**From Thought to Senses: Assessing the Relationship Between the Generation Effect and Multisensory Facilitation**

Michaela Ritchie and Jonathan Wilbiks

Department of Psychology, University of New Brunswick Saint John

**Author Note**

Department of Psychology, University of New Brunswick Saint John

The present study was reviewed by the Research Ethics Board (REB File #012-2024) at

the University of New Brunswick.

Correspondence concerning this article should be addressed to Michaela Ritchie,

University of New Brunswick Saint John, 100 Tucker Park Rd, Saint John, NB E2K 5E2,

Canada.

Email: [michaela.ritchie@unb.ca](mailto:michaela.ritchie@unb.ca)

**Abstract**

The proposed study will investigate whether the generation effect, a memory advantage for self-generated verbal information, is enhanced in multisensory conditions. Such a finding would be consistent with the multisensory facilitation effect, a phenomenon wherein multiple sensory inputs may reduce the cognitive load required to process and respond to co-occurring stimuli from multiple senses. Although extensive research has explored the generation effect, the multisensory aspect of this task has been overlooked. The proposed study aims to determine if the generation effect is modulated by the multisensory facilitation effect. Multisensory facilitation involves the brain’s ability to efficiently process information from diverse sensory sources simultaneously, enhancing cognitive performance. The proposed study will employ a 2 (Task Type: generate, read) X 3 (Sensory Modality: auditory, visual, multisensory) factorial design, utilizing word pair lists and a cued recall test. It is hypothesized that a generation effect will emerge in both unisensory conditions (auditory, visual) and a multisensory condition of encoding, but that the magnitude of the generation effect will be greater for the multisensory encoding condition. This research will contribute to understanding the interplay between self-generation, multisensory processing, and memory, providing insights for educational settings. The proposed study will address a gap in existing literature by examining the multisensory component of generation tasks and will extend the current understanding of multisensory learning by incorporating verbal stimuli as the visual component.

*Keywords:* generation effect, multisensory facilitation effect, multisensory learning, verbal memory, cued recall

**From Thought to Senses: Assessing the Relationship Between the Generation Effect and Multisensory Facilitation**

The generation effect refers to the advantage in memory performance for verbal information when it is generated by one’s own mind, rather than passively read (Slamecka & Graf, 1978). In previous studies, an experimenter would cue participants by presenting them with a rule and asking them to generate a target word. The ability to recall a greater number of generated target words than words that are read, constitutes the generation effect (Slamecka & Graf, 1978). The phenomenon is a robust effect, improving both item memory (word-specific) and context memory (information surrounding the word), compared to reading (McCurdy et al., 2020). Although the generation task often inherently involves a multisensory process of interpreting visual stimuli, and vocalizing the text to produce its auditory representation, this particular aspect of the task has not been explicitly investigated in research. One specific aim of the proposed study is to examine whether the generation effect is moderated by multisensory processing.

**Multisensory Facilitation Effect**

Contemporary research in cognitive psychology, educational psychology, and neuroscience, has shed light on the adaptability of the human brain to efficiently process multisensory information, a capacity finely tuned to accommodate humans’ multisensory way of living (Shams & Seitz, 2008). Although the processing of information from the senses was initially thought as a unisensory process (e.g., Penfield & Rasmussen, 1950; Treisman & Gelade, 1980), the presence of multisensory activations and complex neural networks throughout the brain have since been discovered (Burton et al., 2002; Falchier et al., 2002; Ghazanfar & Schroeder, 2006; Rockland & Ojima, 2003; Röder et al., 2002). This intricate connectivity underlines the efficiency with which the brain integrates and processes information from diverse sensory sources simultaneously, which allows humans to respond and adapt to the multiple sources of information in the environment and reduce cognitive load by pairing congruent stimuli together and processing them using overall fewer cognitive resources than when processing incongruent stimuli (Fahey et al., 2018; Shams & Seitz, 2008). Shams and Seitz (2008) define *congruent stimuli* as “[…] the relationship between stimuli that are consistent with the prior experience of the individual or relationships between the senses found in nature” (p.2). By this definition, arbitrary associations between auditory and visual stimuli will not bring about a multisensory facilitation effect. Moran et al. (2013) elaborated on this proposal: if two stimuli are paired for the purpose of a specific task that are not semantically co-occurring in nature (e.g., an auditory input of a dog barking paired with an image of a lion), it will still be possible to integrate these two stimuli during encoding, but will likely result in only partial or occasional integration at the subconscious level later on, leading to deleterious effects of multisensory input of stimuli in such a scenario (Moran et al., 2013; Shams et al., 2011). Congruent pairings of stimuli are developed within an individual’s semantic network through repeated exposure or experience to the pairings in nature. When such ecologically associated visual and auditory stimuli are presented together, processing of either stimulus will be facilitated as a result of an enhanced, and more elaborate representation of the information across the semantic network and multiple cortical regions in the brain (Moran et al., 2013; Shams et al., 2011). The integrated perception of percepts from multiple senses (i.e., multisensory processing of a single experience) is contingent on *temporal binding*, whereby information acquired from each sense be provided within a certain window of time so that the percepts may be unified as a coherent perception of the stimuli (Meredith et al., 1992; Wallace et al., 1996; Wallace & Stevenson, 2014). Failure to adhere to the temporal binding window could result in the perception of two unintegrated unisensory experiences in lieu of one multisensory percept (Wallace et al., 2004; Stevenson et al., 2012). When auditory and visual cues coincide within a specific temporal window, the result is the perception of a unified and cohesive experience (Wallace et al., 2004). The precise time period during which stimuli from various modalities are prone to perceptual binding is debated as being as little as 50–100 ms (Slutsky & Recanzone, 2001), 200 ms (Lewald et al., 2001) to as great as 800 ms (Wallace et al., 2004). Ultimately, the simultaneous presentation of multiple stimuli ensures that auditory and visual stimuli are perceived concordantly, as the temporal window between stimuli provide statistical cues to their relationship (Wallace & Stevenson, 2014).

A pivotal aspect of multisensory processing is its apparent facilitation effect, wherein unisensory items exhibit enhanced recognition when paired with congruent sensory inputs. Seitz et al. (2006) demonstrated that participants who learned to recognize visual stimuli with congruent auditory stimuli required approximately 60% fewer training sessions than those who studied visual stimuli in isolation. Connections are established through the multisensory learning conditions (Seitz et al., 2006) and any aspect of the multisensory representation can be elicited by merely activating a unisensory component (Friston, 2005; Rao & Ballard, 1999). Thus, the representation of unisensory components is also improved as a result of an enhancement in its representation, which is distributed across a multisensory network (Shams & Seitz, 2008). This is empirically supported by studies that have demonstrated a positive relationship between multisensory processing and improved accuracy, precision, and reaction times in various tasks (Alais & Burr, 2004; Ernst & Banks, 2002; Gingras et al., 2009; Kim et al., 2008). Gingras et al. (2009) focused on the detection and localization of stimuli among animals, demonstrating that multisensory stimuli significantly improved accuracy and precision of responses compared to unisensory stimuli. In their study, No-Go errors were 56% lower and incorrect localization errors were 29% lower for multisensory stimuli, indicating that multisensory integration improves likelihood of response, as well as response accuracy. Ernst and Banks (2002) found that the integrated presentation of visual and haptic information predicts more precise estimates of object properties (e.g., height, size), as compared to when either presentation is permitted alone, suggesting the multisensory processing improves precision of responses. Alais and Burr (2004) explored the ventriloquist effect, which is a specific instance of multisensory integration where visual and auditory cues are combined. When participants had to determine the position of a stimulus, the integration of visual and auditory features was related to improved accuracy and precision of spatial judgments. Taken together, these studies demonstrate that multisensory processing can significantly enhance accuracy, recognition, and precision, in various perceptual tasks by integrating sensory information from different modalities. This integration tends to reduce the overall uncertainty in the sensory estimates, leading to more reliable and precise responses (Alais & Burr, 2004).

Although much of the research on multisensory facilitation is centered around non-verbal stimuli, as has been reviewed in this section, a more recent focus on the verbal stimuli has also come about in the literature. Kim (2021) found that the pairing of e-book reading with the same passage in auditory format (multisensory), improved listening comprehension compared to the e-book (visual) and audiobook (auditory) conditions alone. Interestingly, reading comprehension was only better in the multisensory condition compared to the auditory condition only, but not than the visual only condition. In fact, the auditory condition yielded worse reading comprehension than the visual condition as well (Kim, 2021).

Kalyuga (2000) proposed that the potential negative impact of multisensory input on learning arises from the substantial cognitive load incurred when concurrently processing multiple related sensory sources of information. Consequently, the utilization of redundant sensory modalities may prove ineffective for learning in some cases (Kalyuga, 2000). Other researchers have found that there is no effect of multisensory learning; for example, Rogowsky et al. (2016) observed no differences in comprehension for individuals in either unisensory condition (visual, auditory) compared to the multisensory condition (multisensory).

Diao and Sweller (2007) assessed the impact of simultaneous verbatim spoken presentation alongside written materials on learning to read in English for non-native speakers. Two instructional techniques were compared: reading materials presented in written form only and reading materials presented in written form, accompanied by auditory articulation of the materials. In their study, concurrent spoken presentation hindered understanding and recall of the main ideas of the text compared to the written-only condition (Diao & Sweller, 2007). More recent research contradicts these findings and indicates that multisensory learning improves comprehension among both first language learners (e.g., Kim, 2021) and second language learners (e.g., Mohamed, 2018). For example, Mohamed (2018) found that listening comprehension was improved among individuals in a multisensory learning condition, when learning English as a foreign language. As aforementioned, Kim (2021) observed an advantage for multisensory learning on listening comprehension compared to visual or auditory learning, and an advantage in reading comprehension compared to the auditory learning condition. Further, Liu et al. (2019) found that the multisensory learning of verbal stimuli was related to improved reading comprehension compared to either auditory or visual learning modalities alone. Chang (2009) found that most students perceived that listening in the multisensory mode facilitated easier comprehension, shortened the perceived duration of tasks, enhanced the appeal of learning, and improved their overall sustained attention (Chang, 2009). Kartal and Simsek (2017) reported similar results, and their qualitative findings indicated that visual learning of verbal stimuli accompanied by auditory articulation of that stimuli motivated students to continue learning.

Overall, the literature regarding the advantage of multisensory learning for verbal stimuli is mixed, with many asserting its benefits (Chang, 2009; Kartal & Simsek, 2017; Liu et al., 2019; Mohamed, 2018), and others finding a null (Rogowsky et al., 2016) or negative effect (Diao & Sweller, 2007; Kalyuga, 2000). Given the pervasiveness of multisensory learning in academic settings, these indeterminate results underline the need for further examination.

Multisensory processing of information may allow more efficient processing and response to stimuli in the environment (Wallace, 2009). When presented with congruent stimuli, the brain aims to integrate sensory information from its multiple sources, which may lead to a more accurate perception of the information when it is presented in multiple forms (Wallace, 2009). Although congruent stimuli from multiple senses may allow for more efficient processing of information, redundant multisensory information is also useful (MacLeod, 2010; Wallace, 2009). When the same information is presented to multiple senses at one time (e.g., the visual presentation of a word accompanied by its auditory articulation), the presence of the information from multiple sources increases the probability that the stimuli will be recognized and responded to. Redundant multisensory information may serve to enhance the distinctiveness of the target stimuli from either sense, a process that has been thought to underlie the generation effect (Kinoshita, 1989). Therefore, the simultaneous presentation of cue and target stimuli in a typical generation task may promote its distinctiveness for later recall (Kinoshita, 1989).

**The Generation Effect**

Through a mechanism similar to that proposed for the multisensory facilitation effect, the generation effect may emerge by enhancing memory for contextual information related to the generated target (e.g., the cue word; Mulligan, 2004). Consequently, the mere presentation of a cue word may be sufficient to elicit a multisensory facilitation effect, although this remains to be confirmed. It has been suggested that the generation effect is more pronounced in overt tasks than in covert tasks (McCurdy et al., 2020), despite the stimuli not being congruent, which warrants further investigation.

The concurrent perception of co-occurring stimuli should enhance processing compared to the presentation of multiple stimuli that do not co-occur within the temporal binding window (Meredith et al., 1992; Wallace et al., 1996; Wallace & Stevenson, 2014), and compared to the presentation of stimuli by a single sensory modality alone (MacLeod, 2010; Wallace, 2009). Although not yet directly examined, a typical generation task incorporates an element of auditory production alongside the visual perception of the presented words. The multisensory aspect of the generation task can be characterized as the provision of redundant multisensory information. Despite this, the generation effect is a robust phenomenon, prompting inquiry into whether the multisensory component of the task enhances its effect (increases its magnitude) compared to when an encoding task is exclusively unisensory. For instance, there is limited but intriguing meta-analytic evidence from McCurdy et al. (2020) indicating that covert generation tasks (i.e., tasks completed silently) yield slightly smaller effects as compared to overt generation tasks (i.e., tasks completed aloud). Investigating the impact of multisensory processing on the size of the generation effect is crucial for understanding the factors influencing its occurrence and magnitude.

**The Role of Semantic Relatedness of Stimuli**

Semantically related information shares commonalities in meaning, creating a network of interconnected concepts in the brain (Gardiner & Hampton, 1985; Kinoshita, 1989; Slamecka & Graf, 1978). The distributed network model suggests that the neural representation of a concept is not isolated but involves the activation of a network of nodes that are related to that concept, and the activation of one element spreads across multiple related nodes of information simultaneously to process and retrieve related information more efficiently and successfully (Fuster, 1997; Heisz et al., 2014; Morais et al., 2013). As a result, more elaborate retrieval pathways are established, and the distributed nature of memory allows for multiple pathways to access the same information (Morais et al., 2013). With regard to the recall of information, semantically related pieces of information benefit from a more robust and interconnected representation within this distributed network, leading to enhanced recall compared to the recall of unrelated information, which is a more effortful process that involves the activation of multiple individual regions (Heisz et al., 2014; Morais et al., 2013). Analogous to the mechanisms underlying the multisensory facilitation effect, the activation of one word contributes to the activation of related words, enhancing the efficiency of information retrieval for semantically related stimuli when generated (Gardiner & Hampton, 1985; McCurdy et al., 2020; Slamecka & Graf, 1978).

The semantic activation hypothesis suggests that the brain’s ability to recall information is optimized when that information is semantically related, which has been found to underlie the magnitude of the generation effect (Kinoshita, 1989; McCurdy et al., 2020). For instance, meta-analytic findings suggest that a significant generation effect emerged when words or numbers were used as stimuli, but not when nonwords were used (McCurdy et al., 2020). This finding indicates that meaningful information (e.g., words or numbers) may be a necessary condition to obtain a generation effect (McCurdy et al., 2020). The semantic activation theory of the generation effect posits that the act of generation promotes access to semantic information through a number of mechanisms. First, the act of generating a semantically related response to a given cue requires elaborative encoding of the cue word in attempt to create a connection between the new information and existing knowledge (Craik & Lockhart, 1972). The process of generation is therefore more effortful than merely reading, and this increase in cognitive effort may in part contribute to the benefit of generation in memory (Tyler et al., 1979). Non-semantic generation tasks (e.g., letter transposition) lead to smaller benefits (McCurdy et al., 2020) or non-significant differences (Kinoshita, 1989) in memory. Free recall is also improved by self-generation during encoding for high-meaningful words but not for low-meaningful words (Kinoshita, 1989; Nairne et al., 1985, 1988). Taken together, these findings suggest that self-generation during encoding enhances the distinctiveness of stimuli when the cue and target stimuli are semantically related, which enhances subsequent recall of the generated information (McCurdy et al., 2020; Kinoshita, 1989).

**Hypotheses**

The following hypotheses were derived from a comprehensive examination of relevant literature.

Hypothesis 1 (H1): Based on the robust extant literature on the generation effect, we predict a generation effect, with participants performing better in the generate condition compared to targets that were read, as reflected by better cued recall performance for target items that were generated (e.g., Bertsch et al., 2007; McCurdy et al., 2020; Slamecka & Graf, 1978).

Hypothesis 2 (H2): The generation effect hypothesized in H1 will be more pronounced in multisensory conditions than in unisensory conditions. We expect greater recall of target items when they are generated in the multisensory condition compared to when they are generated in the auditory or visual only condition. This hypothesis was formulated in consideration of the literature that multisensory processing improves recognition rates and accuracy of responses, and the limited but intriguing evidence that overt generation tasks may yield larger effects than covert (e.g., silent) tasks (Alais & Burr, 2004; Ernst & Banks, 2002; Gingras, 2009; McCurdy et al., 2020).

**Method**

**Participants**

The proposed study has been approved by the University of New Brunswick’s institutional review board (REB # 2024-012). Using G\*Power (Faul et al., 2009), and specifying “Repeated measures, between factors, within factors” per our analysis plan proposed below, it was determined that 69 total participants are required to achieve power of 0.90 to detect a medium effect size (*d* = 0.37), which would be consistent with past work on the phenomenon (see Bertsch et al., 2007, for a review). The effect size included in the power analysis reflects the main effect of task type (read versus generate) which is expected to be observed. Actual power of all analyses will be calculated once the data are obtained.

To ensure that the order of the tasks is counterbalanced evenly across participants, a target sample size of 72 participants was determined to be suitable for the proposed study. To ascertain that the quality of participant data, participants will be recruited until we have obtained useable data from 72 participants. Participants will be excluded from analysis if they terminate the study early, there is a failure of any equipment or materials during testing, or any of the tasks are incorrectly completed due to non-compliance with the study’s instructions. For instance, participants are required to respond to all cue words during the encoding phase, providing either a target word, or indicating that they have no response. This ensures that any missing information in the task is due to lack of knowledge rather than accidental omission. Participants who fail to comply with the instructions, whether to due to error or any other reason, will be excluded from the analysis. This decision is based on the need to ensure that any missing responses are attributable solely to a lack of knowledge, rather than to an inadvertent oversight. Further, if a participant vocalizes a response during a reading task, their data will be excluded. Such criteria will be checked by the researcher during the testing phase of the study. Participants will only be included in the analysis if all tasks are appropriately completed as outlined in the procedure.

Participants will be recruited from the researchers’ university population of undergraduate students enrolled in a Psychology course. In compensation for their participation, participants will be awarded one bonus credit toward their Psychology course. Participants must be of at least 19 years of age, and will be asked to report their age, gender, and first language.

**Material**

***Word Pair Lists***

For the proposed study, participants will use a word pair list containing 120 word pairs (synonyms). For each word pair, the target word will be a synonym for the cue word (e.g., chilly-cold). Using a counterbalanced blocked design, some participants will first be exposed to the completed word pair lists under the reading conditions, and the other participants will first receive the incomplete word pair lists—which will contain only the cue words and the first letter of each target word—for the generation tasks. Each participant will complete generation and a reading task in a blocked design which will be counterbalanced across participants.

***PsychoPy***

All task conditions will be created and administered using PsychoPy, a freely available software for conducting psychological studies (Peirce et al., 2019). The auditory and visual stimuli will be presented using this software, in addition to instructions about whether the stimuli are presented through an auditory, visual, or multisensory modality. In the multisensory encoding conditions, the auditory presentation of the word(s) will co-occur with its visual presentation on the screen. The study will be administered in person on a Dell monitor (53.78 X 16.60 cm). Instructions and examples for each task will be presented on the screen at the beginning of the study.

Each word will be displayed in PsychoPy in Arial black font, with a grey background. Most letters are 25.40 mm in height, with the exception of stemmed letters (e.g., d, b, p). For conditions where a visual display is accompanied by an auditory articulation of the cue word, a pre-recorded articulation of that word will accompany its visual presentation and will be synchronized using PsychoPy and presented at full volume. A textbox will be included on a blank grey screen for the visual condition, where participants can input their target word response in real time on the screen. Auditory responses will be recorded with a microphone and saved within the experiment. Where relevant, the screen will be displayed at full brightness.

***Auditory Reading Condition***

An auditory articulation of the cue word, which will be pre-recorded by the researcher, will be played via PsychoPy at full volume. The auditory articulation of the target word will follow, also played via PsychoPy at full volume. A grey screen will accompany the auditory presentations presented for 4 s, followed by a blank interstimulus interval (ISI) of 1 s.

***Visual Reading Condition***

Both the cue word and the target word will be presented on the screen. The display will automatically advance to the next word pair after 4 s (1 s ISI). Participants will be instructed to read each word of the word pair silently, and a researcher will be present to ensure that the participant does not articulate either word aloud.

***Multisensory Reading Condition***

Both the cue word and the target word will be presented on the screen for 4 s (1 s ISI) and accompanied simultaneously by an auditory articulation of each word which will be pre-recorded by the researcher.

***Auditory Generation Condition***

An auditory articulation of the cue word, which will be pre-recorded by the researcher, will be played via PsychoPy at full volume. The auditory articulation of the first letter of the anticipated target word will follow, also played via PsychoPy at full volume. A grey screen will accompany the auditory presentations. The participant will have 4 s to vocalize their response (i.e., the target word), which will be recorded in PsychoPy. After 4 s (1 s ISI), the screen will automatically advance to the next word pair.

***Visual Generation Condition***

The cue word, and the first letter of the anticipated target word, will be presented on the screen. Participants will have 4 s to generate their response by typing out the target word, which will then be displayed on the screen. The display will be advanced to the next screen after the allotted time and a 1 s ISI.

***Multisensory Generation Condition***

The cue word will be presented on the screen and accompanied simultaneously by an auditory articulation of that word which will be pre-recorded by the researcher. Participants will have 4 s (1 s ISI) to generate their response by typing out the target word, which will then be displayed on the screen, and should then be articulated aloud by the participant.

***Distraction Task***

Participants will be given a number on the screen, and will be instructed to continually subtract seven from the given number for 30 seconds. Each of their responses will be recorded in PsychoPy.

***Cued Recall Test***

An untimed cued recall test will be administered following the distraction phase. One at a time, a cue word will be displayed on the screen, in a randomized order. For each cue word that is displayed, the participant must type in the associated target word, and their response will appear on the monitor. Once they have submitted their response, they will be advanced to respond to the next cue word after a 1 s interstimulus interval.

**Procedure**

Following a within-subjects blocked design, participants will be exposed to 60 word pairs under the reading condition, and 60 word pairs under the generation condition, and the order in which they are exposed to either task will be counterbalanced across participants. Once their informed consent has been received, they will be given explicit instructions in PsychoPy about how each task will be completed. Instructions will vary based on whether the participant is assigned to the auditory condition, visual condition, or multisensory condition. Before participants begin the tasks, there will be a brief practice round to ensure that participants are familiar with the expectations of each task. Participants will then encode a total of 120 word pairs, half of which will be self-generated. Following the encoding phase of all 120 word pairs, participants will engage in a brief distraction task before completing a cued recall test on all word pairs. The presentation of the cue words will be in randomized order.

**Proposed Analysis**

The following analysis is proposed: A 2 (Task Type: generate, read) X 3 (Sensory Modality: auditory, visual, multisensory), factorial analysis of variance (ANOVA) will be conducted, with task type assessed within subjects, and sensory modality between subjects.

H1 specifies that a main effect of task type will be observed, with participants performing better in the generate condition compared to targets that were read, as reflected by greater recall for target items that were generated. To examine whether there is support for this hypothesis, we will assess whether there is a main effect of task type in the omnibus test.

Given that H2 specifies a generation effect will emerge across all sensory conditions, but will be larger in magnitude for the multisensory condition, an ordinal interaction will be assessed (Elias, 2006; Park, 2023). First, we will analyze whether an ordinal interaction is present between sensory condition (coded as: 1 = auditory, 2 = visual, 3 = multisensory) and task type (coded as 0 = read, 1 = generate). Then, the 2 X 3 factorial ANOVA specified above will be conducted with the data to assess whether sensory condition moderates the strength of the relationship between task type (read/generate) and cued recall (outcome of interest). It is expected that a generation effect will emerge at all levels of the sensory condition, but that there will be a significant interaction between sensory modalities and task type, such that the magnitude of the generation effect will increase in the multisensory condition, but the direction of the effect will not change at any level of sensory modality. To ensure that significant interaction effects are not an artifact of the scale of measurement, a logit transformation will be applied onto the dependent variable and if the interaction does not survive the transformation, it will be considered an artifact of the original scale of measurement (Labaronne, 2023; Wagenmakers, 2012).

**Implications**

Despite having a vast body of literature to rely on when examining the effect of self-generation on memory under a host of different conditions (for reviews of this literature, see Bertsch et al., 2007; McCurdy et al., 2020), the multisensory component inherent to most generation tasks has been neglected as an area of focus in this research. To help understand the mechanisms underlying the generation effect, assessing the multisensory component of the task will be necessary to elucidate whether generation tasks are exemplary of the multisensory facilitation effect, or whether the generation effect is as pronounced when the multisensory component of generation tasks is removed. The importance of clarifying the multisensory role of encoding under such conditions extends beyond the purpose of understanding the mechanisms which underlie the generation effect. Beyond this, the findings of this study will contribute to the growing literature on multisensory learning, and will contribute evidence toward the debate regarding when a multisensory facilitation effect may be observed. Further, existing research supporting the multisensory facilitation effect examines the relationship between auditory and visual processing using auditory stimuli and images as visual stimuli. In the proposed study, verbal stimuli will be used as the visual component, which will have implications for the role of semantic awareness in the multisensory facilitation effect for verbal visual stimuli paired with congruent auditory stimuli. The proposed study is the first of its kind to examine the role of multisensory learning in the generation effect. For a summary of the study design and planned analyses, see Table 1.

**References**

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14* (3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23*(1), 3–38. <https://doi.org/10.1163/187847510X488603>

Barutchu, A., Freestone, D. R., Innes-Brown, H., Crewther, D. P., & Crewther, S. G. (2013). Evidence for enhanced multisensory facilitation with stimulus relevance: An electrophysiological investigation. *PLoS One, 8*(1), e52978. <https://doi.org/10.1371/journal.pone.0052978>

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. <https://doi.org/10.3758/bf03193441>

Burton, H., Snyder, A. Z., Diamond, J. B. and Raichle, M. E. (2002). Adaptive changes in early and late blind: A fMRI study of verb generation to heard nouns, *Journal of Neurophysiology 88*, 3359–3371. <https://doi.org/10.1152/jn.00129.2002>

Chang, A. C. S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652–663. <https://doi.org/10.1016/j.system.2009.09.009>

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and instruction, 17*(1), 78–88. <https://doi.org/10.1016/j.learninstruc.2006.11.007>

Ernst, M. and Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433. <https://doi.org/10.1038/415429a>

Fahey, S., Charette, L., Francis, C., & Zheng, Z. (2018). Multisensory integration of signals for bodily self-awareness requires minimal cognitive effort. *Canadian Journal of Experimental Psychology, 72*(4), 244–252. [https://doi.org/10.1037/cep0000152](https://psycnet.apa.org/doi/10.1037/cep0000152)

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience 22,* 5749–5759. <https://doi.org/10.1523/JNEUROSCI.22-13-05749.2002>

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate multisensory integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*(6), 1247–1257. <https://doi.org/10.1111/j.1460-9568.2009.06688.x>

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods 41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147*(3), 332–343. <https://doi.org/10.1007/s00221-002-1262-y>

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society, 360*, 815–836. <https://doi.org/10.1098/rstb.2005.1622>

Fuster, J.M. (1997). Network memory. *Trends in Neurosciences, 20*(10), 451–459. <https://doi.org/10.1016/S0166-2236(97)01128-4>

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 732–741. <https://doi.org/10.1037/0278-7393.11.1-4.732>

Ghazanfar, A.A., & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends in Cognitive Science 10*, 278–85. <https://doi.org/10.1016/j.tics.2006.04.008>

Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *Journal of Neuroscience, 29*, 4897–4902.

Gottfried, J. A., Smith, A. P. R., Rugg, M. D. and Dolan, R. J. (2004). Remembrance of odors

past: human olfactory cortex in cross-modal recognition memory. *Neuron, 42*, 687–695. <https://doi.org/10.1016/S0896-6273(04)00270-3>

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425–2430. <https://10.0.4.102/science.1063736>

Heisz, J. J., Vakorin, V., Ross, B., Levine, B., & McIntosh, A. R. (2014). A trade-off between local and distributed information processing associated with remote episodic versus semantic memory. *Journal of Cognitive Neuroscience, 26*(1), 41–53. <https://doi.org/10.1162/jocn_a_00466>

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation.* Wiley, New York, 1966.

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australasian Journal of Educational Technology, 16*(2). <https://doi.org/10.14742/ajet.1829>

Kartal, G., & Simsek, H. (2017). The effects of audiobooks on EFL students’ listening comprehension. *The Reading Matrix: An International Online Journal, 17*(1), 112–123.

Kim, N. Y. (2021). E-Book, Audio-Book, or E-Audio-Book: The Effects of Multiple Modalities on EFL Comprehension. *English Teaching, 76*(4), 33–52.

Kim RS, Seitz AR, Shams L (2008) Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE 3*(1), e1532. <https://doi.org/010.1371/journal.pone.0001532>

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*(5), 563–571. <https://doi.org/10.3758/bf03197079>

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research, 121*(1–2), 69–79. <https://doi.org/10.1016/s0166-4328(00)00386-7>

Liu, H., Cao, S., & Wu, S. (2019). An experimental comparison on reading comprehension effect of visual, audio and dual channels. *Proceedings of the Association for Information Science and Technology, 56*(1), 716–718. <https://doi.org/10.1002/pra2.148>

Mayer, R. E. (2001). *Multimedia Learning.* New York: Cambridge University Press.

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review, 27,*1139–1165. <https://doi.org/10.3758/s13423-020-01762-3>

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906–908. <https://doi.org/10.1038/35038085>

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. <https://doi.org/10.1038/264746a0>

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition.* Psychology Press.

Meredith, M. A., Wallace, M. T., & Stein, B. E. (1992). Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: Multisensory properties of the tecto-reticulo-spinal projection. *Experimental Brain Research, 88*(1), 181–186. <https://doi.org/10.1007/bf02259139>

Mohamed, M. M. K. (2018). Using audiobooks for developing listening comprehension among Saudi EFL preparatory year students. *Journal of Language Teaching and Research, 9*(1), 64–73. <http://dx.doi.org/10.17507/jltr.0901.08>

Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the structure of semantic memory. *Cognitive Science, 37*(1), 125–145. <https://doi.org/10.1111/cogs.12013>

Moran, Z. D., Bachman, P., Pham, P., Cho, S. H., Cannon, T. D., & Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisensory Research, 26*(6), 581–592. <https://doi.org/10.1163/22134808-00002436>

Nyberg, L., Habib, R., McIntosh, A. R. and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval, *Proceedings of the National Academy of Sciences, 97*, 11120–11124. <https://doi.org/10.1073/pnas.97.20.11120>

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man; a clinical study of localization of function.* Macmillan.

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). Cognitive load theory.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience, 2*, 79–87. <https://doi.org/10.1038/4580>

Rockland K.S., Ojima, H. (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology 50,* 19–26. <https://doi.org/10.1016/S0167-8760(03)00121-1>

Röder, B., Stock, O., Bien, S., Neville, H. and Rosler, F. (2002). Speech processing activates visual cortex in congenitally blind humans, *European Journal of Neuroscience 16*, 930–936. <https://doi.org/10.1046/j.1460-9568.2002.02147.x>

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does modality matter? The effects of reading, listening, and dual modality on comprehension. *Sage Open, 6*(3). <https://doi.org/10.1177/2158244016669550>

Rolls, E. T., Treves, A. and Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research, 114*, 149–162. <https://doi.org/10.1007/PL00005615>

Seitz, A., Dinse, H. (2007) A common framework for perceptual learning. *Current Opinion of Neurobiology 17*(2), 148–153. <https://doi.org/10.1016/j.conb.2007.02.004>

Seitz, A.R., Kim, R., & Shams, L. (2006) Sound facilitates visual learning. *Current Biology, 16*, 1422–1427. <https://doi.org/10.1016/j.cub.2006.05.048>

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>

Shams, L., Wozny, D. R., Kim, R. & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology, 2*, 264. <https://doi.org/10.3389/fpsyg.2011.00264>

Sheffert, S. M., & Olson, E. (2004). Multisensory speech facilitates voice learning. *Perception in* *Psychophysics, 66*, 352–362. <https://doi.org/10.3758/BF03194884>

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604. <https://doi.org/10.1037/0278-7393.4.6.592>

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7–10. <https://doi.org/10.1097/00001756-200101220-00009>

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to multisensory illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517–1529. <https://doi.org/10.1037/a0027339>

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A, 36*(1), 51–74. <https://doi.org/10.1080/14640748408401503>

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. <https://doi.org/10.1016/0010-0285(80)90005-5>

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature, 383*(6597), 254–256. <https://doi.org/10.1038/383254a0>

Von Kriegstein, K. & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, 1809–1820. <https://doi.org/10.1371/journal.pbio.0040326>

Wallace, M., Roberson, G., Hairston, W., Stein, B., Vaughan, J., & Schirillo, J. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158*(2). <https://doi.org/10.1007/s00221-004-1899-9>

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123. <https://doi.org/10.1016/j.neuropsychologia.2014.08.005>

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246–1266. <https://doi.org/10.1152/jn.1996.76.2.1246>

Wheeler, M. E., Petersen, S. E. and Buckner, R. L. (2000). Memory’s echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*, 11125–11129. <https://doi.org/10.1073/pnas.97.20.11125>

Wyk, B. C., Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., Lee, S. M., & Pelphrey, K. A. (2010). Cortical integration of audio–visual speech and non-speech stimuli. *Brain and Cognition, 74*(2), 97–106. <https://doi.org/10.1016/j.bandc.2010.07.002>

**Table 1**

*Study Design*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question** | **Hypothesis**  (broken down in manuscript under “Hypotheses”) | **Sampling plan** | **Analysis Plan** | **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis** | **Interpretation given different outcomes** | **Theory that could be shown wrong by the outcomes** |
| Will task type predict cued recall performance? | Hypothesis 1 (H1): We predict a generation effect, with participants performing better in the generate condition compared to targets that were read, as reflected by better cued recall performance for target items that were generated | A sample size of 72 participants was determined using G\*Power to achieve a power of 0.90 to detect a medium main effect size of task type, consistent with effect sizes observed in past work. | The following analysis is proposed: A 2 (Task Type: generate, read) X 3 (Sensory Modality: auditory, visual, multisensory), factorial analysis of variance (ANOVA) will be conducted, with task type assessed within subjects, and sensory modality between subjects.  H1 specifies that a main effect of task type will be observed, with participants performing better in the generate condition compared to targets that were read, as reflected by greater recall for target items that were generated. | Effect size determined based on prior research suggesting a medium effect size for the hypothesized differences in encoding condition. | A main effect of task type in favour of generation > read, indicates a generation effect. A main effect in favour of the reading condition indicates a negative generation effect. A null effect indicates that encoding was not improved through the process of generation in the tasks. | **Theoretical framework:** Explores how the generation effect and multisensory facilitation impact memory.  **Generation effect:** self-generated info > passive reading (Slamecka & Graf, 1978).  **Multisensory facilitation:** multiple senses > better learning (Shams & Seitz, 2008).  **Potential Challenge to Existing Theories:**  Given the robustness of the generation effect in extant research, it is not anticipated that a null or negative generation effect will be observed. If either of such outcomes are observed, it may most likely indicate that the task at hand did not provide sufficient opportunity to encode the target material in the learning trials for the generation tasks.  If H2 is not supported, this may indicate that marginal effects of overt generation tasks (i.e., reading the cue word and generating the target aloud) that have been reported in the past (e.g., McCurdy et al., 2020), may be better explicated by chance error variance, or other factors. Ultimately, lack of support for H2 indicates that the multisensory component of the generation task does not substantially improve retention of the target items compared to unisensory generation tasks. |
| Will the sensory encoding condition moderate the generation effect? | Hypothesis 2 (H2): The generation effect hypothesized in H1 will be more pronounced in multisensory conditions than in unisensory conditions. We expect greater recall of target items when they are generated in the multisensory condition compared to when they are generated in the auditory or visual only condition. | See above. | H2 predicts a generation effect across all sensory conditions, with a greater effect in the multisensory condition. An ordinal interaction will be assessed between sensory condition (auditory = 1, visual = 2, multisensory = 3) and task type (read = 0, generate = 1). A 2 x 3 factorial ANOVA will then test if sensory condition moderates the relationship between task type and cued recall. The generation effect is expected at all sensory levels, with a larger effect in multisensory, but no change in effect direction. To ensure the interaction is not due to measurement scale, a logit transformation will be applied to the dependent variable. Survivability and transformation guidelines were derived from Labaronne, 2023; Wagenmakers , 2012. | See above. Actual power will also be calculated after the data are collected. Given the limited but intriguing evidence that overt generation tasks may yield larger effects than covert (e.g., silent) tasks, we expect to detect at least a medium effect size.  Survivability and transformation guidelines were derived from Labaronne, 2023; Wagenmakers , 2012. | A significant ordinal interaction once the data are transformed indicates support for H2. Non-significant findings indicates insufficient support for the hypothesis that the generation effect will be more pronounced in multisensory encoding conditions than for either unisensory encoding condition. |

**Appendix**

**List of Word Pairs**

1. Woods-forest
2. Pleasant-nice
3. Require-need
4. Wealthy-rich
5. Sad-upset
6. Cash-money
7. Chilly-cold
8. Shack-cabin
9. Wail-weep
10. Terminate-finish
11. Filthy-dirty
12. Identical-same
13. War-battle
14. Safe-secure
15. Start-begin
16. Shove-push
17. Ship-boat
18. Cab-taxi
19. Woman-lady
20. Throw-toss
21. Employ-hire
22. Mad-angry
23. Advance-progress
24. Informal-casual
25. Attain-achieve
26. Intent-purpose
27. Elaborate-complex
28. Rescue-save
29. Advertise-promote
30. Difficult-hard
31. Water - Fluid
32. River - Stream
33. Clock - Timer
34. Car-vehicle
35. Beverage - Drink
36. Chuckle - Laugh
37. Waltz - Dance
38. House - Home
39. Smile - Grin
40. Fashion - Style
41. Earth - World
42. Pony - Horse
43. Sound - Noise
44. Furnace-heater
45. Night - Evening
46. Drizzle - Rain
47. Amazing – Incredible
48. Liquor-alcohol
49. Huge-Enormous
50. Radiant-Bright
51. Tranquil-Peaceful
52. Carve-Slice
53. Danger-hazard
54. Idea-thought
55. Help-assist
56. Murder-kill
57. Book-novel
58. Love-admire
59. Peculiar-unusual
60. Rabbit-bunny
61. Instructor-teacher
62. Little-small
63. Despise-hate
64. Under-below
65. Connect-link
66. Physician-doctor
67. Grass-lawn
68. Mistake-error
69. Enemy-opponent
70. Harm-damage
71. Infant-baby
72. Depart-leave
73. Robber-thief
74. Paste-glue
75. Drowsy-sleepy
76. Quiet-silent
77. Steer-drive
78. Come-arrive
79. Lose-fail
80. Expand-stretch
81. Delicate-fragile
82. Untrue-False
83. Question-inquiry
84. Admit-confess
85. Vacant-empty
86. Close-shut
87. Broad-wide
88. Respond-Answer
89. Stop-quit
90. Gift-present
91. Above-over
92. Occur-happen
93. Illness-disease
94. Scream-yell
95. Stone-rock
96. Approve-accept
97. Illicit-illegal
98. Humble-modest
99. Plausible-possible
100. Gain-acquire
101. Hefty-heavy
102. Thankful-grateful
103. cheerful-joyful
104. Agitated-annoyed
105. Unsure-doubtful
106. Cautious-careful
107. Vulnerable-sensitive
108. Blame-accuse
109. Combat-fight
110. Divorce – separate
111. Comprehend-understand
112. Obligate-force
113. Artificial-fake
114. Beautiful-gorgeous
115. Barbecue-grill
116. Waiter-server
117. Spouse-partner
118. Duvet-comforter
119. Priest-pastor
120. Speak-talk