RAVLT and were then asked to recall everything they could remember from the first narrative, then the second narrative, and finally the third narrative. Free recall was always completed first, followed by specific recall. Participants were asked to freely recall each of the stories in the order they were read at encoding. Recall was measured in this way following past studies that used a similar recall format as a measure of comprehension (Morrow, 1988; Pearman, 2008).

Next, the experimenter asked the participants the specific recall questions for each narrative (see Appendix 2). Unlike the Preliminary Study, where specific recall questions were asked in a random order, in Study 1, all participants were asked the six questions within each narrative in the same order, which mirrored the order of the events in the narrative. As there were three separate, but parallel, stories, asking questions in a random order was deemed too confusing for participants, who would be prone to forgetting which story was which, and what happened in what order. Finally, participants were asked what they believed the study was about, followed by debriefing.

Coding

Both free and specific recall responses were recorded for subsequent transcribing and coding. Coding was conducted using computer software ELAN (2022, Version 6.4). Firstly, to check the gesture manipulation, gesturing during reading aloud was coded. Participants were ascribed one point each time they produced a gesture. Gestures were coded if made by the hands or by other body parts including the head, accompanying speech. Self-maintenance movements such as fiddling with the hair or clothes were not included in the gesture count. Gestures produced were tallied to calculate a total gesture produced score.

Participants were given points for each item of information they could correctly recall from each of the narratives. For free recall, the negative and neutral narratives were scored out of 49 points, and the positive narrative was scored out of 50 points. In ELAN, participant answers were transcribed, and each transcription allocated points for how many units of correct information it contained. An item or unit of correct information was defined as a conceptual unit, rather than as verbatim words or phrases. For example, "woke up" was one unit of correct information despite being two words, as was the single word "Rory" (a character's name).

For specific recall, answers were coded as either correct (1), incorrect (0), or partially correct (0.5). For example, for the question "What job did Rory's parents have?," the correct answer was "convenience store workers," which would receive (1). An answer of "store worker" would receive (0.5) as the type of store is not specified. A completely unrelated occupation would be marked (0), for example "hairdressers."

To evaluate inter-rater reliability, a second coder independently coded 20 randomly selected participants (20%), 10 from each gesture condition. Single-measure mixed effect intraclass correlations (ICC) were used to measure reliability using an absolute agreement model. For gestures produced, ICC $= 0.88$, $p < 0.001$. For free recall, ICC $= 0.98$, $p < 0.001$, and for specific recall, ICC $= 0.92$, $p < 0.001$.

Results

Analysis Plan

Stata version 17 was used to conduct all analyses. The dependent variables, free and specific recall, were assessed for normality within groups and most were skewed. Analyses were therefore bootstrapped with 2000 replications. Linear mixed effects models with an unstructured covariance matrix were run with emotional valence as a within-subjects fixed factor, immediate VM as a mean-centered numeric fixed predictor, and gesture condition as a between-subjects fixed factor. Emotion valence was nested within the random effect of participant. All fixed main effects and interactions were tested. Follow-up tests requiring multiple comparisons were Bonferroni adjusted.

Manipulation Check

A mixed design ANOVA was undertaken to compare the number of gestures produced between gesture conditions, by emotion valence. There was a significant main effect of gesture condition, such that those in the instructed gesture condition ($M = 25.88$, $SD = 2.62$) produced significantly more gestures than those in the spontaneous gesture condition ($M = 0.18$, $SD = 0.16$), $F(1, 98) = 95.68$, $p < 0.001$, with a large effect size, $\eta_p^2 = 0.49$. There was no significant main effect of emotion valence, $F(2, 196) = 0.76$, $p = 0.469$, $\eta_p^2 = 0.008$, and no interaction between gesture condition and emotion valence, $F(2, 196) = 0.96$, $p = 0.38$, $\eta_p^2 = 0.01$. Therefore, the gesture manipulation was effective.

A series of one-sample *t*-tests on the first specific recall question ("How did [character] feel about the day?") were conducted to validate the emotional polarity of each narrative. Participants correctly responded that the character in the positive narrative felt a positive emotion, $M = 0.90$ out of 1, $SD = 0.27$, $t(99) = 13.77$, $p < 0.001$. Participants similarly correctly responded that the character in the negative narrative felt a negative emotion, $M = 0.96$, $SD = 0.20$, $t(99) = 23.56$, $p < 0.001$. Conversely, participants were unable to consistently select an emotion label for the neutral narrative, $M = 0.53$, $SD = 0.49$, $t(99) = 0.51$, $p = 0.61$. Together with the equivalent readability metrics noted above, these findings suggest the narratives evoked different emotional states, but were otherwise equivalent in comprehensibility.

Free Recall

The overall model predicting free recall was significant, Wald $\chi^2(11) = 156.47$, $p < 0.001$. There was a significant main effect of the gesture condition (see Table 1 for descriptive statistics), such that the spontaneous condition performed better than the instructed gesture condition, $\chi^2(1) = 11.84$, $p < 0.001$.

Table 1 Study 1 free recall scores by emotional valence and gesture condition

Gesture condition	Emotional valence					
	Positive		Negative		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*
Instructed	9.48	4.63	10.90	5.21	6.24	3.60
Spontaneous	11.18	4.85	12.12	5.96	7.46	4.55

There was a significant positive effect of immediate VM ($M = 11.20$, $SD = 2.53$), $z = 4.85$, $p < 0.001$, 95% CI [0.23, 0.53]. Finally, there was a significant main effect of emotional valence, $\chi^2(2) = 91.33$, $p < 0.001$. Pairwise comparisons showed that recall of the positive narrative was significantly worse than the negative narrative, $z = -2.44$, $p = 0.01$, 95% CI [−2.12, −0.23]. The neutral narrative was significantly worse recalled than the negative narrative, $z = -8.87$, $p < 0.001$, 95% CI [−5.69, −3.63]. Finally, recall of the neutral narrative was significantly worse than the positive narrative, $z = -7.69$, $p < 0.001$, 95% CI [−4.37, −2.60].

There was no significant interaction between gesture and emotional valence, $\chi^2(2) = 0.44$, $p = 0.802$, or between VM and emotional valence, $\chi^2(2) = 3.17$, $p = 0.205$. However, there was a significant interaction between gesture and VM, $\chi^2(1) = 23.38$, $p < 0.001$ (see Fig. 1). For VM one standard deviation below the mean, gesture condition did not impact recall, $\chi^2(1) = 1.41$, $p = 0.234$. For mean VM, however, those in the spontaneous gesture condition had significantly better recall than those in instructed, $\chi^2(1) = 11.84$, $p < 0.001$. Similarly, for VM one standard deviation above the mean, spontaneous gesture had significantly better recall than instructed gesture, $\chi^2(1) = 31.66$, $p < 0.001$. The three-way interaction between emotion, gesture, and VM was not significant, $\chi^2(2) = 0.71$, $p = 0.703$.

Specific Recall

The overall model predicting specific recall was significant, Wald $\chi^2(11) = 114.69$, $p < 0.001$. There was no significant main effect of gesture condition, $\chi^2 = 1.48$, $p = 0.224$ (see Table 2 for descriptive statistics), and no significant main effect of VM, $z = 1.35$, $p = 0.177$, 95% CI [−0.02, 0.09]. There was a significant main effect of emotional valence, $\chi^2(2) = 92.83$, $p < 0.001$. Pairwise comparisons

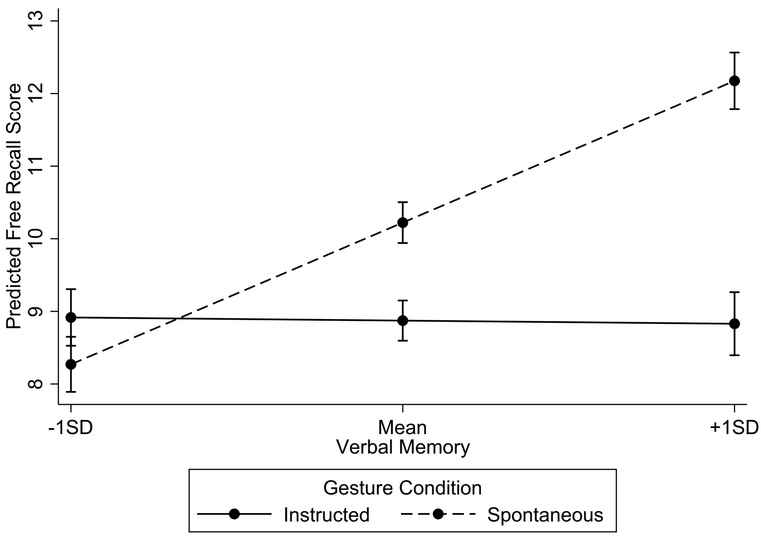

Fig. 1 Interaction between gesture condition and VM, predicting free recall

Table 2 Study 1 specific recall scores by emotional valence and gesture condition

Gesture condition	Emotional valence					
	Positive		Negative		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*
Instructed	4.72	1.13	4.95	1.08	3.19	1.80
Spontaneous	5.19	0.84	4.86	1.03	3.31	1.87

showed no difference between the positive and negative narratives, $z=0.33$, $p=0.744$, 95% CI [−0.24, 0.34]. However, recall of the neutral narrative was significantly worse than recall of both the negative narrative, $z=-8.58$, $p<0.001$, 95% CI [−2.04, −1.28], and the positive narrative, $z=-9.12$, $p<0.001$, 95% CI [−2.07, −1.34].

The interactions between gesture and emotional valence, $\chi^2(2)=3.86$, $p=0.145$, and between VM and emotional valence were not significant, $\chi^2(2)=5.44$, $p=0.066$. The interaction between gesture and VM was significant, however, $\chi^2(1)=4.23$, $p=0.040$ (see Fig. 2). Gesture condition did not affect recall for those with VM one standard deviation below the mean, $\chi^2(1)=0.39$, $p=0.531$, or those with mean VM, $\chi^2(1)=1.48$, $p=0.224$. For those with VM one standard deviation above the mean, however, participants in the spontaneous gesture condition had significantly better recall than those in the instructed gesture condition, $\chi^2(1)=6.39$, $p=0.012$. Finally, the three-way interaction between emotion, gesture, and VM was not significant, $\chi^2(2)=0.73$, $p=0.693$.

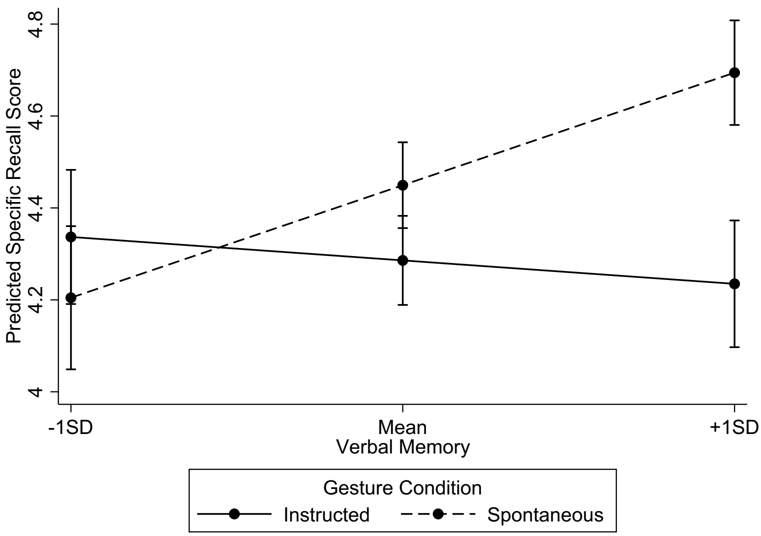

Fig. 2 Interaction between gesture condition and VM, predicting specific recall

Discussion

Study 1 investigated the effect of gesture, emotional valence, and VM on narrative recall, with separate narratives for each emotion valence. As expected, VM and emotional valence positively predicted recall. Unexpectedly, however, the instruction to gesture was detrimental to recall, but not for those low in VM.

Gesture and VM

We expected that the beneficial effects of gesture would be stronger for those with lower VM scores. We found significant interactions for both free and specific recall, however, in the opposite direction to that which was expected. For free recall, gesture condition did not influence recall for those with lower VM scores. However, for those with average and above-average VM scores, those in the spontaneous gesture condition performed better than those in the instructed gesture condition. Similarly, for specific recall, recall of individuals with low and average VM did not differ between the two gesture conditions. Yet, for those with above-average VM scores, the gesture condition was again influential, with those in the spontaneous condition performing better than those in the instructed condition.

These unexpected results may stem from differences in naturalistic gesturing. As noted above, depending on task factors, those with lower VM abilities tend to gesture more frequently, whereas those with higher VM abilities gesture less frequently (Gillespie et al., 2014; Smithson & Nicoladis, 2013). It is possible therefore that when those in the instructed gesture condition were asked to gesture, these instructions had divergent effects on those with differing VM abilities, as a result of their differing baseline propensity to gesture. Gesturing is perhaps more habitual for individuals with lower VM scores, meaning they were able to perform gestures as normal. Simply enhancing an activity they are naturally prone to performing may have meant they could provide adequate attention to the narratives.

Participants with higher VM and therefore less natural propensity to gesture however may have viewed gesturing as an additional, effortful task, preventing it from lightening the cognitive load. Participants' attention may have been divided between reading and gesturing, resulting in worse performance than the spontaneous gesture group, who were only focused on reading. It has been shown that when attention is divided between a primary and secondary task during encoding, performance on later memory tasks is impaired (Naveh-Benjamin et al., 2014). Thus, as the spontaneous gesture condition was only focused on reading, they were able to provide undivided attention to the narratives, allowing them to have better recall than the instructed gesture condition.

Emotional Valence

We predicted emotional valence would benefit recall. Free recall results supported this expectation, with the negative story recalled best, followed by positive, then neutral. For specific recall, while there was no significant difference between the positive and negative narratives, the neutral narrative was again recalled worse than both emotive narratives. These findings align with previous research, providing support for the beneficial effects of emotional valence on recall (Earles et al., 2016; Van Bergen et al., 2015).

Study 2

Study 2 attempted to replicate some of the unexpected findings of Study 1 and to further explore the possible association between naturalistic propensity to gesture and verbal memory. It also removed the potential problems associated with conducting a gesture study online, in which individuals' gesturing is perhaps less naturalistic. As Zoom video frames do not usually include people's hands, participants may not be used to gesturing, and therefore seeing their own gestures, while on Zoom. By the time of running Study 2, the social distancing measures enforced during the COVID-19 pandemic that required the Preliminary Study and Study 1 to be run remotely had been removed, and Study 2 was conducted face to face. Finally, given the literature reviewed above suggesting tasks of appropriate difficulty are crucial to see effects of gesture on cognitive load and therefore on task performance, Study 2 attempted to tap into deeper narrative comprehension rather than just recall, as was measured in Study 1.

A negative relationship was expected between immediate VM and individual propensity to gesture. Following from this expected negative relationship, predictions regarding the relationship between propensity to gesture and recall scores are reversed from those noted in Study 1 for VM and recall. Specifically, a negative relationship was expected between propensity to gesture and recall, and interactions should similarly be reversed in direction: we expected the effect of instruction to gesture on narrative comprehension to be stronger for those with a higher propensity to gesture and the emotional valence effect to be stronger for individuals with a higher propensity to gesture.

Method

Participants

The study recruited 102 participants through the University Psychology Participant pool (Ethics approval reference code 52023954446613). Participants were enrolled in first-year psychology or second-year cognitive psychology units and self-selected into the study for course credit. Study inclusion criteria included normal to corrected to normal vision and fluency in English. Four participants were excluded from the analysis due to experimental error and failure to meet inclusion criteria, resulting in a final sample of $N = 98$ (73 females, 22 males, one non-binary, and two preferred not to specify their gender). Participant age ranged from 17 to 58 years ($M = 20.61$, $SD = 5.69$).

Experimental Design

This study used the same 2×3 design as Study 1. Participants were quasi-randomized into gesture condition groups to ensure equal allocation ($n = 49$ in each condition). Narrative presentation order was randomized between participants using an online random number generator. Two continuous predictors of VM and propensity to gesture were measured for each participant. Narrative comprehension was again measured by both free and specific recall, with an additional measure of extended comprehension outlined below.

Materials and Procedure

Unlike Study 1, this study was conducted face to face, in which the experimenter and participant sat together in individual testing sessions. Following informed consent, the experimenter instructed the participant, "I am now going to show you a short video," and administered the propensity to gesture stimulus on a laptop screen. The propensity to gesture stimulus was a 2-min 38-s animated video titled "Donald Duck's Garden," depicting a cartoon duck attempting to water his prize-winning watermelons. This video was used in previous research as an effective gesture elicitation stimulus (Dargue & Sweller, 2018). After viewing the cartoon, participants were asked to stand up and face the experimenter, who informed them they would begin audio and video recording. The experimenter obtained verbal consent to begin audio and video recording for the entirety of the study. Once the recording began, participants were asked four open-ended questions about the cartoon drawn from Stevanoni and Salmon (2005) (i.e., "tell me everything that happened in the video that you just saw," "tell me everything Donald Duck did in the video," "tell me everything Donald Duck saw in the video," and "is there anything else you remember about the video?"). No instruction to gesture was given, as we were interested in participants' naturalistic gesturing. However, if the participant's hands were not visible to the camera (i.e., behind their back) or in a position that prevented possible gestures (i.e., in their pockets), then they were asked not to place their hands in that position without any further instruction.

Participants were requested to remain standing for the duration of the study and were given the first narrative, displayed on a laptop screen in front of them. Slight modifications were made to the narratives for Study 2, such as changing words (i.e., coffee, curtain) to prevent overlap with the RAVLT, and to ensure the narratives had an equal number of words (168 words), and story elements (49 elements). The modified narratives can be found in Appendix 3. The narratives were of similar readability levels as measured by the Flesch Reading Ease metric: positive = 75.1, negative = 69.5, neutral = 81.3. All three narratives had Flesch-Kincaid grade levels in the elementary school range.

Participants read each narrative aloud from a laptop screen, with each narrative presented on a separate PowerPoint slide. Instructions for each gesture condition mirrored those of Study 1. Those in the instructed gesture condition were told to "read the story aloud while showing and telling the story with your hands," whereas those in the spontaneous gesture condition were asked to "read the story aloud" without further instructions. However, those in the spontaneous gesture condition were asked to remove their hands from positions that may have suppressed or hidden gestures, without explicitly mentioning gesture use. The experimenter then removed the laptop from the desk and then administered the RAVLT according to standard protocol, as in Study 1.

The narrative comprehension questions were then administered. Free recall was measured by stating, "tell me everything you remember" about the first story, then the second story, and the final story in the order they initially read them aloud. Participants then completed the specific questions pertaining to each narrative, again in the order of the narratives the participant initially read aloud. No feedback was given regarding which questions corresponded with which story; the experimenter said "okay" to each response without further comments.

Within the set of questions, six were drawn from Study 1, and six were newly created to reflect deeper, extended narrative comprehension (see Appendix 4). The newly created extended comprehension questions were adapted from Paris and Paris (2003) and involved drawing together multiple elements of the story, rather than simple direct recall of one story element. No prompts or feedback were given to participants, and the experimenter only responded with "okay" after each response.

Coding

Audio and video recordings were coded using ELAN (Version 6.5). All gestures and words produced during the propensity to gesture task were individually counted, such that each gesture was given 1 point and each word was given 1 point. Propensity to gesture was then calculated as a gesture rate per 100 words spoken.

Free recall was scored separately for each narrative as a whole number out of 49, corresponding to each of the 49 elements in each narrative. Specific recall questions were scored as either correct (2 points), partially correct (1 point), or incorrect (0 points). Scoring was altered to be out of 2 per question rather than out of 1 per question as in Study 1, to accommodate the newly adapted comprehension questions, which were scored similarly to the Paris and Paris (2003) scoring guide. For example, when asked "Why do you think Rory reacted that way?," responses that connected these reactions to two elements of the narrative received 2 points, responses that connected these reactions to one element of the narrative were given 1 point, and response that failed to connect to any narrative elements were given 0 points. Two separate dependent variables for specific questions were calculated: one to directly mirror Study 1 involving the six questions that paralleled those in Study 1 (called specific recall below) and one with the deeper, extended narrative comprehension questions (extended comprehension).

A second rater independently coded 20% of all participant data (20 videos) to assess inter-rater reliability, and ICCs were again calculated. For propensity to gesture, ICC = 0.94, $p < 0.001$. For gestures produced at encoding, ICC = 0.98, $p < 0.001$. For free recall, ICC = 0.75, $p < 0.001$. For specific questions, ICC = 0.95, $p < 0.001$.

Results

Analysis Plan

Analyses were run using Stata (version 18). As the assumption of normality was not consistently met, mixed models were run with 2000 bootstrapped repetitions with the within-subjects fixed factor of emotional valence, between-subject fixed variables of gesture condition, propensity to gesture and VM, and emotion valence nested within the random effect of participant. Propensity to gesture and VM were mean centered prior to analysis. Initially, the full four-way models including all main effects and interactions were run for each of free recall, specific recall, and extended comprehension questions. In all three models, the four-way interaction was not significant. It was therefore dropped from all models, to keep analyses consistent with the three-way models conducted in Study 1. Three-way models are reported below. Post-hoc pairwise comparisons were run for any significant effects of emotional valence, and simple effects to follow-up interactions. As sphericity was violated for the specific recall dependent variable, the models predicting specific recall were run with an unstructured covariance matrix for the within-subjects factor emotional valence. Follow-up tests requiring multiple comparisons were Bonferroni adjusted.

Preliminary Analyses and Manipulation Check

As in Study 1, a mixed design ANOVA was used to compare the difference in mean gestures produced between conditions. The manipulation was effective, as participants in the instructed gesture condition produced significantly more gestures ($M = 75.25$, $SD = 21.49$) than those in the spontaneous gesture condition ($M = 3.59$, $SD = 8.47$), $F(1, 96) = 471.39$, $p < 0.001$, with a very large effect size, $\eta_p^2 = 0.83$. Although there was a main effect of emotion valence, $F(2, 192) = 4.27$, $p = 0.015$, $\eta_p^2 = 0.04$, such that participants produced more gestures while encoding the positive narrative ($M = 13.64$, $SD = 13.66$) than the neutral narrative ($M = 12.27$, $SD = 12.86$), Bonferroni adjusted $p = 0.031$, there was no significant interaction between gesture condition and emotion valence, $F(2, 192) = 1.06$, $p = 0.35$, $\eta_p^2 = 0.01$.

As in Study 1, a series of one-sample *t*-tests on the first specific recall question ("How did [character] feel about the day?") were conducted to validate the emotional polarity of each narrative. Participants correctly responded that the character in the positive narrative felt a positive emotion, $M = 1.88$ out of 2, $SD = 0.41$, $t(97) = 21.04$, $p < 0.001$. Participants similarly correctly responded that the character in the negative narrative felt a negative emotion, $M = 1.93$, $SD = 0.33$, $t(97) = 27.94$, $p < 0.001$. Unlike in Study 1, participants did select a neutral emotion label significantly above chance for the neutral narrative, $M = 1.27$, $SD = 0.79$, $t(97) = 3.31$, $p = 0.001$. However, this is a much smaller difference from chance responding than for the positive and negative narratives, which together with the readability metrics noted above, again suggests the narratives evoked different emotional states, but were otherwise equivalent in comprehensibility.

To address the relationship between propensity to gesture and verbal memory, a Pearson's correlation was run. There was no significant correlation between these variables, $r(96) = -0.02$, $p = 0.83$.

Free Recall

Overall, the model predicting free recall was significant Wald $\chi^2(21) = 126.04$, $p < 001$. The main effect of the gesture condition was non-significant $\chi^2(1) = 3.57$, $p = 0.059$. However, there were significant positive main effects of propensity to gesture ($M = 13.58$, $SD = 8.10$), $z = 3.09$, $p = 0.002$, 95% CI [0.03, 0.12], and VM ($M = 10.53$, $SD = 3.13$), $z = 3.66$, $p < 0.001$, 95% CI [0.11, 0.37]. There was a significant main effect of emotional valence, $\chi^2(2) = 84.08$, $p < 0.001$ (see Table 3), such that the neutral narrative was recalled significantly worse than both the positive narrative, $z = -6.86$, $p < 0.001$, 95% CI [−4.06, −2.26], and the negative narrative, $z = -8.65$, $p < 0.001$, 95% CI [−5.07, −3.20]. There was no significant difference between the positive and negative narratives following the Bonferroni adjustment, $z = 2.06$, $p = 0.040$, 95% CI [0.05, 1.91].

All two-way interactions were non-significant: gesture condition by propensity to gesture, $\chi^2(1) = 1.08$, $p = 0.299$; gesture condition by emotional valence, $\chi^2(2) = 0.60$, $p = 0.74$; emotional valence by propensity to gesture $\chi^2(2) = 4.51$, $p = 0.105$; gesture condition by VM, $\chi^2(1) = 0.05$, $p = 0.828$; emotional valence by VM, $\chi^2(2) = 0.33$, $p = 0.849$; and propensity to gesture by VM, $z = 0.34$, $p = 0.735$. Finally, no three-way interactions between gesture condition, emotional valence, propensity to gesture, and VM were significant (all $p > 0.200$).

Specific Recall

The model predicting specific recall was significant, Wald $\chi^2(21) = 216.66$, $p < 0.001$. There was a significant main effect of the gesture condition, such that those in the spontaneous condition recalled more than those in the instructed gesture condition $\chi^2(1) = 9.47$, $p = 0.002$ (see Table 4). There was a significant positive main effect of propensity to gesture, $z = 2.98$, $p = 0.003$, 95% CI [0.01, 0.07]. There was a significant main effect of emotional valence, $\chi^2(2) = 148.31$, $p < 0.001$, such that participants had significantly worse recall for the neutral narrative than both the positive, $z = -9.76$, $p < 0.001$, 95% CI [−3.61, −2.40], and the negative narrative, $z = -11.90$, $p < 0.001$, 95% CI [−4.30, −3.08]. Participants

Table 3 Study 2 free recall of narratives by emotional valence and gesture condition

Gesture condition	Emotional valence					
	Positive		Negative		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*
Instructed	11.51	5.16	12.65	5.50	8.37	4.28
Spontaneous	11.16	4.08	11.88	4.79	8.06	3.82

Table 4 Study 2 specific recall of narratives by emotional valence and gesture condition

Gesture condition	Emotional valence					
	Positive		Negative		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*
Instructed	9.20	1.95	9.76	1.71	5.49	3.09
Spontaneous	9.43	2.28	10.20	1.86	7.16	2.93

recalled significantly more in the negative narrative than in the positive narrative, $z = 2.68$, $p = 0.007$, 95% CI [0.19, 1.19]. There was no significant effect of VM, $z = 1.70$, $p = 0.089$, 95% CI [−0.01, 0.14].

The interaction between gesture condition and propensity to gesture was significant $\chi^2(1) = 6.89$, $p = 0.008$ (see Fig. 3). For propensity to gesture one standard deviation below the mean, $\chi^2(1) = 16.57$, $p < 0.001$, and for mean propensity to gesture, $\chi^2(1) = 9.47$, $p = 0.002$, participants in the spontaneous gesture condition recalled significantly more than participants in the instructed gesture condition. There was no significant difference between gesture conditions for propensity to gesture one standard deviation above the mean, $\chi^2(1) = 0.20$, $p = 0.654$.

The interaction between gesture condition and emotional valence was significant, $\chi^2(2) = 6.90$, $p = 0.032$ (see Fig. 4). Follow-up simple effects showed an effect of gesture condition for the neutral narrative only, $\chi^2(1) = 11.25$, $p < 0.001$, with the spontaneous condition showing greater specific recall than the instructed gesture condition. There were no significant differences between gesture conditions for the positive, $\chi^2(1) = 0.09$, $p = 0.766$, or negative narratives $\chi^2(1) = 1.08$, $p = 0.300$.

There was no significant interaction between emotional valence and propensity to gesture, $\chi^2(2) = 1.69$, $p = 0.429$, between emotional valence and VM, $\chi^2(1) = 1.10$,

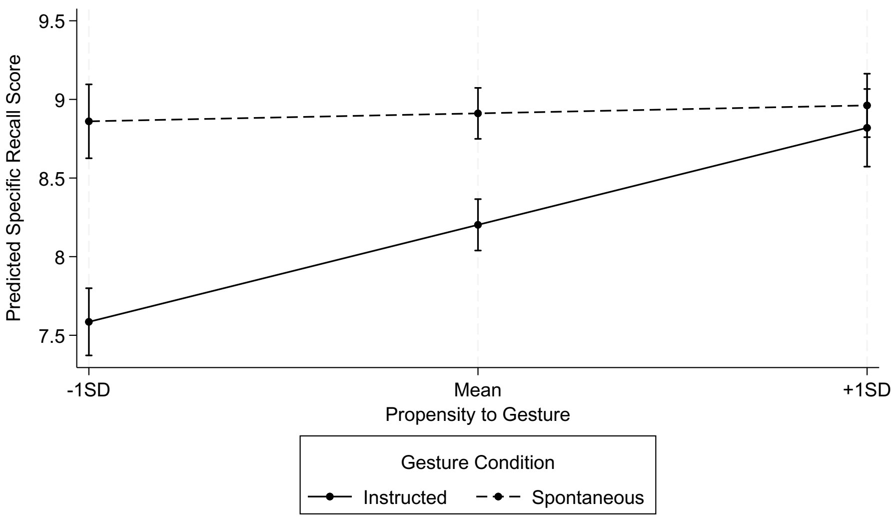

Fig. 3 Interaction between gesture condition and propensity to gesture, predicting specific recall

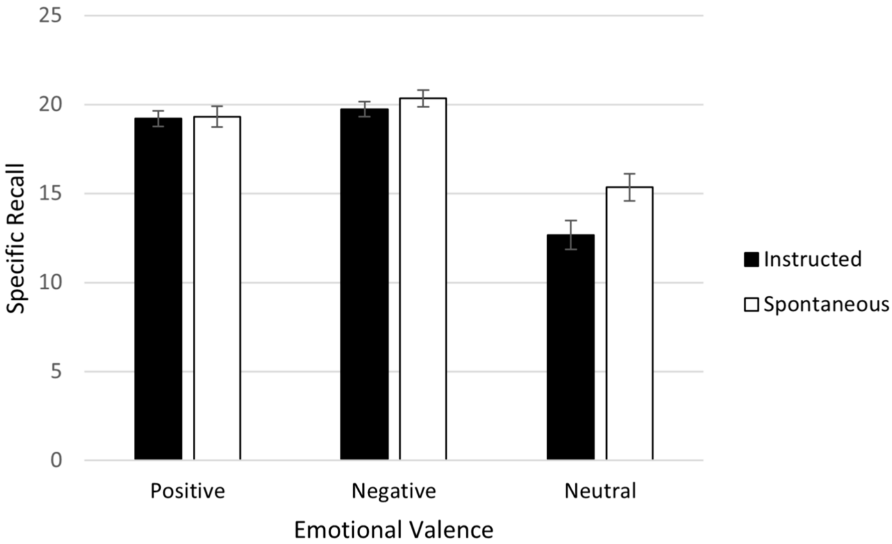

Fig. 4 Interaction between gesture condition and emotional valence on specific recall of narratives

$p = 0.577$; between gesture condition and VM, $\chi^2(1) = 0.99$, $p = 0.319$; or between VM and propensity to gesture, $z = 0.48$, $p = 0.632$. All three-way interactions were non-significant (all $p > 0.200$).

Extended Comprehension

Overall, the model predicting extended comprehension was significant, Wald $\chi^2(21) = 95.16$, $p < 0.001$. The main effect of the gesture condition was non-significant, $\chi^2(1) = 1.11$, $p = 0.293$. However, there were significant positive main effects of VM, $z = 2.18$, $p = 0.029$, 95% CI [0.01, 0.14], and propensity to gesture, $z = 2.48$, $p = 0.013$, 95% CI [0.01, 0.07]. There was a significant main effect of emotional valence, $\chi^2(2) = 63.49$, $p < 0.001$, such that the neutral narrative was recalled significantly worse than both the positive narrative, $z = -7.11$, $p < 0.001$, 95% CI [−2.92, −1.66], and the negative narrative, $z = -7.68$, $p < 0.001$, 95% CI [−3.06, −1.82] (see Table 5). There was no significant difference between the positive and negative narratives, $z = 0.64$, $p = 0.519$, 95% CI [−0.30, 0.60]. All two-way interactions were non-significant (all $p > 0.100$), as were all three-way interactions (all $p > 0.200$).

Table 5 Study 2 extended comprehension of narratives by emotional valence and gesture condition

Gesture condition	Emotional valence					
	Positive		Negative		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*
Instructed	10.00	1.67	9.98	1.53	7.16	3.11
Spontaneous	9.86	2.22	10.12	2.04	8.16	2.88

Discussion

Study 2 expanded on Study 1 by investigating whether the propensity to gesture was an underlying factor in the unexpected results found in Study 1. Specifically, it examined whether a relationship between verbal memory and propensity to gesture existed and whether the effect of gesture condition on narrative comprehension was moderated by emotional valence and propensity to gesture, as well as VM. Results suggest that task-related factors, namely, gesture condition and emotional valence, had independent and interactive effects on direct narrative recall, but only emotional valence had consistent effects across recall measures including extended narrative comprehension. Furthermore, results found that the individual-related factor of propensity to gesture influenced narrative comprehension, but that there was no relationship between propensity to gesture and verbal memory. Finally, for specific narrative recall, the propensity to gesture interacted with the instruction to gesture manipulation, such that instructions to gesture were detrimental to narrative recall, but not for those with above-average propensity to gesture.

Study 2 further attempted to replicate the results of Study 1 regarding the effects of verbal memory on narrative recall and interactive effects of verbal memory with gesture condition and emotional valence. While there were significant main effects of VM on both free recall and specific extended comprehension, there was no significant main effect of VM on specific recall, and the significant interaction between verbal memory and gesture condition found in Study 1 was not replicated.

Gesture Condition and Emotion Valence

As expected, emotional narratives were recalled better than neutral narratives on both free and specific recall measures, as well as on extended comprehension. However, there was no significant difference in free recall or extended comprehension between the positive and negative narratives, with the negative narrative recalled better than the positive narrative for specific recall only. As in Study 1, instructing participants to gesture was detrimental to the specific recall of narrative content, but not consistently across other tasks or individual participant characteristics. Unlike Study 1, this effect did not interact with VM, and nor was the effect present for free recall or extended comprehension questions. However, for specific recall, the instruction to gesture was detrimental for (a) neutral narratives only and (b) participants who were low to average on the individual propensity to gesture. We discuss the interaction with emotional content here and return to the propensity to gesture below.

Instructions to gesture were detrimental for specific recall of narratives, however, only for the neutral narrative. While prior research shows that gestures enhance narrative comprehension and recall (Vilà-Giménez & Prieto, 2020), the neutral narrative task may have been too difficult to process. While the positive and negative narratives cover recognizable schemas for university students (learning whether they had been accepted into university), the neutral narrative is somewhat more vague. The main character may have opened unrelated mail at the end of the narrative, making this task more difficult to comprehend. Given this difficulty, the instruction to gesture may have added cognitive load, rather than reducing it. These results mirror those of Overoye and Wilson (2020), who showed similar detrimental effects of gesture and found those who were instructed to gesture performed worse on a subsequent task with higher levels of difficulty compared to common tasks used in gesture literature.

Propensity to Gesture and Verbal Memory

Contrary to expectations, propensity to gesture was not associated with verbal memory and was positively associated with narrative recall on all three measures, rather than the anticipated negative relationship. On specific recall, however, this effect interacted with the instruction to gesture manipulation, such that although the instruction to gesture did not affect recall for those with a high propensity to gesture, it was detrimental for those with below average, or average, propensity to gesture. This finding provides some, albeit inconsistent, support for the suggestion above following Study 1, that an instruction to gesture may constitute an additional, effortful task for those who are less prone to gesture naturally. In turn, this additional task perhaps hinders encoding and recall. This result should be treated with caution, however, as it was not replicated for free recall or for extended comprehension.

As in Study 1, verbal memory was positively associated with free recall, as well as with extended comprehension. Unlike Study 1, however, there was no significant effect for specific recall, and VM did not interact with gesture condition. The lack of a relationship between gesture and VM skills was unexpected and contrary to previous research suggesting that those with higher memory skills might have less tendency to gesture (Gillespie et al., 2014; Smithson & Nicoladis, 2013). As we note previously, however, the relationship between gesture production and VM skills has been found to varying extents and may be modulated by task-specific factors. It is possible that the specific gesture elicitation task used in Study 2 underlies this lack of a relationship. The video was chosen as it is high action and spatially based and has been shown in the past to elicit gestures (Dargue & Sweller, 2018). However, given the nature of the video content, it is possible that all speakers, regardless of VM abilities, used the gestures to help explain the visual stimulus. Indeed, previous research has shown that participants who viewed an animated video produced more gestures than those who read a written description of the same video (Hostetter & Hopkins, 2002). Future research might consider a gesture propensity measure that more closely relies on VM and therefore more closely mirrors the task of interest, namely the narrative recall task in the current study.

General Discussion

The current studies investigated the effects of instructions to gesture, emotional valence, and individual differences in VM and propensity to gesture on recall of a written narrative. While results supported the theory of emotion-enhanced memory (Buchanan & Lovallo, 2001), with the emotive narratives producing better recall than the neutral narrative, our negative effect of instructions to gesture on narrative recall was unexpected. These findings provide minimal support for the idea that gestures reduced cognitive load for the current tasks. In Study 1, this effect appeared only for individuals with higher VM, while in Study 2, it appeared only for individuals with a lower propensity to gesture and for recall of the neutral narrative task.

Both studies elicited the unexpected finding that instructing some adults to gesture led to significantly worse recall than giving no gesture instructions. This finding suggests that gesturing may be detrimental for some adults' recall, at least when learning by reading aloud narrative texts. We do not propose that previous theories suggesting gesture production can reduce cognitive load were put forward in error (e.g., Goldin-Meadow et al., 2001). Rather, it appears that for a task such as that used in the current studies, namely learning a written narrative by reading it aloud, that gestures may not decrease cognitive load. Indeed, it has been previously proposed that gestures may not always decrease cognitive load. For example, Cook (2018) notes that gesture may engage certain memory processes for a given task. When other methods activate those same memory processes, gesture should not enhance learning. In the current studies, if gesturing engaged memory processes involved in imagery for example, but the act of reading aloud the narrative likewise involved the activation of mental imagery, then gestures may not serve to benefit narrative learning despite imagery being beneficial for learning (Paivio, 1990).

Further, we noted above that tasks such as learning route directions, recalling animated vignettes, and verb learning, which had previously been shown to benefit from gesturing during encoding, are qualitatively different tasks than the encoding written narratives tasks used in the current studies. It is possible that the task of explicitly attempting to elicit gesture production during the encoding of a verbal narrative is an unnatural task for adults to undertake. Indeed, the very low gesture production during encoding seen in the spontaneous gesture conditions suggests that gesturing while reading aloud a narrative may be an unnatural task. Future research should examine the effects of instructions to produce gesture on narrative recall for more naturalistic tasks, such as in parents reading stories aloud to their children.

The individual differences perspective that we took in designing the current studies (see Özer & Göksun, 2020) was partially supported. Although VM was positively associated with recall in both studies and propensity gesture was (albeit unexpectedly) positively associated with recall in Study 2, our primary interest was in the interactions between these measures of individual differences with emotion valence and gesture condition. In Study 1, the instruction to gesture appeared detrimental to those with higher VM, and in Study 2, the instruction to gesture was detrimental to those with a lower propensity to gesture. These results should be treated with caution, however, as the patterns of interactions are not precisely mirrored across studies. The current findings require replication, with a focus on when, and for whom, an instruction to gesture during learning may affect subsequent recall. We propose here that careful attention must be paid to the alignment between a gesture elicitation task and tasks related to cognitive abilities, as discussed above.

A potential solution to the alignment between gesture elicitation and primary tasks is an alteration of the study design. Many gesture studies use a within-subjects manipulation for gesture production, where participants first receive no instructions regarding gesture production (spontaneous gesture) and subsequently are encouraged to gesture (e.g., Alibali & DiRusso, 1999; Broaders et al., 2007). In this way, individual differences in gesture production in the specific task in question can be taken into account. The current Study 2 used a separate task to establish individual gesture production rates, which as noted above, likely has less alignment with the target task than a within-subjects manipulation would provide. Future research should consider a replication of the current studies with a within-subjects manipulation of gesture condition, to better account for individual differences in gesture production.

A closer examination of individual differences in some cognitive abilities is warranted. For example, we note that in the current studies, we did not measure participants' cognitive load. Measuring individual experiences of cognitive load, in addition to individual pre-existing differences in VM, may aide in determining if gesturing increases or decreases load for individuals with varying VM during a narrative encoding task. While higher VM was beneficial for narrative recall, the pattern of interactions suggests that our measure of pre-existing VM perhaps does not relate to the cognitive load that different individuals may experience when encoding the written narrative. Further, we did not include any measures of participants' mood states or general emotion understanding, which are important individual differences that may interact with individuals' gesture use. In addition to measures of cognitive load noted above, future research might include variables on individual mood states and emotion understanding.

The most consistent findings from the current studies are those involving the effects of emotional valence on recall. The negative narrative was recalled best across studies and across measures of recall, though the extent to which it differed from the positive narrative varied between free and specific recall across Studies 1 and 2. The neutral narrative, however, was consistently recalled most poorly, across both studies and all recall measures. As noted above, these results support the theory of emotion-enhanced memory (Buchanan & Lovallo, 2001) and have clear implications for educational settings when seeking to improve comprehension. The current findings suggest incorporating emotional valence into instructional content when teaching from written text may benefit comprehension, and for content in which emotional content cannot be added, additional instructional supports may be needed.

Although instructions to gesture have in the past been shown to be beneficial to learners, careful consideration should be given to instructions to gesture while learning narrative-based tasks. It is as yet unclear whether the current findings may generalize to other tasks that involve reading aloud, and as noted above, replication of the current inconsistent results is needed before drawing definitive conclusions. It is a limitation of the current study's chosen tasks that in educational settings, it would be unusual to find students expected to read aloud a narrative and shortly after recall it, with minimal rehearsal time. Other tasks that involve reading aloud however may be more common, such as in foreign language learning. The current negative effects, combined with previous findings showing beneficial effects of gesture production, highlight that from a practical perspective, educators need to carefully consider the specific task students are undertaking, before deciding if instructions to produce gestures should be implemented.

Conclusion

The current studies investigated the relationship between gesture production, emotional valence, individual differences (VM and propensity to gesture), and narrative recall. Results showed gesture production was detrimental for recall, but not equivalently across levels of VM in Study 1, or across individual propensity to gesture in Study 2. Emotional valence however was beneficial for learning in both studies. We recommend caution in encouraging gesture production during learning of narrative-based tasks, while noting that additional learning supports may be beneficial in tasks that cannot incorporate emotional content.

Appendix 1. Study 1 narratives

Positive: Rory woke up and the butterflies in her stomach put a smile on her face. She enthusiastically jumped out of bed and lifted her blinds. The day she had been looking forward to was finally here! It was the day of university acceptance offers. She had tried so hard in her last year of school and studied her butt off. Her past papers and relaxation techniques helped her in exams. Her studying had been rewarded as the work became more understandable, her marks increased, and she felt closer to her achieving her dream. Her parents' laughter echoed through the house. They had worked hard at their convenience store, and she wanted to be a first-generation doctor. She skipped out of her room and exclaimed to her family that today was the big day. With a spring in her step, she made her way to the post-box. She opened it without hesitation. A big envelope. She squealed with laughter and tore open the envelope. "We are happy to confirm."

Negative: Lucy woke up, feeling the pit in her stomach. She reluctantly crawled out of bed, not bothering to open her blinds. She was not looking forward to what the day would bring. It was the day she would find out if she had gotten into university. She had studied relentlessly. No matter how hard she had studied, she forgot everything when she went into exams. The work became harder, her marks lowered, and she could feel her dream university slipping away. She heard her parents getting ready for their mornings and her breathing quickened. Her parents were architects, her brother a pilot, and she was going to be the unsuccessful lawyer. She begrudgingly left her room and made her way to the back door to avoid her family. Every step felt heavier than the last until she finally reached her post-box. She opened it with caution. A small envelope. She felt the tears in her eyes, and she slowly opened the envelope. Opening it confirmed her nightmare "We regret to inform you."

Neutral: Brooke woke up and felt her stomach grumble with hunger. She stepped out of bed into her slippers, opened her curtains, then made her bed. It was a normal day of the week for her. She was expecting a post in the mail. She had finished her exams and had completed her final year of high school. It had involved studying for exams as she was considering going to university. She was undecided as she wanted to be an influencer and she may not need a degree for it. She heard her parents making coffee and her sister was brushing her teeth. Her parents worked as teachers at the local primary school and her sister was still in high school. She left her room and walked to the kitchen where she greeted her parents. She then opened the front door. She walked at a moderate pace to the post-box. She opened the post-box. Inside was an envelope which she opened and squinted to read "Dear Applicant, please find enclosed."

Appendix 2. Study 1 specific recall questions

Positive narrative

1. How did Rory feel about the day?

   Answer: Happy/excited
2. What job do Rory's parents have?

   A: Convenience store workers
3. How did Rory feel about her future after school?

   A: Optimistic/excited
4. What did Rory want to do as a career?

   A: Doctor
5. How did Rory react when she opened the post box?

   A: Excited/squealed
6. What did Rory's letter say?

   A: She was accepted

Negative narrative

1. How did Lucy feel about the day?

   A: Sad dreading
2. What job do Lucy's parents have?

   A: Architects
3. How did Lucy feel about her future after school?

   A: Pessimistic/not looking forward to it
4. What did Lucy want to do as a career?

   A: Lawyer

5. How did Lucy react when she opened the post box?

   A: Cried/sad

6. What did Lucy's letter say?

   A: She was rejected

Neutral narrative

1. How did Brooke feel about the day?

   A: No strong feelings/neutral

2. What job do Brooke's parents have?

   A: Teachers

3. How did Brooke feel about her future after school?

   Undecided/unbothered

4. What did Brooke want to do as a career?

   A: Influencer

5. How did Brooke react when she opened the post box?

   A: No reaction

6. What did Brooke's letter say?

   A: Don't find out in the story

Appendix 3. Study 2 narratives

Positive: Rory woke up with a smile on her face and butterflies in her stomach. She enthusiastically jumped out of bed and lifted her blinds. The day she had been looking forward to was finally here! It was the day of university acceptance offers. She had tried so hard in her last year of school and studied her butt off. Her past papers and relaxation techniques helped her in exams. Her studying had been rewarded as the work became more understandable, her marks increased, and she felt closer to her achieving her dream. Her parents' laughter echoed through the hallway. They had worked really hard at their convenience store, and she wanted to be a first-generation doctor. She skipped out of her room and exclaimed to her family that today was the big day. With a spring in her step, she made her way to the post-box. She opened it without hesitation. A big envelope. She squealed with laughter and tore open the envelope. "We are happy to confirm."

Negative: Lucy woke up, feeling the pit in her stomach. She reluctantly crawled out of bed, not bothering to open her blinds. She was not looking forward to today.

It was the day she would find out if she had gotten into university. She had studied relentlessly. No matter how hard she had studied, she forgot everything when she went into exams. The work became harder, her marks lowered, and she could feel her dream university slipping away. She heard her parents getting ready for their mornings and her breathing quickened. Her parents were architects, her brother a pilot, and she was going to be the unsuccessful lawyer. She begrudgingly left her room and made her way to the back door to avoid her family. Every step felt heavier than the last until she finally reached her post-box. She opened it with caution. A small envelope. She felt the tears in her eyes, and she slowly opened the envelope. Opening it confirmed her nightmare "We regret to inform you."

Neutral: Brooke woke up and felt her stomach grumble with hunger. She stepped out of bed into her slippers, opened her blinds, then made her bed. It was a normal day of the week for her. She was expecting a post in the mail. She had finished her exams and completed her final year of high school. It had involved studying for exams as she was considering going to university. She was undecided as she wanted to be a stylist and she may not need a degree for it. She heard her parents making tea and her sister was brushing her teeth. Her parents worked as teachers at the local primary school and her sister was still in high school. She left her room and walked to the kitchen where she greeted her parents. She then opened the front door. She walked at a moderate pace to the post-box. She opened the post-box. Inside was an envelope which she opened and squinted to read "Dear Applicant, please find enclosed."

Appendix 4. Study 2 specific recall and extended comprehension questions and scoring guide

This scoring guide was applied to each narrative (*Rory* = positive, *Lucy* = negative, *Brooke* = neutral). However, if participants simply confused the character names but retained holistic understanding of each separate narrative, no marks were deducted, to avoid participants receiving an overall score of 0 simply because they misremembered the names, rather than being an accurate representation of their narrative comprehension. For example, if a participant recalled Rory (positive story) as Lucy (negative story) and answered all the questions as though they were negative, their responses were coded as being negative with the negative story.

1. How did Rory/Lucy/Brooke feel about the day?

- Rory

  2 points: response indicates that Rory was feeling any positive emotion (e.g., happy/excited)

  1 point: response is an ambivalent response but somewhat identifies a positive valence (e.g., kind of okay, nervous but still confident)

  0 points: response is an incorrect inference of Rory's feelings

- Lucy

  2 points: response indicates that Lucy was feeling any negative emotion (e.g., upset, dreading the day)
  1 point: response is an ambivalent response but somewhat identifies a negative valence (e.g., kind of sad)
  0 points: response is an incorrect inference of Lucy's feelings

- Brooke

  2 points: response indicates that Brooke was feeling neutral (or normal, indifferent)
  1 point: response is an ambivalent response but somewhat identifies a middle ground or balance between positive and negative valence (e.g., nervous but also excited, she was happy but not as happy as the [positive] story or happier than the [negative] story)
  0 points: response is an incorrect inference of Brooke's feeling

2. Why do you think Rory/Lucy/Brooke felt that way?

   2 points: response correctly connects feelings to two elements of the narrative
   1 point: response correctly connects feelings to one element of the narrative
   0 points: response fails to connect feelings to any elements of the narrative

3. What job do Rory's/Lucy's/Brooke's parents have?

- Rory

  2 points: response indicates that Rory's parents are convenience store workers
  1 point: response only mentions the word "convenience" or "store workers" when describing Rory's parents job
  0 point: response fails to identify Rory's parents job

- Lucy

  2 points: response indicates that Lucy's parents are architects
  1 point: response only mentions a similar occupation or broad field broad when describing Lucy's parents job (e.g., designers, builders)
  0 point: response fails to identify Lucy's parents job

- Brooke

  2 points: response indicates that Brooke's parents are primary school teachers
  1 point: response only mentions the word "primary school" or "teachers" when describing Brooke's parents job
  0 point: response fails to identify Brooke's parents job

4. How did Rory/Lucy/Brooke feel about her future after school?

- Rory

  2 points: response indicates that Rory felt positively about her future after school (e.g., good, confident, excited, happy)

1 point: response is ambivalent but correctly identifies somewhat positive valence
0 point: response is not a correct inference of Rory's feeling about her future after school

- Lucy

2 points: response indicates that Lucy felt negatively about her future after school (e.g., anxious, worried)
1 point: response is ambivalent but correctly identifies somewhat negative valence (e.g., she was okay but knew it was limited)
1 point: response is ambivalent but correctly identifies somewhat negative valence (e.g., she was okay but knew it was limited)

- Brooke

2 points: response indicates that Brooke felt neutral or undecided about her future after school
1 point: response is ambivalent (e.g., "she was motivated to go to university but unsure if she would get in", "quite happy but nervous")
0 point: response is not a correct inference of Brooke's feeling about her future after school

5. What did Rory/Lucy/Brooke want to do as a career?

- Rory

2 points: response indicates that Rory wanted to become a doctor
1 point: response only provides a broad or similar occupation (i.e., indicating medical field or medicine)
0 point: response is an incorrect inference of what Rory wants to do as career

- Lucy

2 points: response indicates that Lucy wanted to become a lawyer or indicates law
1 point: response only provides a broad or similar occupation (e.g., an arts degree, or a"successful profession")
0 point: response is an incorrect inference of what Lucy wants to do as career

- Brooke

2 points: response indicates that Brooke wanted to become a stylist
1 point: response only provides a broad or similar occupation (i.e., fashion, designer, hairdresser, beauty industry)
0 point: response is an incorrect inference of what Brooke wants to do as career

6. How did Rory/Lucy/Brooke react when she opened the post box?

- Rory

2 points: response indicates that Rory reaction was overly positive (e.g., squealed with laughter)

1 point: response is ambivalent but correctly identified that Rory's reaction was somewhat positively valanced (e.g., she was kind of happy)
0 points: response is an incorrect inference of Rory's reaction when she opened the post box

- Lucy

2 points: response indicates that Lucy's reaction was overly negative (e.g., tears in her eyes, opened letter with caution)
1 point: response is ambivalent but correctly identified that Lucy's reaction was somewhat negatively valanced (e.g., she was okay but kind of upset)
0 points: response is an incorrect inference of Lucy's reaction when she opened the post box

- Brooke

2 points: response indicates that Brooke's reaction was neutral (e.g., squinted to read)
1 point: response does not explicitly state that Brooke's reaction was neutral but is able to infer that she was feeling "in the middle" of the other narratives or response is ambivalent (e.g., " a bit happy but also okay")
0 points: response is an incorrect inference of Brooke's reaction when she opened the post box

7. Why do you think Rory/Lucy/Brooke reacted that way?

2 points: response connects these reactions to two elements of the narrative
1 point: response connects these reactions to one element of the narrative
0 points: response fails to connect them to any elements of the narrative

8. What did Rory's/Lucy's/Brooke's letter say?

- Rory

2 points: response indicates letter stated, "we are happy to confirm…" (note* similar variations also correct e.g., "pleased to inform you")
1 point: response was a general inference of the letter (i.e., indicates that Rory was accepted)
0 points: response is an incorrect inference about what Rory's letter said

- Lucy

2 points: response indicates that the letter stated "we regret to inform you…"
1 point: only a partial response was provided (i.e., indicates that Lucy was not accepted
0 points: response is an incorrect inference about what Lucy's letter said

- Brooke

2 points: response indicates we are not told about Brooke's outcome or what was written (i.e.,"Dear applicant please find enclosed…")

1 point: only a partial response was provided (e.g., she was on the waiting list, or unsure if she got in or not, but states she is most likely to get in/probably got in)
0 points: response in an incorrect inference about what Brooke's letter said

9. What do you think happens next?

   2 points: response indicates two predictions that extend the narrative
   1 point: response only provides only one prediction that extends the narrative
   0 points: response fails to make a prediction that follows the narrative.

10. Why do you think that happens next?

    2 points: response integrates prediction with at least two elements of the narrative
    1 point: response integrates prediction with one element of the narrative
    1 point: response integrates prediction with one element of the narrative

11. If you knew someone in a similar situation as Rory/Lucy/Brooke, what would you say to them?

    2 points: response provides two or more examples that correctly draw from the narrative's overarching emotional valence
    1 point: response provides only one example that correctly draw from the narrative's overarching emotional valence
    0 point: response does not draw from the narrative's overarching emotional valence

12. Why would you say that to them?

    2 points: response integrates examples with narrative's correct emotional valence
    1 point: response integrates examples with vague or ambivalent emotional valence
    0 points: response does not integrate example with the narrative emotional valence

Author Contribution The studies were initially conducted by the first, second, and third authors as research theses. All authors contributed to the studies' conception and design. Material preparation, data collection, and analyses were performed by Kavya Thakore, Trisha Das, and Shamma Jahan. The first draft of the manuscript was written by Naomi Sweller and adapted from the theses of Kavya Thakore, Trisha Das, and Shamma Jahan. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability Data are available from the authors on request.

Declarations

Ethics Approval All procedures performed in both studies were in accordance with the ethical standards of the institutional Human Research Ethics Committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare no competing interests.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(3), 530–535. https://doi.org/10.1016/j.cognition.2013.08.014

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

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

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Baddeley, A. (2021). Developing the concept of working memory: The role of neuropsychology. *Archives of Clinical Neuropsychology, 36*(6), 861–873. https://doi.org/10.1093/arclin/acab060

Bellaera, L., Debney, L., & Baker, S. T. (2016). Subject comprehension and critical thinking: An intervention for subject comprehension and critical thinking in mixed-academic-ability university students. *The Journal of General Education, 65*(3–4), 264–282. https://doi.org/10.5325/jgeneeduc.65.3-4.0264

Bireta, T. J., Guitard, D., Neath, I., & Surprenant, A. M. (2021). Valence does not affect serial recall. *Canadian Journal of Experimental Psychology, 75*(1), 35–47. https://doi.org/10.1037/cep0000239

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

Buchanan, T. W., & Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology, 26*(3), 307–317. https://doi.org/10.1016/s0306-4530(00)00058-5

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

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General, 143*(2), 694–709. https://doi.org/10.1037/a0033861

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*(8), 1106–1116. https://doi.org/10.1177/0956797610375449

Cook, S. W. (2018). Enhancing learning with hand gestures: Potential mechanisms. *Psychology of Learning and Motivation - Advances in Research and Theory, 69*, 107–133. https://doi.org/10.1016/bs.plm.2018.10.001
Cook, S. W., Yip, T. K., & 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
Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546
Dargue, N., & Sweller, N. (2018). 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
Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand. *Psychological Bulletin, 145*(8), 765–784. https://doi.org/10.1037/bul0000202
Davidson, D., & Vanegas, S. B. (2015). The role of emotion on the recall of central and peripheral information from script-based text. *Cognition and Emotion, 29*(1), 76–94. https://doi.org/10.1080/02699931.2014.896319
de Stefani, E., & de Marco, D. (2019). Language, gesture, and emotional communication: An embodied view of social interaction. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02063
Earles, J. L., Kersten, A. W., Vernon, L. L., & Starkings, R. (2016). Memory for positive, negative and neutral events in younger and older adults: Does emotion influence binding in event memory? *Cognition & Emotion, 30*(2), 378–388. https://doi.org/10.1080/02699931.2014.996530
Easterbrook, J. A. (1959). The effects of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*(3), 183–201. https://doi.org/10.1037/h0047707
ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan
García-Madruga, J. A., Vila, J. O., Gómez-Veiga, I., Duque, G., & Elosúa, M. R. (2014). Executive processes, reading comprehension and academic achievement in 3rd grade primary students. *Learning and Individual Differences, 35*, 41–48. https://doi.org/10.1016/j.lindif.2014.07.013
Gardner, M. (1941). The Dinner Party. The Saturday Review of Literature. https://my.hrw.com/support/hos/hostpdf/host_text_103.pdf
Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition, 132*(2), 174–180. https://doi.org/10.1016/j.cognition.2014.03.012
Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education, 4*(1), 5–18.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395
Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*(3), 267–272. https://doi.org/10.1111/j.1467-9280.2009.02297.x
Guilbert, D., Sweller, N., & Van Bergen, P. (2021). Emotion and gesture effects on narrative recall in young children and adults. *Applied Cognitive Psychology, 35*(4), 873–889. https://doi.org/10.1002/acp.3815
Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language, 82*(1), 22–29. https://doi.org/10.1016/S0093-934X(02)00009-3
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.
Kelly, S. D., & Tran, Q.-A.N. (2023). Exploring the emotional functions of co-speech hand gesture in language and communication. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12657
Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 62–69. https://doi.org/10.1177/2372732215624707
Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1*(2), 99–113. https://doi.org/10.1177/1754073908100432
Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language, 56*(4), 575–591. https://doi.org/10.1016/j.jml.2006.05.004
Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*(6156), 377–380. https://doi.org/10.1126/science.1239918
Koskinen, P. S., Gambrell, L. B., Kapinus, B. A., & Heathington, B. S. (1988). Retelling: A strategy for enhancing students' reading comprehension. *The Reading Teacher, 41*(9), 892–896.

Levy, R. S., & Kelly, S. D. (2020). Emotion matters: The effect of hand gesture on emotionally valenced sentences. *Gesture, 19*(1), 41–71. https://doi.org/10.1075/gest.19029.lev
Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin & Review, 22*, 1328–1333. https://doi.org/10.3758/s13423-014-0791-2
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. The University of Chicago Press.
McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141*(2), 302–320. https://doi.org/10.1037/a0025250
Morrow, L.M. (1988). Retelling stories as a diagnostic tool. In S.M. Glazer, L.W. Searfoss, & L.M. Gentile (Eds.), *Re-examining reading diagnosis* (pp. 128–149). Newark, DE: International Reading Association
Naveh-Benjamin, M., Guez, J., Hara, Y., Brubaker, M. S., & Lowenschuss-Erlich, I. (2014). The effects of divided attention on encoding processes under incidental and intentional learning instructions: Underlying mechanisms? *Quarterly Journal of Experimental Psychology, 67*(9), 1682–1696. https://doi.org/10.1080/17470218.2013.867517
Overoye, A. L., & Wilson, M. (2020). Does gesture lighten the load? The case of verbal analogies. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.571109
Özder, L. E., Özer, D., & Göksun, T. (2023). Gesture use in L1-Turkish and L2-English: Evidence from emotional narrative retellings. *Quarterly Journal of Experimental Psychology, 76*(8), 1797–1816. https://doi.org/10.1177/17470218221126685
Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.573555
Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*, 36–76. https://doi.org/10.1598/RRQ.38.1.3
Pearman, C. J. (2008). Independent reading of CD-ROM storybooks: Measuring comprehension with oral retellings. *The Reading Teacher, 61*(8), 594–602. https://doi.org/10.1598/RT.61.8.1
Pyers, J. E., Magid, R., Gollan, T. H., & Emmorey, K. (2021). Gesture helps, only if you need it: Inhibiting gesture reduces tip-of-the-tongue resolution for those with weak short-term memory. *Cognitive Science, 45*(1), e12914. https://doi.org/10.1111/cogs.12914
Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-specific manner. *Applied Cognitive Psychology, 29*(5), 680–690. https://doi.org/10.1002/acp.3150
Sadoski, M. (2018). Reading comprehension is embodied: Theoretical and practical considerations. *Educational Psychology Review, 30*(2), 331–349. https://doi.org/10.1007/s10648-017-9412-8
Sadoski, M., & Quast, Z. (1990). Reader response and long-term recall for journalistic text: The roles of imagery, affect, and importance. *Reading Research Quarterly, 25*(4), 256–272.
Sadoski, M., Goetz, E. T., & Kangiser, S. (1988). Imagination in story response: Relationships between imagery, affect, and structural importance. *Reading Research Quarterly, 23*, 320–336.
Saryazdi, R., & Chambers, C. G. (2022). Gesture and reference to objects in the here-and-now. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(4), 583–597. https://doi.org/10.1037/xlm0001035
Siddiqui, A. P., & Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition, 39*, 1387–1400. https://doi.org/10.3758/s13421-011-0125-9
Sivashankar, Y., & Fernandes, M. A. (2022). Enhancing memory using enactment: Does meaning matter in action production? *Memory, 30*(2), 147–160. https://doi.org/10.1080/09658211.2021.1995877
Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition, 16*(4), 934–944. https://doi.org/10.1017/S1366728913000175
So, W. C., Ching, T.H.-W., Lim, P. E., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PLoS ONE, 9*(11), e112543. https://doi.org/10.1371/journal.pone.0112543
Sparapani, N., Connor, C. M., McLean, L., Wood, T., Toste, J., & Day, S. (2018). Direct and reciprocal effects among social skills, vocabulary, and reading comprehension in first grade. *Contemporary Educational Psychology, 53*, 159–167. https://doi.org/10.1016/j.cedpsych.2018.03.003
Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.

Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior, 29*(4), 217–233. https://doi.org/10.1007/s10919-005-7721-y

Tatsumi, I. F., Watanabe, M. (2009). Verbal Memory. In: Binder, M.D., Hirokawa, N., Windhorst, U. (Eds.), *Encyclopedia of neuroscience* (pp. 4176–4178). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_6266

Tournier, I., Jordan, O., & Ferring, D. (2016). Motivation and memory: Impact of emotional content and age relevance on recall. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry, 29*(3), 147–154. https://doi.org/10.1024/1662-9647/a000153

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*, 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 832–842. https://doi.org/10.1037/a0029669

Van Bergen, P., Wall, J., & Salmon, K. (2015). The good, the bad, and the neutral: The influence of emotional valence on young children's recall. *Journal of Applied Research in Memory and Cognition, 4*(1), 29–35. https://doi.org/10.1016/j.jarmac.2014.11.001

van Wermeskerken, F., & 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

Vilà-Giménez, I., & Prieto, P. (2020). Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Developmental Science, 23*(6), 12967. https://doi.org/10.1111/desc

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*(4), 395–407. https://doi.org/10.1016/j.jml.2004.01.002

Ward, C. M., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2016). Effects of age, acoustic challenge, and verbal working memory on recall of narrative speech. *Experimental Aging Research, 42*(1), 97–111. https://doi.org/10.1080/0361073X.2016.1108785

Yiend, J. (2010). The effects if emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion, 24*, 3–47. https://doi.org/10.1080/02699930903205698

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.