

# Multimodal communication in instructional settings: An investigation of the functional roles of gestures and arrows

*Comunicación multimodal en contextos educativos: Una investigación de los roles funcionales de los gestos y las flechas*

**Cengiz Acartürk**

JAGIELLONIAN UNIVERSITY  
 POLAND  
 MIDDLE EAST TECHNICAL UNIVERSITY  
 TURKEY  
 acarturk@acm.org

**Melda Coskun**

MCMASTER UNIVERSITY  
 CANADA  
 MIDDLE EAST TECHNICAL UNIVERSITY  
 TURKEY  
 coskunks@mcmaster.ca

**Serap Emil**

MIDDLE EAST TECHNICAL UNIVERSITY  
 TURKEY  
 semil@metu.edu.tr

**Recibido:** 13-VII-2021 / **Aceptado:** 22-10-2021

**DOI:** 10.4067/S0718-09342021000300867

## Abstract

In daily communication and instructional settings, arrows and spontaneous gestures usually convey similar spatial information to interlocutors. They can visualize spatial information in 2D or 3D space based on shared conceptual and spatial representations. The present study investigates the relationship between spontaneous gestures and arrows from the perspective of generation and comprehension in instructional settings. We introduce a corpus of arrows in written multimodal text and a corpus of gestures generated during communication through multimodal instructional material. We also report an experimental investigation that used corpus material to study the impact of gestures on comprehension learning. A comparison of the corpora reveals functional similarities and differences between arrows and gestures and their complementary role in multimodal communication. The findings suggest that arrows and gestures may share a common conceptual space during communication, having a significant impact on comprehension and learning.

**Key Words:** Modality, deixis, arrows, gestures, multimodality.

## Resumen

En la comunicación diaria y en entornos educativos, las flechas y los gestos espontáneos suelen transmitir información espacial similar a los interlocutores. Ellos pueden visualizar información espacial en el espacio 2D o 3D basándose en representaciones conceptuales y espaciales compartidas. El presente estudio investiga la relación entre los gestos espontáneos y las flechas desde la perspectiva de la generación y la comprensión en

contextos de enseñanza. Introducimos un corpus de flechas en un texto multimodal escrito y un corpus de gestos generados durante la comunicación a través de material instructivo multimodal. También presentamos una investigación experimental que utilizó un corpus para estudiar el impacto de los gestos en el aprendizaje de la comprensión. La comparación de los corpus revela similitudes y diferencias funcionales entre las flechas y los gestos y su papel complementario en la comunicación multimodal. Los resultados sugieren que las flechas y los gestos pueden compartir un espacio conceptual común durante la comunicación, teniendo un impacto significativo en la comprensión y el aprendizaje.

**Palabras Clave:** Modalidad, deixis, flechas, gestos, multimodalidad.

## INTRODUCTION

The term ‘modality’ has different uses in various fields of research. ‘Sensory modality’ usually addresses senses of various types, such as vision, audition, and touch (Calvert, Spence & Stein, 2004). On the other hand, ‘communication modality’ addresses speech, gestures, body posture, and emotions conveyed by facial expressions (Holler & Levinson, 2019; Wahlster, 1991). A relevant term is ‘representational modality,’ frequently used for text, figures, and diagrams of various types (Acartürk, 2014; Bernsen, 1994; Bateman, 2008, cf. ‘multimodal documents’; Parodi, Julio & Recio, 2018). Accordingly, the research on multimodal communication emphasizes one aspect of multimodality or the other, depending on the research context and the specific target of the study. In the present study, we assume that multimodality is a cover term that represents various forms of communication (e.g., gestures and speech), representations (e.g., figures and text), and senses (e.g., vision and touch). In daily settings, such as instructional one in a classroom environment, we employ those multiple aspects of multimodality for specific purposes (Bernsen, 1994, 2008). Multimodal aspects of communication comprise instructional settings, also having some unique benefits on comprehension and learning (Ainsworth & Scheiter, 2021; Eilam & Poyas, 2008; Gates, 2018; Mayer, 2005; Ping, Church, Decatur, Larson, Zinchenko & Godin-Meadow, 2021). For instance, teaching and learning involve frequent use of gestures, bodily postures, and gaze contact besides verbal communication. The instructional material usually involves multiple representations, such as text and figures aiming to transform the material into comprehensible chunks in order to facilitate learning, depending on the level of learner (cf. ‘intermodal connections’ in Parodi, et al., 2018). Nevertheless, as Parodi (2012) states, the research on specific types of multimodal interaction, such as a systematic investigation of multisemiotic descriptions of written texts and diagrams, is scarce. More generally, large-scale corpus studies are needed beyond the case studies (Bateman, 2021).

A significant challenge in the study of multimodality in daily settings, such as educational, is the complexity of the environment. Bernsen (1994) presented a taxonomy of multimodal representations divided into visual and graphical qualities/ vision, sound qualities/ auditory abilities, and tactile and kinesthetic qualities/ touch.

The taxonomy also presented an alternative dimension for classifying multimodal representations (*viz.* common sense' classification) by emphasizing multiple ways of classifying modalities. The classification included six categories, namely language, picture of something, non-visual pictures or analog representations, graphs, representations, and explicitly rendered structuring of information.

A further challenge is that multimodal aspects of communication form a virtually Gestalt whole, presenting an environment where the whole is larger than the sum of its parts. The richness in the variety of nonverbal representations (e.g., figures) and the flexibility in their integration to linguistic expressions enables achieving a systematic understanding of multimodal comprehension and, more generally, multimodal learning. For example, from an information processing perspective, a significant challenge is the integration of information in different representational modalities, such as text and figures. A coherent understanding of the instructional material may require effective use of limited cognitive resources, which is a common challenge for learners (Habel & Acartürk, 2007; cf. cognitive load theory by Sweller, 1994; also see multimedia learning theory by Mayer, 2005).

The complexity of the environment in natural, multimodal communication settings has also led researchers to focus on specific characteristics of a modality, such as gesture production (e.g., Goldin-Meadow, 1999). Those limitations have resulted in divergent findings in the literature on the impact of various aspects of multimodality in daily settings. The present study is also subject to those limitations due to the high-level approach to the complex research domain. Therefore, we approach the problem by narrowing down the question of multimodality by seeking the common characteristics among the concepts that allow coherent multimodal communication. In particular, we focus on the common concepts that share a conceptual space while expressing themselves in multiple modalities. Accordingly, our approach emphasizes the cognitive aspects of multimodal communication by making the operational assumption about the need for abstract conceptual and spatial structures to achieve successful communication (Habel & Acartürk, 2007). The basic proposal is that those conceptual and spatial structures allow humans to construct a coherent understanding of the environment, eventually leading to comprehension and learning by constructing memory representations coherently. Accordingly, multimodal communication is established on a coherent construction of them in the mind/brain. Recently, our knowledge has been limited about the inner characteristics of those structures, whether they are inherently multimodal or loosely connected modular structures, which is also beyond the scope of the present study. Instead, we present a partial answer to the big question of multimodality, suggesting that gestures and specific diagram types (in this case, arrows) share common characteristics when investigated in terms of their functionality in discourse.

Gestures and arrows may both exhibit ‘representational’ functions in that they depict a veridical representation of the object of interest. A relevant common concept shared by gestures and arrows is their ‘deictic’ functionality. In verbal communication, deixis refers to words and phrases that cannot be fully interpreted without additional contextual information. It is usually expressed by deictic expressions, such as ‘this’, ‘that’, ‘these’, and ‘those’. Deictic expressions are essential in developing our conceptualization of the references in the external world (Levinson, 2006). Visualizations are cognitive tools, like written language, designed for augmenting the capacity of the human mind (Tversky, Agrawala, Heiser, Lee, Hanrahan, Doantam, Stolte & Daniel, 2020). Deixis has been recognized as fundamental to visualization (Hill & Hollan, 1991). Deictic gestures have also been categorized as a major gesture type (McNeill, 1992, 2005). From the perspective of multimodal communication, what makes those functions interesting is their expressions as a nonverbal communication modality by various means. In diagrammatic representations, designers use deictic cues, such as arrows, for pointing at a particular part of the figure or the whole figure. Similar deictic functionality is frequently observed in daily communication instances, where an interlocutor uses a pointing gesture. Non-verbal expressions of deixis are also observed in gaze direction and gaze gestures, when the interlocutor points at an object of interest in the environment, even in head gestures.

The commonalities among different forms of multimodality are interpreted as shared, underlying conceptual and spatial structures that express themselves in various modalities, in a complementary or a redundant manner. Gestures may be conceived as virtual diagrams in the air, whereas diagrams may be viewed as permanent traces of gestures on a surface. Therefore, gestures and diagrams visualize the message in communication by employing a set of content-free geometric forms (e.g., arrows) that convey information utilizing their geometric properties (Tversky, Jamalain, Giardino, Kang & Kessell, 2013). The present study proposes a cognitive approach that allows sketching as a simplified framework to study multimodality under various forms. In particular, we aim to understand the role of two non-verbal modalities, namely gestures and arrows, regarding their functional roles in multimodal communication. For this, we first present a survey of the use of arrows. Next, we introduce a corpus of arrows in the written multimodal text, including excerpts from academic journals, newspapers, magazines, and textbooks. Then, we introduce a corpus of gestures and arrows generated during communication through multimodal instructional materials. Finally, we report an experimental investigation to analyze how gestures influence learning outcomes in an instructional setting. We expect an interaction between the modalities to share a common conceptual space throughout communication in instructional settings, known to be rich in both modalities. The following section presents the background for the study and a review of the relevant work.

## 1. Background and relevant work

Gestures may substitute for the speech using conventionalized forms, or gestures may accompany speech through spontaneous hand movements. Speech-substituting gestures (e.g., American Sign Language, ASL) may facilitate communication as the primary modality by assuming a language-like form. On the other hand, speech-accompanying gestures (henceforth, spontaneous gestures) play a significant role in communication without assuming a language-like form (Goldin-Meadow, 1999). In terms of their functions in multimodal communication, gestures and diagrams are different in that gestures are momentary actions, whereas diagrams are relatively permanent visual representations. They also exhibit functional differences, as well as similarities.

In the present study, we focus on the relationship between spontaneous gestures and arrows rather than the differences between two. In particular, gestures and diagrams share a general two-fold division: Representational and deictic functions. McNeill (1992) stated that representational gestures represent picturable aspects of information, like diagrammatical elements. On the other hand, deictic gestures point out entities in the domain of discourse, as accomplished by certain diagrammatic elements, such as arrows. A literature review shows that the similarities between gestures and diagrammatic elements have been studied by focusing on different aspects of the relationship. Heiser, Tversky and Silverman (2004) asked the participants to work in pairs on a map to find the most efficient route to rescue a certain number of injuries. They used gestures in a typical diagrammatic element form; pointing a place (e.g., dot), tracing a path between places (e.g., line), and tracing a place (e.g., box). Another study (Alaçam, Habel & Acartürk, 2013) focused on spontaneous gestures in verbal descriptions of statistical graphs. The findings revealed that gestures highlight certain aspects of the information represented by statistical graphs. For example, when describing time-series line graphs, the participants used vertically oriented gestures to refer to an increase or a decrease in the value of the domain variable. In contrast, they used bidirectional horizontal gestures referring to durative states of the value (e.g., the average temperature over time). In time-series bar graphs, directional gestures accompany verbal descriptions of trend information similarly. More generally, both gestures and diagrams address various spatial aspects of communicative content in multimodal communication.

Moreover, the studies investigating the relationship between speech and gestures emphasize the role of spatial cognitive processes. For instance, Kita and Özyürek (2003) propose that gesture generation pertains to spatio-motoric processes interacting with speech production. These findings suggest that alongside the language, gestures and diagrams may be conceived as communication modalities that externalize common conceptual and spatial mental representations (Acartürk, 2014).

The previous research has revealed further findings of how spontaneous gestures play a role in conveying information about a task by analyzing systematic relationships between gestures and task-specific aspects (e.g., Alibali & Nathan, 2007; Goldin-Meadow, 1999; Ping et al., 2021). Spontaneous gestures have been proposed to enrich communication by supporting information in a second modality, usually conveying information not represented in speech. It has also been claimed that gestures promote communication by externalizing thoughts that are not formed well enough to express verbally (Goldin-Meadow, 1999). This role of gesture in communication has been conceived to reduce the speaker's cognitive load (Rauscher, Krauss & Chen, 1996). The facilitating effect of gestures in communication has been shown not only for objects available in the visual scene during the speaker's gesturing but also for objects not present in the scene (Ping & Goldin-Meadow, 2010). The findings suggest that gestures and communication modalities employ conceptual-level structures that go beyond a single modality.

Methodologically, we investigate the relationship between spontaneous gestures and arrows from the perspective of generation and comprehension. We follow Duncan, Cassell & Levy (2007) annotation practice based on McNeill's (1992, 2005) classification of gestures for the analyses. McNeill (1992) identified an ontology of speech-accompanying gestures, frequently employed in gesture research. McNeill proposes that spontaneous gestures are classified into four types: iconic, metaphoric, deictic, and beat gestures. Iconic gestures and metaphoric gestures visualize picturable aspects of speech. Accordingly, they bear a structural resemblance to what they represent. Deictic gestures consist of pointing movements of the arm or the fingers, and they usually refer to concrete entities in the environment. Thus, deictic gestures function as a bridge between speech and the entity referred to by speech. Iconic gestures and metaphoric gestures have been conceived as closely related to the information content of the communication. On the other hand, beat gestures are only speech-related rhythmic movements, which have not been conceived as related to communicative content. Therefore, in the present study, we focus on the first three gesture types by keeping beat gestures out of the scope of the study. Moreover, we merge iconic and metaphoric gestures under a single type of gesture, namely 'representational' gestures, since they serve similar purposes in multimodal communication.

The usual communication setting for gesture use is a conversation. For example, McNeill's (1992) descriptions of gestures were contextualized in a dialogue environment, in which a participant used spontaneous gestures in verbal communication as an example. Similarly, Alibali (2005) examined a conversation as a contextual environment where a man held two bags with his hands, needed to drop the bags, and freed his hands when the conversation started, thus highlighting the role of gestures in daily communication. In the present study, we employed an instructional setting, also a gesture-rich environment, as our experimental investigations revealed.

As for diagrams, they are also crucial elements of instructional settings. Learning with multiple representations (i.e., pictorial illustrations and text) has been examined in multiple domains of research since the past several decades (e.g., Glenberg & Langston, 1992; Hegarty & Just, 1993; Zahner & Corter, 2010), in particular, instructional science (e.g., Mayer, 2005; Nys, Hickmann & Gyselinck, 2018; Richter, Scheiter & Eitel, 2017). The results unveiled the facilitating role of diagrams in learning directed to the circumstances. More generally, instructional science has been a significant domain of research for gesture research and diagrams. Regarding gestures, the previous research on learning uncovered that nonverbal behavior has a significant role in instructional settings (e.g., Goldin-Meadow, 1999; McNeill, 1992; Rueckert, Church, Avila & Trejo, 2017). For example, teachers' nonverbal behavior might influence learning processes by conveying teachers' attitudes toward students. Research on gesture in instructional science has varying topics, from foreign language education (e.g., Sime, 2006) to mathematics education (Gerofsky, 2011). Klerleft (2007) emphasized the significant role of gestures in technical instruction. They found that the interaction between a preschool teacher and a child primarily relied on the gestures of the instructor along with the simple verbal expressions, like 'here' and 'it', rather than complex verbal expressions or other nonverbal expressions, such as making eye contact during a story creation task in which they performed together at the computer. The enhancing effect of gestures on learning is not only limited to human gestures, but a computer-generated instructional avatar's gestures can also have a positive effect on learning (Cook, et al., 2017).

The medium of communication (aka. the mode of instruction) is another factor that impacts learning.<sup>1</sup> Educational technology design is fundamental to efficiently implementing basic learning principles (Graesser, Sabatini & Li, 2021). Nevertheless, the studies conducted in the past decade have revealed divergent findings. Chen and Sager (2011) pointed out three functions of presentation technologies: to show the process of problem-solving, to support multisensory learning, and to keep a record of instructional material. They compared the use of tablet PCs with the earlier generation of presentation techniques (overhead projectors, blackboards, or presentation slides). They found that tablet PCs appear to support all these three pedagogical functions more than the previous presentation techniques, either alone or in combination. On the other hand, Skibinski, DeBenedetti, Ortol-Bloch and Hines (2015) argue that blackboards are cheap, flexible, and simple tools to help lecturers in teaching. However, blackboards are limited by human anatomy since the lecturer faces away from the classroom and the lecturer's body blocking the students' view. Instead, they suggest using a light board projection system, allowing a lecturer to write on a transparent plate and reflect on a bigger screen through a projector. This system allows a lecturer's writing to appear on the screen and allows students to see a lecturer's gestures and facial expressions, helping to convey the information. The findings in the relevant literature are divergent, in particular, about the role of the mode of instruction. Since the mode of instruction is

an indispensable part of an instructional setting, we specified the mode of instruction as a factor in the present study. In particular, we employed three modes of instruction allowing sketching and gestures: ‘board-and-board marker’, ‘paper-and-pencil’, ‘tablet-and-pen’.

In summary, the review of the relevant work reveals functional similarities between gestures and certain diagrammatic constituents under various experimental settings. We investigate those similarities in instructional settings in the present study, since they are rich in gestures and diagrams. For this, we designed two corpora, one for arrows and the other for gestures and arrows. Then, we conducted an experimental investigation on the role of gestures in comprehension and learning by selecting the video recordings from the corpus to show that gestures, as non-verbal constituents in multimodal communication, significantly impact multimodal communication. In the next section, we present the methodological framework of the study.

## **2. Methodological framework**

Concerning the background and related work in the literature, the central research question of the present study focuses on the role of arrows in multimodal text and its projection on gestures. Accordingly, the research included three parts. First, through the survey, we examined the presence of arrows in multimodal text in instructional materials. Secondly, an experiment was carried out to explore arrows and gestures in three modes of instruction (board, paper, and tablet). Finally, the role of gestures in learning was explored through experimental design. Accordingly, we present the methodological framework of the study in three sections. The first section includes the procedures for the survey on a corpus of arrows in multimodal instructional material. The second section presents the methodology for designing a corpus of gestures and arrows generated during communication through instructional materials. We discuss similarities and differences between the two corpora regarding the functional roles of arrows and gestures. Finally, we report the experimental investigation on the role of gestures in learning, conducted using selected videos from the second corpus.

### ***2.1. A corpus of arrows in multimodal text***

This section aims to present a corpus of arrows in multimodal text based on a survey to identify the types of arrows in instructional material. For this, we established a collection of arrow samples from various sources, including academic journals, newspaper magazines, and textbooks, classified based on their functional similarities. A total of 1,015 arrow samples (i.e., sketches) were collected from nine sources: five books, two academic journals, and two weekly magazines (approximately 5,800 pages). The domain of the source material included physics, engineering, chemistry, biology, cognitive science, psychology, and popular science. Journals and magazines were selected randomly not to limit the survey to books.

In diagrammatic comprehension studies, diagrams are typically to be analyzed subject to their syntactic and semantic characteristics. Kosslyn (1989) conducted a similar syntactic-role and semantic-role analysis to understand charts and graphs, defined basic-level constituents, and their interrelations. In Kosslyn's (1989) framework, the syntactic analysis describes the inner and outer organization of the constituents without considering what they represent. The semantic analysis focuses on the meaning of the display, which varies based on the configuration of the constituents. Finally, the pragmatic level focuses on the meaning of the display, which is different from direct semantic interpretations of symbols.

In the present study, we follow a similar approach by considering both syntactic and semantic characteristics of the arrows, though by emphasizing their communicative functions. Each arrow sample was classified based on its contextual function. We identified more than ten functions of arrows. Those groups were then further identified in three major classes of arrows: 'Representational,' 'Relational,' and 'Deictic.' The arrows, which did not fit into any class, were classified as 'Other.'

## **2.2. A corpus of gestures and arrows in instructional settings**

Coşkun and Acartürk (2015) reported an experimental study in which 24 content specialist instructors (15 female, mean age  $M = 30.0$ ,  $SD = 3.7$ , range 25-38) presented their subject knowledge in math or science. The experiment design consisted of three within-subject conditions, namely the mode of instruction: 'board,' 'paper,' and 'tablet.' The instructors used a whiteboard in the board condition, an A4-size paper in the paper condition, and a SMART Podium ID422w interactive pen display in the tablet condition. They were asked to select a topic to teach in the three different instructional modes. The order of the instructional mode was randomized across the participants to counterbalance variance. They were told that their video recordings would be watched by a group of learners later without providing any information about our interest in gestures or arrows. The instructors were requested to complete the session in 5 to 10 minutes. They selected topics from four math subjects (Geometry, Algebra-I, Pre-calculus, Trigonometry) and three general science subjects (Physics, Chemistry, Biology). We manually annotated spontaneous gestures and arrows in the video recordings. Figure 1 shows example gestures of specific types from the video recordings.



**Figure 1.** Example gestures from the video recordings. A representational gesture illustrating the “pull and push” action (left), a deictic gesture pointing to the particle on the table (middle), a beat gesture which is a speech-related rhythmic movement (right). The photos are used upon the permission of the participant.

### ***2.3. An experimental investigation on the role of gestures in an instructional setting***

The previous research has already acknowledged that gestures are an indispensable aspect of instructional communication. Long ago, Goldin-Meadow, Kim and Singer (1999) identified two functions of gesture in classroom interaction: (a) gestures can inform a listener about the motivation and attitudes of the speaker during the communication between teacher and pupils and (b) gestures play a role in the acquisition of the content of the lesson itself. The first function of gestures may affect a student’s attitude towards learning, whereas the second function of gestures has a direct impact on cognitive aspects of learning. The previous findings support the idea that gestures are crucial for learning. Based on this motivation, we conducted an experimental investigation. The selected video lectures of the gesture and arrow corpus were presented to students for comprehension and learning. We focused on gesture types, using gesture-rich videos and gesture-poor videos in the experimental stimuli, leaving arrow-rich videos and arrow-poor videos for further research.

#### ***2.3.1. Participants, materials, and design***

A total of 61 undergraduate students (43 female) participated in the experiment for course credit. The participants were divided into two groups. Each group was presented with two short video lectures (4-7 minutes) from one of the two gesture conditions (gesture-rich videos vs. gesture-poor videos). The group watching the gesture-rich videos consisted of 27 undergraduate students (19 females,  $M_{age} = 22.62$ ,  $SD = 2.01$ ) and the group watching the gesture-poor videos consisted of 34 students (24 females,  $M_{age} = 20.60$ ,  $SD = 5.26$ ). Each group also watched two videos from each of the two instructional conditions (board vs. tablet). Therefore, the design included one within-subject factor (the mode of instruction: board vs. tablet) and one between-subject factor (the gesture condition: rich vs. poor). We did not include the paper-and-pencil instructional model as a condition since Coşkun and Acartürk (2015) revealed a more profound contrast between the board and tablet conditions in terms of gesture types.

The design of the experiment required selecting four videos, one for each condition (1-gesture-rich tablet, 2-gesture-rich board, 3-gesture-poor tablet, 4-gesture-poor board, with 12 videos in each condition) from the recordings in the gesture and arrow corpus (Section 2.2.2). The richness of the videos was determined based on the number of gestures they contained; the functions of the gestures were disregarded. For this purpose, the videos were ordered by the number of gestures used by the instructor. The two (of 24) videos at the top of the list and the two (of 24) at the bottom were omitted, assuming that those videos might be the outliers. The following two videos from the top and bottom were selected for the gesture-rich (one for the board condition and the other for the tablet condition) and two gesture-poor (again, one for the board condition and the other for the tablet condition) conditions, respectively. The topics of the videos included the following: ‘what is speed,’ ‘the law of conservation mass,’ ‘electric field,’ ‘the factors affecting resistance’). The topic of the study was treated as a random factor in the experiment.

In the instructional sessions, the participants were informed that the goal of the experiment was to study how educational videos influenced learning so that they did not focus either on their gestures or the mode of instruction. The videos were presented in the classroom on a projector screen in a single session for each group of participants. A multiple-choice pretest questionnaire (eight questions) was administered before presenting the videos to the participants to identify their prior knowledge on the topic of the study. The learning assessment was conducted by using the same questionnaire at the end of the session. The post-test questionnaire consisted of the same questions but with different order of the multiple-choice answers. The order of the videos was also counterbalanced between the two groups of participants. The students spent approximately five minutes on the pretest questionnaire and slightly shorter for the post-test questionnaire. The total duration of the session was approximately 50 minutes.

### **3. Results**

#### ***3.1. Arrows in multimodal text***

Table 1 shows the major arrow classes in terms of their distribution in the corpus of arrows in multimodal text. All arrow classifications were initially performed by the authors of this study. Then, randomly selected 24.6% of the arrows (250 of 1015) were classified independently by a second coder, informed about the classification system employed in the survey study for reliability analysis. Cohen’s kappa was used to calculate inter-rater reliability between the two annotators. The inter-rater agreements of annotations were calculated as .75. The value between .61 and .81 indicates substantial inter-rater agreement based on Landis and Koch (1977).

**Table 1.** The distribution of the arrows according to the classes.

	Type	Count
Representational	605	59.61%
Relational	378	37.24%
Deictic	14	1.38%
Other	18	1.77%
Total	1,015	100%

**Representational Arrows.** Representational arrows were the most frequently observed arrow type in the corpus. Representational arrows convey information about the picturable aspects of the content. Therefore, the syntactic properties of representational arrows depend on the aspects of the represented entity. All components of the arrow may be of importance for an appropriate functional interpretation. For instance, the angle of the arrowhead may represent resistance (Frutiger, 1989), the length of the arrow may show the distance between two entities, or the shape of the arrow body may represent the trajectory of the motion. In addition, they depict visuospatial aspects of the content (e.g., motion, distance, axes) by using space and geometric properties. We classified the following arrow functions as ‘representational’ (see Figure A1 in the Appendix for samples):

- Motion, Path, or Manner: Arrows in this group represent a function of a mechanism (e.g., to convey twisting forces effect in a mechanism), the behavior of physical concepts that may not be observed with the naked eye (e.g., electric field, ray), or dynamic behavior (e.g., force, pressure, velocity). Path and manner of motion can also be represented by manipulating the arrow's body slot (shape).
- Distance: Arrows in this group refer to a distance between two entities. The distance is interpreted as the distance between the starting point of the arrow and the end. The distance is typically assessed by an additional attribute, for instance, letter symbols located nearby, or by a dashed line that highlights the boundaries of the entities.
- Object Representation: Arrows in this group are analogical demonstrations of an object. Arrow constituents represent a part of the represented entity by preserving spatial configurations. For instance, the length of the arrow may represent the length of the represented object. Similarly, the start point of the arrow may refer to the beginning of the represented entity.
- Ordinal or Ratio Scale Representation: Arrows on an axis or arrows on a number line are classified as the members of this group.
- Symbolic: Arrows in this group are used for symbolic representations, such as an arrow in a photon symbol, which shows the trajectory of its traveling path.
- Vector: Arrows in this group represent directed magnitudes, such as velocity or force. The direction of the vector is essential in vectors.

**Relational Arrows.** Relational arrows were the second most frequently observed arrow type in the collected samples. Relational arrows depict the asymmetric relation between two entities such as closure, similarity, consist-of relation, and part-of relation. Syntactically (cf. Kosslyn, 1989), relational arrows connect two entities. Therefore, they require two nodes if conceived as a node-link style diagrammatic representation. In addition, a label may be attached nearby the arrow to indicate what kind of relationship exists between the entities. The relational arrows that convey various aspects of the relationship between entities are presented below. This classification should be taken as a snapshot-view of different subclasses of the relational arrows, varying depending on their use of context (see Figure A2 in Appendix A):

- Textual Link: Textual link arrows attach text to an entity through labeling.
- Abstract Causal Link: The arrow shows an abstract causal relation between the entities.
- Transform: Transform arrows depict a change in sequence, such as a temporal change.
- Zoom: Zoom arrows represent a close-up and detailed view of a particular part of the represented entity.
- Process: Process arrows are used in diagrams, such as Unified Modelling Language (UML) diagrams and flowcharts.

**Deictic Arrows.** Deictic arrows were the least frequently observed arrow type in the collected samples of instructional material. Deictic arrows point an entity by directing the viewer's attention towards the direction of the arrow. Hence, the direction and its extension are the crucial properties of a deictic arrow. The main difference between a deictic arrow and a relational arrow is that the latter requires two concrete items in a representation to show the connection between them. In contrast, a deictic arrow connects one item with the other, which is not represented by inference (see Figure A3 in Appendix A).

In summary, a survey of collected samples of instructional materials (1,015 arrow tokens) returned numerous possible functions of arrows, which were further sized down for the present study. In the next section, we present a corpus, based on the gestures and arrows used by the participants in an instructional setting.

### ***3.2. Gestures and arrows in instructional settings***

In this section, we present the results for gesture types and arrow types obtained in the analysis of the video recordings from the instructor participants, aiming to reveal the communicative role of spontaneous gestures and arrows in live instructional settings.

### 3.2.1. Annotation of spontaneous gestures

The methodology by Duncan et al. (2007) was used for speech-accompanying, spontaneous gesture annotation, based on the approach suggested by McNeill (1992, 2005). In the analysis, gestures without speech were not included. The corpora consisted of 3525 gestures. The type of gesture was determined based on a set of guidelines, as presented below. Table 2 shows the distribution of the gestures in the study (see Appendix B for sample video recording frames).

**Table 2.** The distribution of the gestures according to the types and the modes of instruction (Coskun & Acartürk, 2015). The numbers show the mean number of instances of each type in the corpus.

	Deictic	Beat	Representational
Board	31.25 (13.58)	20.79 (20.16)	3.54 (3.71)
Paper	29.91 (16.22)	17.00 (14.55)	2.13 (2.19)
Tablet	26.75 (11.63)	14.38 (10.03)	1.13 (1.23)

\*Zero values were included in the analysis. The numbers in parentheses show Standard Deviation (SD) associated with the mean deviation between the participants.

**Representational Gestures.** Gestures include picturable aspects of semantic content. The numerous aspects of gestures, such as form, direction, and motion trajectory of the hands, play a particular role in displaying the semantic content (Alibali & Nathan, 2007). The gestures displaying an action - a concept of a drawing or an object - were regarded as representational gestures. An example can be the speaker's hand moved around a particle representation on the board to show an electric field. Besides, the gestures pointing a direction were accepted as representational gestures, such as the instructor moving his hand forward while showing an upward palm to represent an outward direction.

**Deictic Gestures.** The pointing movements that indicate physical, available objects and physically unavailable ones at the time of gesturing were regarded as deictic gestures. In the present study, gestures that pointed to a particular part or all parts of a diagram were annotated as deictic gestures. More specifically, the words about objects such as 'this, here, there' were often accompanied by deictic gestures. Furthermore, the gestures that traced a particular part or all parts of a drawing and demonstrated a group of text, drawings, or formulas were classified as deictic gestures. Speakers referring to things inside a particular area to create an impression on the audience or the boundary marked by the speaker's hand movements are some examples of deictic gestures, which are typically performed with the extended finger or hand in the board session, whereas mainly demonstrated by the pen in paper and tablet sessions.

**Beat Gestures.** These can be defined as speech-related rhythmic hand movements, usually classified in two forms: discrete and continuous (McNeill, 1992). Beat gestures are demonstrated through a syllable, word, or clause that are stressed and disappear right after the utterance. In other words, beat gestures in continuous form were used

throughout the speech (word, clause, and sentence). Particular and series of hand movements – circular or continuous ones - can lead to these gestures.

### 3.2.2 Annotation of arrows

The instructors also generated arrows during their teaching sessions. The corpora consisted of 547 arrows, classified into three types by following the annotation scheme presented in section 2.1: ‘representational,’ ‘relational,’ and ‘deictic.’ Briefly, the arrows that represented motion (e.g., movements of particles in a cap), force (e.g., the pressure on the piston), physical representations (e.g., electric field), and processes (e.g., increase, decrease) were assumed to be representational arrows depicting the relation between two entities were annotated to be relational arrows. Finally, the arrows that were used for pointing purposes were annotated as deictic arrows. Each arrow token was assumed to belong to only one of the classes. Table 3 shows the distribution of the arrows in the study.

**Table 3.** The distribution of the arrows according to the types and the modes of instruction. The numbers show the mean number of instances of each type in the corpus.

	Relational	Representational	Deictic
Board	7.00 (7.50)	0.96 (2.61)	0.21 (0.59)
Paper	6.92 (7.37)	1.67 (3.77)	0.13 (0.45)
Tablet	6.33 (6.90)	1.33 (2.91)	0.25 (0.61)

\*Zero values were included in the analysis. The numbers in parentheses show Standard Deviation (SD) associated with the mean deviation between the participants.

All the gesture annotations and the arrow annotations were initially performed by the two authors of the study. All annotated gestures and arrows fit one of the three categories presented above. A total of 25% of all the randomly selected gestures and 20% of all arrows were annotated independently by a second coder, who was informed about the study, for reliability analysis. The inter-rater reliability among two coders was calculated with Cohen’s kappa, where the inter-rater agreements of initial annotations were found as .75 and 1.0 for arrow annotation. Based on Landis and Koch (1977), the value above .61 indicating substantial inter-rater agreement, and the value between .81 and .99 indicating almost perfect agreement were used. Upon discussion, the coders re-annotated the gesture data, and the agreement was found .96.

In summary, this section presented a corpus of gestures and arrows based on the data collected and partially reported in Coşkun and Acartürk (2015). In the written multimodal corpus, presented in Section 2.2.1, the representational arrow category was the most frequently observed one, followed by the relational arrow and the deictic arrow categories. In the second corpus, in which we investigated the gestures and the arrows generated by a group of instructors, we observed a different distribution of the arrows than the initial corpus of written multimodal text. The most frequently observed category of arrows was the relational, followed by representational and deictic

categories. Given that those arrows are gesture and speech-accompanying arrows, this difference between the two corpora reflects the impact of the generation context in live instruction.

Overall, the findings of the corpora reveal a complementary role of arrows and gestures in multimodal communication. Furthermore, gestures had a significant impact on the distribution of arrow types in the instructional material, suggesting a division of labor between gestures and arrows in a generation context. Nevertheless, those findings do not provide robust evidence for the role of gestures in a learning context. In the next section, we present the experimental investigation on the role of gestures in a natural instructional setting by using a set of gesture-rich and gesture-poor instructional material from the gesture and arrow corpus presented in the previous section.

### ***3.3. The experimental investigation on the role of gestures in an instructional setting***

One of the challenges in the design of the experiment was the selection of the videos, in particular, the dependency of the selection to the topic of study in the videos. We assumed that the topic of study was a random factor in the experiment design, since it was virtually impossible to find gesture-rich videos of the same topic of study. This resulted in significant differences in the pretest scores of the participants between the conditions. We resolved this issue by considering the pretest scores as a covariate in the analysis of post-test scores.

A multilevel logistic regression model was conducted to account for variability across participants and variability in the difficulty of topics. The log of the odds of each correctly responded question was predicted from an interaction of the gesture condition (gesture-rich vs. gesture-poor videos) and the mode of instruction (board vs. tablet) with a binary fixed effect of the pretest score (depending on whether the question was correctly responded or not in the pretest). Participants, questions, and topics were introduced to the model as random intercepts. This model revealed that the pretest score had a positive effect on the posttest results, ( $\beta_{Pre} = 0.69$ ,  $\chi = 3.42$ ,  $p < .05$ ). There was also a significant effect of the mode of instruction, with the board setting resulting in better understanding than the tablet setting ( $\beta_{Tablet} = -1.15$ ,  $\chi = -2.54$ ,  $p < .05$ ). Findings also demonstrated a significant interaction between the gesture condition and the mode of instruction, ( $\beta_{Tablet-GestureRich \times Board-GesturePoor} = 2.40$ ,  $\chi = 3.71$ ,  $p < .05$ ). Specifically, the tablet with more gestures led to a better understanding, whereas the richness of gestures in the board setting had a reverse effect. However, the gesture condition was not significant, ( $\beta_{GestureRich} = -0.64$ ,  $\chi = -1.2$ ,  $p = .23$ ). Overall, the participants watching gesture-poor videos correctly responded to only 64% of post-test questions. In contrast, the participants watching gesture-rich videos correctly responded to 72% of post-test questions (Table C1 in Appendix C).

## 4. Discussion

Gestures and arrows convey similar information content, in particular, spatial aspects of the communicative content. We assumed that gestures and arrows carry similar functional roles in multimodal communication. To investigate the similarities between them, besides the differences, we developed two corpora. We employed a frequently used gesture classification from the literature (McNeill, 1992, 2005; Duncan et al., 2007). The literature is relatively limited in the classification of arrows according to their communication functions. Therefore, we developed a classification for the functional roles of arrows by considering the similarities and differences compared to gestures. A comparison of the two corpora revealed the complementary role of gestures and arrows, in terms of the impact of gestures on the distribution of the arrows, besides their functional similarities, proposed as an operational assumption in the present study. The two corpora presented a generation perspective, one in written multimodal text, the other in a live instructional setting. We then conducted an experimental investigation to study the comprehension perspective by presenting gesture-rich and gesture-poor videos to participants. The results of the experimental investigation (Section 2.3) showed that the richness of gestures had a significant, positive impact on learning outcomes.

Overall, the board setting had a significant impact on participants' learning of the material compared to the tablet setting. Interestingly, gestures' more substantial facilitating effect was observed when the instructor used a tablet rather than the board. This finding is surprising given that the gesture and arrow corpus (Section 2.2.2, also see Coşkun & Acartürk, 2015) shows that the instructors exhibited a tendency to use fewer gestures when using the tablet device. A likely reason may be that the participants of the experimental investigation were influenced by the use of a tablet as a novel technology compared to the classical setting, which may have resulted in more attachment to the topic. Another likely reason is the influence of more vigorous split attention between information sources (Kalyuga, Chandler & Sweller, 1999) in the board setting compared to the tablet setting. In other words, the gestures and the content of the instruction, as the two information sources, are closer to each other in the tablet setting compared to the board setting.

Previous studies reported similar results related to the split attention effect. More recently, Post, Van Gong, Paas and Zwaan (2013) found that asking learners to use gestures alongside the gestures in the instructional animation might not facilitate learning but even hinder particularly those with low competence in that domain. A likely reason for this finding is that students might have paid more attention to gestures and fewer resources left to comprehend the subject. According to Baylor and Kim (2009), a single nonverbal communication mode accompanying verbal expressions might improve learning than those with two or more. Specifically, they investigated the

effectiveness of facial expressions and gestures in two different tasks: a technical task where participants were taught to use a software program (procedural learning outcome); the other task aimed to change the emotional attitudes of the listener towards the content taught (attitudinal learning outcome). The results showed that gestures lead to a better understanding of procedural instruction when facial expressions are missing. In contrast, facial expressions are more effective in influencing the listener emotionally when gestures are not used. In this case, the presence of two nonverbal expressions may split a listener's attention regarding the purpose of the task. Further research is necessary to understand better the likely sources of the findings obtained in the present study.

## **CONCLUSIONS**

Spontaneous gestures are an integrated part of speech. The relationships between verbal and nonverbal communication modalities, such as language and gestures or language and diagrams, have been subject to intense research for the past several decades. However, further research is needed to extend research on multimodal communication by focusing on relationships between nonverbal modalities and gestures under specific contextual environments. We selected an instructional setting as a contextual environment for our study due to its richness in the number and variety of diagrammatic representations and gestures. The results reveal a relationship between gestures and arrows, suggesting a trade-off between the two modalities.

The arrow symbol is one of the ubiquitous elements of visual communication. An arrow can be used for various purposes depending on context through its geometric and Gestalt properties such as causal and temporal relations, motion and force concepts in physics, direction, and label. Deictic arrows play a crucial role in communication, in contrast to their occasional use in written material. Gestures may also represent a wide range of information content, as arrows do. The information conveyed by gestures may overlap with those of arrows since they share similar conceptualization aspects. Accordingly, gestures and arrows can convey the same information by using similar diagrammatic features. The relationship between spontaneous gestures and arrows may be conceived as supporting evidence that spontaneous gestures and diagrams share a conceptual space for representation during multimodal communication.

In future work, we plan to address a set of limitations in the present study, which resulted primarily due to the assumptions in identifying specific arrow types, specific gesture types, and the constraints imposed by practical aspects of experiment design. A major limitation of the present study was that the participants were asked to teach the topic of their choice to a hypothetical classroom audience. The topic of choice likely impacts the results, which is virtually impossible to control experimentally. An alternative approach might be to use the same topic; however, that would require a between-subject design. Future research should aim at improving the experimental control of the stimuli presented to the participants. A natural classroom setting might

have influenced the patterns of gesture use. Also, the gestures in the tablet condition and the gestures in the paper-and-pencil conditions were constrained by the sitting position of the participants. The design strategy for keeping the natural bodily position of the presenters as natural as possible (i.e., the sitting position in the tablet condition and the paper-and-pencil condition vs. the standing position in the board condition) resulted in different nature of gesture patterns.

Another limitation of the study was the set of assumptions made in identifying the types of arrows. We largely classified arrows based on their functional similarities for the present study. There are alternative methods of classification in the literature. For instance, Kurata and Egenhofer (2008) first classified arrows according to their syntactic properties (e.g., the number of required components), and then they re-analyzed the semantic roles of the arrows.

Regarding tablet use, we conjecture that future intrusion of tablet use in daily life may result in changes in gesture patterns in tablets. The present study also conceives spontaneous gestures as a domain-dependent action. Therefore, a closer look at using domain-dependent spatial terminology (e.g., spatial terms in topics in geometry vs. topics in physics) is necessary.

One of the major findings of this research was that a more interdisciplinary approach needs to be developed in teaching and learning in educational settings. The research that solely focuses on the pedagogical aspect of using technology provides a limited scope for studying the learning process. A cognitive approach to educational settings would allow researchers to understand the complexity of learning better. This need is emphasized by Parodi (2015), stating that the developments in the genre theory could be applied both to linguistics and to the pedagogical approaches of language teaching and learning in academic contexts. This approach was efficiently demonstrated by Boudon and Parodi (2014), which utilized Parodi's (2010) methodological proposal with seven artifacts: Statistics Complex, Diagram, Formula, Graphic, Icon, Illustration, and Table. Their results showed that the high-frequency multisemiotic artifacts - table formula and graphics - in the Economics discourse support the communicative pedagogical purpose of the genre textbook. Further research should extend the analysis by addressing the relationship among speech, gestures, and diagrams and the relationship between gestures and diagrams.

## **REFERENCES**

- Acartürk, C. (2014). Towards a systematic understanding of graphical cues in communication through statistical graphs. *Journal of Visual Languages and Computing*, 25(2), 76-88. DOI: 10.1016/j.jvlc.2013.11.006.

- Ainsworth, S. E. & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science*, 30(1), 61-67.
- Alaçam, Ö., Habel, C. & Acartürk, C. (2013). Gesture and language production in communication through bar graphs. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (Eds.), *Proceedings of the 35<sup>th</sup> Annual Conference of the Cognitive Science Society* (pp. 1714-1719). Austin, TX: Cognitive Science Society.
- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, 5(4), 307-331.
- Alibali, M. W. & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In R. Goldman, R. Pea, B. Barron & S. J. Derry (Eds.), *Video Research in the Learning Sciences* (pp. 349-365). Mahwah, NJ: Erlbaum.
- Bateman, J. A. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. New York: Springer.
- Bateman, J. A. (2021). Growing theory for practice: Empirical multimodality beyond the case study. *Multimodal Communication*, pp. 000010151520210006. DOI: <https://doi.org/10.1515/mc-2021-0006>
- Baylor, A. L. & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior*, 25, 450-457.
- Bernsen, N. O. (1994, 2008). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers*, 6(4), 347-371.
- Bernsen, N. (2008). Multimodality Theory. In D. Tzovaras (Eds.), *Multimodal User Interfaces. Signals and Communication Technologies* (pp. 5-29). Berlin, Heidelberg: Springer. DOI: [https://doi.org/10.1007/978-3-540-78345-9\\_2](https://doi.org/10.1007/978-3-540-78345-9_2).
- Boudon, E. & Parodi, G. (2014). Multisemiotic artifacts and academic discourse of Economics: Knowledge construction in the Textbook genre. *Revista Signos. Estudios de Lingüística*, 47(85), 164.
- Calvert, G., Spence, C. & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. Boston: MIT Press.
- Chen, F. & Sager, J. (2011). Effects of tablet PC use in the classroom on teaching and learning processes. *Journal of Learning in Higher Education*, 7(2), 55-67.

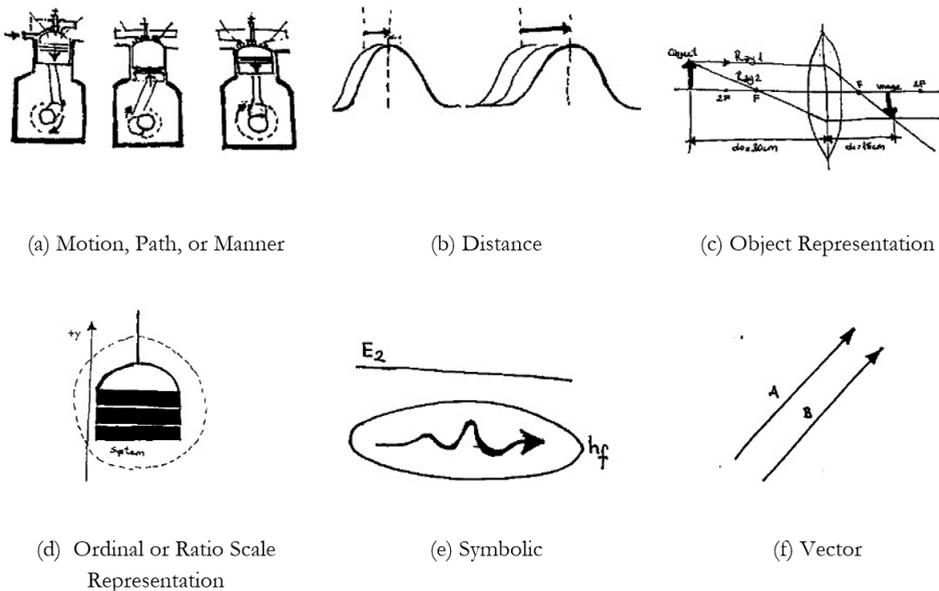
- Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J. & Popescu, V. (2017). Hand gesture and mathematics learning: lessons from an Avatar. *Cognitive Science*, 41(2), 518-535.
- Coşkun, M. & Acartürk, C. (2015). *Gesture production under instructional context: The role of mode of instruction*. In D. C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (Eds.), *Proceedings of the 37<sup>th</sup> Annual Meeting of the Cognitive Science Society* (pp. 459-464). Austin, TX: Cognitive Science Society.
- Duncan, S. D., Cassell, J. & Levy, E. T. (Eds.). (2007). *Gesture and the dynamic dimension of language: Essays in Honor of David McNeill (Vol. 1)*. Amsterdam, Philadelphia: John Benjamins.
- Eilam, B. & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction*, 18(4), 368-378.
- Frutiger, A. (1989). *Signs and symbols*. New York: Van Nostrand Reinhold.
- Gates, P. (2018). The importance of diagrams, graphics and other visual representations in STEM teaching. In R. Jorgensen & K. Larking (Eds.), *STEM Education in the Junior Secondary* (pp. 169-196). Singapore: Springer.
- Gerofsky, S. (2011). Mathematical learning and gesture: Character viewpoint and observer viewpoint in students' gestured graphs of functions. *Gesture*, 10(2-3), 321-343.
- Glenberg, A. M. & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, 31(2), 129-151.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11), 419-429.
- Goldin-Meadow, S., Kim, S. & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, 91(4), 720-730.
- Graesser, A. C., Sabatini, J. P. & Li, H. (2021). Educational psychology is evolving to accommodate technology, multiple disciplines, and Twenty-First-Century skills. *Annual Review of Psychology*, 73(1). DOI: 10.1146/annurev-psych-020821-113042.
- Habel, C. & Acartürk, C. (2007). On reciprocal improvement in multimodal generation: Co-reference by text and information graphics. In I. van der Sluis, M. Theune, E. Reiter & E. Kraemer (Eds.), *Proceedings of MOG 2007: The workshop on multimodal output generation* (pp. 69-80). UK: University of Aberdeen.

- Hegarty, M. & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32(6), 717-742.
- Heiser, J., Tversky, B. & Silverman, M. (2004). Sketches for and from collaboration. In J. S. Gero, B. Tversky & T. Knight (Eds.), *Visual and Spatial Reasoning in Design III* (pp. 69-78). Key Centre of Design Computing and Cognition, University of Sydney.
- Hill, W. C. & Hollan, J. D. (1991). *Deixis and the future of visualization excellence*. In Proceedings of the 2<sup>nd</sup> Conference on Visualization '91 (pp. 314-320). IEEE Computer Society Press.
- Holler, J. & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639-652.
- Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13(4), 351-371.
- Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16-32.
- Klerfelt, A. (2007). Gestures in conversation: The significance of gestures and utterances when children and preschool teachers create stories using the computer. *Computers & Education*, 48(3), 335-361.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185-225.
- Kurata, Y. & Egenhofer, M. (2008). The arrow-semantics interpreter. *Spatial Cognition & Computation*, 8(4), 306-332.
- Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159.
- Levinson, S. C. (2006). *Deixis*. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 97-121). UK: Blackwell Publishing.
- Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. UK: Cambridge University Press.
- McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

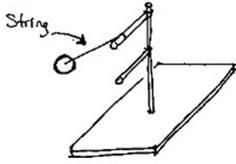
- Nys, M., Hickmann, M. & Gyselinck, V. (2018). The role of verbal and visuo-spatial working memory in the encoding of virtual routes by children and adults. *Journal of Cognitive Psychology*, 30(7), 710-727.
- Parodi, G. (2010). Research challenges for corpus cross-linguistics and multimodal texts. *Information Design Journal*, 18(1), 69-73.
- Parodi, G. (2012). Genre organization in specialized discourse: Disciplinary variation across university textbooks. *Discourse Studies*, 16(1), 65-87.
- Parodi, G. (2015). Variation across university genres in seven disciplines: A corpus-based study on academic written Spanish. *International Journal of Corpus Linguistics*, 20(4), 469-499.
- Parodi, G., Julio, C. & Recio, I. (2018). When words and graphs move the eyes: The processing of multimodal causal relations. *Journal of Eye Movement Research*, 11(1). DOI: <https://doi.org/10.16910/jemr.11.1.5>.
- Ping, R., Church, R. B., Decatur, M. A., Larson, S. W., Zinchenko, E. & Goldin-Meadow, S. (2021). Unpacking the gestures of chemistry learners: What the hands tell us about correct and incorrect conceptions of stereochemistry. *Discourse Processes*, 58(3), 213-232.
- Ping, R. & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34(4), 602-619.
- Post, L. S., Van Gog, T., Paas, F. & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, 29(4), 1450-1455.
- Rauscher, F. H., Krauss, R. M. & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4), 226-231.
- Richter, J., Scheiter, K. & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19-36.
- Rueckert, L., Church, R. B., Avila, A. & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications*, 2(1), 1-6.
- Skibinski, E., DeBenedetti, W., Ortoll-Bloch, A. & Hines, M. (2015). A blackboard for the 21<sup>st</sup> Century: An inexpensive light board projection system for classroom use. *Journal of Chemical Education*, 92(10), 1754-1756.

- Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *Review of Applied Linguistics in Language Teaching*, 44(2), 211-230.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312.
- Tversky, B., Jamalian. A., Giardino. V., Kang. S. & Kessell. A. (2013). Comparing gestures and diagrams. In *Proceedings of TiGeR 2013: The combined meeting of the 10<sup>th</sup> international Gesture Workshop (GW) and the 3rd Gesture and Speech in Interaction (GESPIN) conference*.
- Tversky, B., Agrawala, M., Heiser, J., Lee, P., Hanrahan, P., Doantam, P., Stolte, C. & Daniel, M. P. (2020). Cognitive design principles for automated generation of visualizations. In G. L. Allen (Ed.), *Applied Spatial Cognition* (pp. 53-74). New York: Psychology Press.
- Wahlster, W. (1991). User and discourse models for multimodal communication. In J. W. Sullivan & S. W. Tyler (Eds.), *Intelligent User Interfaces* (pp. 45-67). New York: ACM Press.
- Zahner, D. C. & Corter, J. E. (2010). The process of probability problem solving: Use of external visual representations. *Mathematical Thinking and Learning*, 12(2), 177-204.

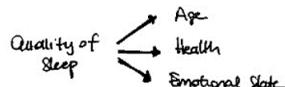
## APPENDIX A



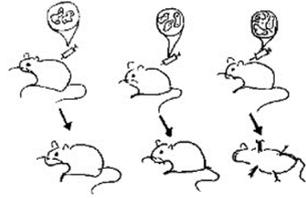
**Figure A1.** Representational arrow samples. The figures have been redrawn by the authors.



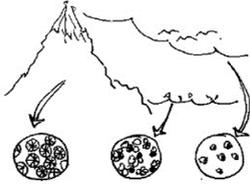
(a) Textual Link



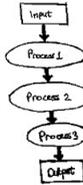
(b) Abstract Causal Link



(c) Transform



(d) Zoom



(e) Process

**Figure A2.** Relational arrow samples. The figures have been redrawn by the authors.



**Figure A3.** A deictic arrow sample. The figures have been redrawn by the authors.

## APPENDIX B



**Figure B1.** The modes of instruction of the experiment (from left to right: board, paper, tablet).



**Figure B2.** A representational gesture illustrating the “pull and push” action (left), a deictic gesture pointing to the particle on the table (middle), a beat gesture which is a speech-related rhythmic movement (right). The photos are used upon the permission of the participant.

## APPENDIX C

**Table C1.** The proportion of correctly solved problems in each condition.

		Pretest (%)	Posttest (%)
Gesture-poor	Tablet	0.49 (0.50)	0.52 (0.50)
	Board	0.80 (0.40)	0.75 (0.43)
Gesture-rich	Tablet	0.50 (0.50)	0.81 (0.39)
	Board	0.50 (0.50)	0.62 (0.49)

## NOTE

<sup>1</sup> We use the terms ‘mode’ and ‘media’ to mean the medium communication, i.e., the method of instructional delivery. This use is in contrast to the use of the term ‘multimedia’ by Mayer (2005) and colleagues. In our terminology, Mayer’s use of the term ‘multimedia learning’ corresponds to ‘multimodal learning’.