The aim of this study is to present experimental evidence in support of the idea that an “embodied-centered” teaching approach based on multimodal sensory grounding should be more effective than a classic verbal-centered teaching approach. The study is based on the comparison between learning and retrieval of sentences presented in a verbal only or verbal plus related visual images modality. Concrete and abstract meanings were used. Participants had to learn and then recognize these sentences (name and verb recognition). The preliminary data showed a general facilitation of the verbal plus visual image of the learning material. Moreover, a difficulty with abstract sentences emerged. The pattern of results supports the effectiveness of an embodied centered teaching based on educational stimulations that have a strong experiential sensory-motor basis. The implications of these results are discussed.

KEYWORDS: teaching, learning, embodied cognition, multimodal approach
1. The theoretical framework

Nowadays the embodied cognition approach has gained much consensus in the field of cognitive science. In contrast with the neo-Cartesian conception where mind, brain and body are considered as separate units, the embodied cognition approach proposes a holistic conception according to which the mind is a complex and dynamic system that emerges from the brain and the body/environment interaction (Wilson, 2002).

There are several theories of embodied cognition in the literature. Some authors emphasize the importance of body characteristics and the role of action for cognition, while others stress the importance of grounding cognition in sensorimotor modalities and situated cognition. In all theories there are several points in common such as:

1. the central notion of the body
2. the interaction between perception-cognition-action
3. critique of symbolic conception of the mind (cognitivism)

Different approaches can be summarised in two basic orientations:

- **embodied cognition** that emphasises the role of the characteristics of the body and the action for cognition (Gibbs, 2006; Rizzolatti & Sinigallia, 2008) according to the idea that cognition is constrained by the physical characteristics of our body. In this approach, the key notion is action;
- **grounded cognition** that, instead, emphasizes that cognition is linked to our sensory-motor system in its interaction with the environment (for reviews: Barsalou, 2010; Iachini, 2011). The key notion is that cognition is based on perceptual and motor processes.

The aim of this study is to present experimental evidence in support of the idea that an “embodied-centered” teaching approach based on multimodal sensory grounding should be more effective than a classic verbal-centered teaching approach. Here preliminary data are presented that are based on the comparison between learning and retrieval of words presented in a verbal only or verbal plus related visual images modality. The outcomes of this study, along with previous evidence (Mayer, 2005), should serve as a starting point for future research aimed at exploring the educational potentialities of the multimodal presentation of the teaching materials.

2. The research design

The research design consists in giving educational stimulations that have a strong experiential sensory-motor basis. Participants were required to memorize sentences describing different kinds of actions or abstract meanings. The sentences could be presented in two different modalities: verbal only and verbal plus image. Many studies have shown that language comprehension and production involve the use of sensory-motor resources (Fischer & Zwaan, 2008). Furthermore, it has
been shown that words indicating concrete (e.g. chair) and abstract (e.g. truth) concepts can contain motor information (Barsalou, 2003) and that words are typically understood and represented against background situations (cfr. Murphy & Medin, 1985). So, an “embodied-centered” teaching approach should take advantage of this operating system of the mind by devising a learning environment as close as possible to the way the mind works. On the basis of this assumption, we can hypothesize that learning materials presented in a multimodal way (i.e. written verbal information plus an image depicting the content of the sentence) should facilitate successive memory performance as compared to materials presented in an abstract manner (written verbal information only).

3. The methods and instruments

3.1 Participants

Thirty students (12 males and 18 females, mean age = 25.4, SD = 5.649) from the Second University of Naples and from the University of Naples Parthenope participated in the experiment. All participants were right-handed and had a normal or corrected to normal vision. They were assigned to two experimental conditions: “verbal” and “verbal+image”.

3.2 Setting and materials

For the experiment the instruments provided by the Laboratory of Cognitive Science and immersive virtual reality (coordinated by prof. T.Iachini) of the Department of Psychology of the Second University of Naples were used. A PC with high resolution graphics was used to present the stimuli and measure the performance.

The verbal materials consisted of 20 sentences, divided in 5 categories: sentences describing relations between objects (e.g. the cup is on the table, OBJECT); sentences describing people assuming food (e.g. the child licks the ice cream, FOOD); sentences describing people making sport (e.g. the woman throws the ball, SPORT); sentences describing animals’ actions (e.g. the dog gnaws the bone, ANIMAL); sentences describing abstract concepts (e.g. the greediness damages the world, ABSTRACT). These sentences were chosen with the following constraints: they had to contain only a verb and two substantives (see above); only relations between one entity and a second entity were described; the descriptions were clear and could be depicted by images without ambiguity.

The pictorial materials consisted of 20 scenes that described the 20 sentences, divided in 5 categories, previously described. These images were in full colour, high resolution and realistic. They depicted a scene in such a way that all the elements presented in the sentence were reproduced. As regards abstract concepts, we used images that prototypically represented the meaning of the sentence. For example, there was a wonderful landscape on the background and a smiling happy girl to illustrate the sentence: the beauty provokes joy. And more: there was a middle-age man seizing banknotes in his hand and squeezing the world.
The testing materials were the same for all conditions. For each sentence or sentence+image, there were five substantives among whom participants had to recognize the second term of each sentence. Finally, a list of 60 verbs was prepared that contained: the 20 original verbs presented in the sentences, 20 semantically similar verbs, 20 completely different verbs.

4. The description of the experiment

The assumption is that learning is facilitated if the mode of presentation of information is as closer as possible to sensory-motor experience. In order to test this hypothesis, we used an experimental paradigm similar to that used by Mayer and colleagues (e.g. Mayer, 2005). It is based on the comparison of two groups of participants who are submitted to a recognition task of words presented according to two conditions:

1. presentation of the stimulus through verbal input, i.e. only written sentences (verbal condition);
2. presentation of the stimulus through visual, i.e. full colour images, and verbal-perceptual input, i.e. written sentences (verbal + image condition).

There was first a training session where the instructions and an example of the entire procedure (by using different sentences and images) were presented. Once the participants reported they fully understood the procedure, the experimental session started.

4.1 Learning phase

Verbal condition. In each trial, 4 sentences were presented on the four quadrants of a PC screen. Each sentence appeared in body 24, character Times New Roman, and belonged to one of the five categories previously described. The presentation lasted 10 sec. The order of presentation of sentences and their location on the quadrants was counterbalanced among participants in order to avoid spurious effects due to primacy or recency.

Verbal + image condition. In each trial, 4 coloured images that depicted the relative sentences were presented on the four quadrants of a PC screen. Below each image, the related sentence appeared (body 24, character Times New Roman). The presentation lasted 10 sec. The order of presentation of sentences-images and their location on the quadrants was counterbalanced among participants.

4.2 Testing phase

After the stimuli disappeared, there was a blank (1 sec) and the testing phase started. This phase was the same for all conditions. Participants were instructed to recognize the second term of each sentence among 5 words. The first term of each sentence was presented (1 sec) and as it disappeared, the 5 words appeared and the participants had to choose one of them. These words comprised: two
semantically similar words, the original target, two different words. Each word had a number on the left side. Participants had to press the number corresponding to the chosen word without time limits. The position of the target was balanced among distractors. Accuracy (number of items correctly reproduced) and response time (msec) measured the performance. At the ending of the five trials, there was a further memory task that tested the recognition of the verbs of each sentence. Participants were asked to recognize the 20 verbs previously studied by pressing two buttons corresponding to yes/no. A list of 60 verbs was presented, including 20 semantically similar, 20 completely different, 20 original targets. Items were presented one at a time, participants had to decide if they recognized or not the item by pressing the yes/no buttons without time limits. Accuracy and response time (msec) measured the performance. To measure accuracy, the following scoring method was adopted: 1 was assigned to each original verb correctly recognized, 0 if it was not recognized; 1 was assigned to each semantically similar verb correctly discarded, 0 if it was erroneously recognized; 1 was assigned to each different verb correctly discarded, 0 if it was erroneously recognized. A mean for original, similar, different verbs was computed by averaging the 20 items for each type of verb.

5. Results

A mixed ANOVA with one between factor (verbal vs verbal+image) and one within factor (5 categories of sentences: object, food, sport, animal, abstract) was used to analyze the data about recognition of names. Furthermore, in order to analyze the data about recognition of verbs, two different ANOVAs were carried out. In the first case, data about the 5 categories of sentences (object, food, sport, animal, abstract) were collapsed and a mixed 2 X 3 ANOVA was carried out with one between factor (verbal vs verbal+image) and one within factor (3 types of verbs: original, similar, different). Instead, in the second analysis data about the two types of verbs (similar vs. original) were collapsed and a mixed 2 X 5 ANOVA was carried out with the presentation modality (verbal vs verbal+image; between) and on the 5 categories of sentences (within) as factors. The Bonferroni test was used to check post-hoc effects. The magnitude of effect sizes was expressed by $\eta^2_p$. For sake of clarity, the analyses are reported in two paragraphs: name recognition, i.e. recognition of the second terms of each sentence; verb recognition, i.e. recognition of the verb of each sentence.

5.1 Name recognition

As regards accuracy, the ANOVA showed a main effect of the learning condition ($F(1, 28) = 8.67, p < .01, \eta^2_p = .24$). As hypothesized, the verbal + image multimodal condition ($M = .870, SD = .181$) was more accurate than the verbal condition ($M = .753, SD = .223$). The ANOVA also revealed a main effect of the categories of sentences ($F(4, 112) = 7, p < .0001, \eta^2_p = .20$). The post hoc analysis revealed that the effect was due to abstract sentences being less accurate than sentences involving animals ($p < .05$), and sentences describing sport actions being less accurate than sentences involving objects ($p < .05$) and animals ($p < .005$).
Finally, an interaction between the two factors appeared (F(4, 112) = 1.619, p < .05, $\eta^2_p = .05$). A facilitation for the verbal + image condition was always present except for “animal” sentences, that were overall very accurate (see Figure 1). The post hoc analysis revealed that the effect was due to the sentences describing actions with food being more accurate in the verbal + image rather than verbal only condition (p < .05).

![Figure 1](image.png)

**Figure 1.** The graph shows the level of accuracy for the “name” recognition as a function of the 5 categories of sentences and the learning condition.

As regards Response Time, the ANOVA revealed a main effect of sentence categories (F(4, 112) = 6.850, p < .001, $\eta^2_p = .20$). The post hoc analysis revealed that the effect was due sentences involving animals being faster than those involving objects, sports and abstract concepts (in all cases p < .05). There was neither main effect of learning conditions (F < 1) nor significant interaction (F(4, 112) = 1.533, p > .05, $\eta^2_p = .05$).

### 5.2 Verb recognition

As regards accuracy, the first ANOVA revealed a main effect of the learning condition (F(1, 28) = 9.958, p < .01, $\eta^2_p = .26$). As hypothesized, also in this case the verbal + image multimodal condition (M = .811, SD = .082) was more accurate than the verbal condition (M = .738, SD = .150). Moreover, there was a main effect of the types of verbs: original, similar, different (F(2, 56) = 41.104, p < .0001, $\eta^2_p = .60$). Figure 2 illustrates these effects. Participants rightly recognized the original verbs in the 64% of the cases, erroneously recognized as original the semantically similar verbs in the 26% of the cases and the different verbs in the 6%
of the cases. The related means were: original = .638, SD = .132, similar = .745, SD = .147, different = .940, SD = .109. The post hoc analysis revealed that the effect was due to significant differences between all pairwise comparisons (in all cases ps < .05). Finally, there was no interaction between the two factors: F (2, 56) = 1.595, p > .05, η²_p = .06. Notwithstanding, in order to deeply understand the effect of the presentation modality on the verb recognition performance we carried out three one-way ANOVAs with the kinds of verbs (original vs. similar vs. different) as a within factor, and presentation modality (verbal vs verbal+image) as a between factor. Results showed that the presentation modality did not affect the mean accuracy in recognizing the “original” verbs (F <1), whereas it influenced the ability to recognize as “not presented” the “similar” (F(1, 28)= 7, p <.05, η²_p = .20) and the “different” (F(1, 28)= 4.6, p < .05, η²_p = .14) verbs. As it is shown in Figure 2, the verbal+image condition mainly improved participants’ performance in judging “similar” and “different” verbs as “not having been previously presented”.

As regards the 2 X 5 ANOVA, it revealed a main effect of the learning condition (F(1, 28) = 8.73, p < .01, η²_p = .24). Also in this case the verbal + image multimodal condition (M = .76, SD = .082) was more accurate than the verbal condition (M = .68, SD = .150). Moreover, there was a main effect of the categories of sentences (F(4, 112) = 22.4, p < .001, η²_p = .44). The post hoc analysis revealed that the effect was due to significant differences between all pairwise comparisons (in all cases ps < .05) except for abstract vs animal sentences. Overall, the most accurate performance was on verbs contained in sentences describing actions with food, whereas the less accurate performance refers to the verbs contained in sentences describing actions with non-food items.

**Figure 2.** The graph shows the level of accuracy for the “verb” recognition of the 3 types of verbs as a function of the learning condition.
in sentences describing abstract concepts. Finally, there was no interaction between the two factors (F<1). Figure 3 illustrates these effects.

![Figure 3](image_url)

**Figure 3.** The graph shows the level of accuracy for the recognition of the verbs contained in the five categories of sentences as a function of the learning condition.

As regards Response Time, the ANOVA showed a main effect of the types of verbs: original, similar, different (F(2, 56) = 7.313, p < .01, η²_p = .21). The related means (in msec) were: original = 1663.440, SD = 428.690, similar = 1799.797, SD = 450.667, different = 1611.580, SD = 418.300. The post hoc analysis revealed that the effect was due to the performance about similar verbs that was slower than both original and different verbs (in both cases ps < .05). However, there were neither main effect of the learning condition (F < 1) nor interaction between the two factors: F (2, 56) = 1.649, p > .05, η²_p = .06.

Finally, the last 2 X 5 ANOVA showed a main effect of the categories of sentences (F(4, 112) = 3.05, p < .05, η²_p = .09). The related means were: food = 1528, SD = 332.42, object = 1593.86, SD = 435.99, sport = 1743.51, SD = 664.18, animal = 1800.93, SD = 555.17, abstract = 1815.14, SD = 662.96. However, the post hoc analysis did not reveal significant differences among the categories. Finally, there were neither main effect of the learning condition (F < 1) nor interaction between the two factors (F<1).
Conclusions

The results showed that the recognition performance of the second name was more accurate when the learning materials were presented in a multimodal way, i.e. by sentences and images depicting them, rather than in a verbal only way. Therefore, they verify the hypothesis that the multimodal presentation of learning materials facilitates successive memory performance as compared to exclusively verbal materials. Notably, when participants were presented with more learning materials, that is both images and sentences, this did not charge the processing time needed to retrieve them.

Consistently, the results about the recognition of the verbs confirmed that the performance was more accurate when the materials included both sentences and images rather than sentences only. In particular, results indicated that the presence of images along with written words indirectly improved recognition performance by avoiding participants to judge as previously learned synonymous or completely different verbs with respect to the actually presented verbs. Moreover, there was no significant response time difference between the learning conditions. This would suggest that people encoded images and sentences as a holistic representation where the meaning of each source of information was mixed together.

Overall, these results are in line with previous evidence about multimedia learning (Mayer, 2005). For example, Mayer and Anderson (1992) found out that participants were more accurate in a retention test when both verbal and images of complex operations were seen with respect to the verbal only presentation. Specifically, they suggest that pictures’ and words’ presentations are most effective when they occur contiguously in time or space. According to Mayer’s cognitive theory of multimedia learning (2001), printed words and pictures are processed in two different channels producing a pictorial and a verbal model respectively. These two models are then integrated with each other and person’s prior knowledge, producing a more robust representation than the verbal only modality.

Although the multimodal presentation was overall facilitating, there also emerged an effect of the different categories of sentences and in particular a difficulty with abstract sentences. This difficulty confirms the well known concrete words superiority effect (Paivio, 1971) that supports the dual coding model proposed by Paivio (1978, 1991). According to Paivio’s theory, verbal stimuli are processed by a verbal subsystem based on abstract and discrete symbols (logogens) that represent linguistic information. Non verbal stimuli, instead, are processed by an imagery subsystem that is based on perceptual-like and continuous symbols that represent perceptual information (imagens). Since concrete words are represented both by imagens and logogens, this would ensure a better memory performance as compared to abstract words that are represented only by logogens. In sum, the Paivio’s model and the experimental evidence he gathered would suggest that learning and retrieval is facilitated when the materials are concrete and close to everyday sensory experience. However, adding an image to represent an abstract concept does not ensure that it has a positive effect on the memorization of that concept. According to Brunyé and colleagues (2007), an image should make the learner effectively able to establish connections be-
tween the written word and the content of the image. It can be easily understood that this operation could be simple if the word is “CHAIR” and the image represents a chair, whereas it is more difficult if the word to be memorized is “TRUTH.” From a more grounded perspective, Schwanenflugel and colleagues (for a review see Schwanenflugel, 1991) suggest that abstract words can be accessed, understood and remembered as well as concrete words only when relevant situations are presented along with them, that is situations in which the learner has experienced the abstract concept. This suggests that the choice of visual images is crucial in helping the learning of abstract sentences. In line with Mayers’ theory (2005) it is possible that the images used for depicting abstract concepts are more effective if participants are enabled to build on useful connections between written words and image contents. According to a more embodied approach, this beneficial effect should be linked to images being representative of participants’ experiences (i.e. situation) with the presented abstract contents. However, Mayer’s theory cannot account for the lack of efficacy of multimedia presentation for sports actions where, instead, the connections between written words and image contents could be built without ambiguity (e.g. in our study, if the sentences was “the girl lifts weights”, the image depicted a girl lifting weights). Possibly, the difficulty we found with sentences describing sport actions may be explained by the fact that participants had a low experience with the implied sports.

From a theoretical point of view, it is important to say that the results of this exploratory study are not in contrast with Paivio’s and Mayer’s models, but they add to this traditional line of research by showing that the involvement of visual stimuli depicting concrete situations facilitates the performance not only when the written words can be connected without ambiguity to the image contents but especially when the images can activate learners’ sensory-motor experience. Indeed, this facilitation effect clearly appears in our study when participants had to retrieve from memory verbs and substantives contained in sentences describing actions with food. In other words, results from the current study would support the effectiveness of an embodied centered teaching that is based on educational stimulations that have a strong experiential sensory-motor basis. Clearly, though, this study only constitutes a preliminary attempt at establishing the role and the weight of sensory-motor experiences in the learning of concrete and abstract concepts. Many issues remain unresolved, and further research remains to be done.

From a teaching point of view, one possible application of the study would be to reflect on the quality of the stimulus, as demonstrated in the literature. According to Paivio (Paivio, 1991) concreteness, imagery, and verbal associative processes play an important role in various educational domains; the representation and understanding of knowledge, the memorization of subject contents together with effective teaching improve the learning of motor skills. Our research shows that the selection of material has important implications in teaching, especially if we look at the multimodal condition of the experiment when participants improve their performance significantly both in terms of speed and accuracy. Therefore, it is useful to use stimuli that have a strong visual impact and are particularly
engaging from the point of view of motor imagination. For example, a lesson that in addition to verbal communication also uses a set of stimulating and engaging images has a better chance to impact successfully not only on the cognitive aspects of the learning process but also on the emotional aspects. Another aspect to be considered for testing the usefulness of the experimental model to teaching are the metacognitive skills of self-evaluation and self-selection to evaluate the quality of the stimuli for facilitating the memorization and storage of the study material. In other words, it may be useful to allow learners to monitor their own choices through a “learning governance process” (Iavarone & Santoianni, 2003) which allows them to select cognitive procedures that are increasingly more relevant and effective for learning purposes, especially when it is possible to give a meaning to individual decisions and to draw interpretative and operational conclusions about them. Indeed, recent contributions on educational resources have pointed out that the effectiveness of the training lies in the ability of individuals to control the learning processes and to guide the process of construction of knowledge as an active and consciously directed process. This approach should inform the setting of learning environments that should be organized in a visually appealing way, as well as teaching materials that should be well structured; even a simple text should be organized, for example, through label reading in such a way that the learner never misses the focal point of the process, both in terms of content and procedures. In conclusion, we believe that although additional research still needs to be done, the results of our study may be applied to increase our understanding of educational phenomena and to strengthen related pedagogical practices in order to improve teacher training programs.

References


