Presentation of Keywords by Means of Interactive Drawings

Alfredo Campos[1], Ángeles Amor[2], and María Ángeles González[2]

[1]University of Santiago de Compostela   [2]University of A Coruña

One of the main outstanding problems in keyword mnemotechnics is whether this technique is more effective when the subjects generate their own keywords, or when the keywords are supplied by the experimenter. Both methods have advantages and disadvantages. An alternative method has recently been suggested, in which the keywords are generated by the subjects' peers. In the present study we aimed to investigate whether immediate or delayed recall are affected by keyword generation method (experimenter or peer generation). We also aimed to determine whether the method of keyword generation affects imaging capacity as evaluated by questionnaires or spatial tests. A total of 377 secondary-school students were selected and divided into four groups. All subjects were presented with 30 Latin words. Additionally, the subjects in Group 1 were presented with bizarre images, while the subjects of Group 2 were presented with normal images, in both cases representing the keywords selected by peers as interacting with the Latin words. The subjects in Groups 3 and 4 were likewise presented with normal or bizarre images, respectively, but representing the keywords selected by the experimenters. The subjects' imaging capacity was evaluated by means of the Vividness of Visual Imagery Questionnaire (VVIQ) and the Spatial Test of Primary Mental Abilities (ST-PMA). The results were analysed by multivariate analysis of variance (MANOVA) with three factors (ST-PMA imaging capacity, 2 levels; VVIQ imaging capacity, 2 levels; and mnemotechnic method, 4 levels) and dependent variables immediate recall and delayed recall. All three factors influenced recall. Subsequent univariate analyses of variance indicated that subjects with high ST-PMA score and subjects with high VVIQ score showed better immediate and delayed recall than subjects with low ST-PMA score and subjects with low VVIQ score. Mnemotechnic method (i.e. whether keywords are generated by the experimenter or by peers) significantly affected immediate recall but not delayed recall.
Key words: keyword method, mnemonic, drawings, imagery

Uno de los problemas sin resolver en la mnemotecnia de la palabra clave es si ésta resulta más eficaz cuando son los sujetos los que generan sus propias palabras clave o cuando son facilitadas por el experimentador. Los dos métodos tienen ventajas e inconvenientes. Recientemente se ha propuesto una alternativa que consiste en que sean compañeros de los sujetos los que elaboren las palabras clave, alternativa que parece eficaz. En esta investigación se deseaba saber si existía diferencia en el recuerdo inmediato y en el demorado cuando la palabra clave, reforzada con dibujos normales y raros, es elaborada por los experimentadores o cuando es elaborada por compañeros de los sujetos que participan en la investigación. También se deseaba saber si afectaba al recuerdo la capacidad de los sujetos de formar imágenes mentales cuando ésta se evalúa a través de cuestionarios y de pruebas espaciales. Se seleccionó una muestra de 377 estudiantes de ESO que se distribuyeron en cuatro grupos. A todos los sujetos se les presentaron 30 palabra latinas, pero además, al primer grupo se le presentaron dibujos raros, y al segundo se le presentaron dibujos normales, confeccionados con las palabras clave generadas por los compañeros en interacción con el significado de las palabras latinas. A los grupos 3 y 4 también se le presentaron dibujos normales y raros, respectivamente, pero confeccionados con las palabras clave generadas por los experimentadores. También se evaluó la capacidad de los sujetos de formar imágenes mentales a través de un cuestionario, el Vividness of Visual Imagery Questionnaire (VVIQ) y de un test espacial, la Escala Espacial del Test de Aptitudes Mentales Primarias (ST-PMA). Se efectuó un Análisis de Varianza (MANOVA) de 2 (ST-PMA capacidad de imagen) x 2 (VVIQ capacidad de imagen) x 4 (método mnemotécnico). Las variables dependientes fueron el recuerdo inmediato y el demorado. Las tres variables influyeron en el recuerdo. Los posteriores ANOVAs indicaron que los sujetos altos en PMA y los sujetos en VVIQ tuvieron un mejor recurdo inmediato y retardado que los sujetos bajos en PMA y en VVIQ. El método mnemotécnico influyó en el recuerdo inmediato, pero no en el retardado.
Palabras clave: método de la palabra clave, mnemónico, dibujos, imagen

The keyword mnemotechnics method (Atkinson, 1975) comprises two basic steps: the first consists of identification of the L1 keyword, which should be as concrete as possible, as familiar as possible for the participant, and as close as possible in sound to the L2 word to be learnt. The second step involves creating a visual image relating the keyword to the target word. Subsequently, by evocation of the keyword and interaction of images, the participant remembers the word's L1 translation.

Since the initial description of this technique, there have been many studies of its effectiveness (see Pressley, Levin, & Delaney, 1982; and Pressley, Levin, & MacDaniel, 1987, for reviews). However, some questions remain unresolved. One such issue is whether the keywords should be supplied by the experimenter or generated by the participants themselves.

Although many authors have expressed opinions on this issue, there have been few experimental studies aimed at resolving it. Some studies have found no significant differences between the two methods (Hall, 1988; McGivern & Levin, 1983; Pressley et al., 1980), whereas others have found that experimenter-supplied keywords are more effective than subject-generated keywords (Hall, Wilson, & Patterson, 1981; King-Sears, Mercer, & Sindelar, 1992). As far as we know, no studies have obtained better results with subject-generated keywords, despite the putative advantages identified on theoretical grounds by a number of authors. For example, Wang, Thomas, and Ouellete (1992) and Thomas and Wang (1996) have hypothesized that subject-generation of keywords may be more effective because it avoids conflict between the experimenter's and the participant's modes of coding.

The view that keywords should be generated by the experimenter is largely based on the supposition that participants will find it difficult to identify keywords as effective as those identified by the experimenter. Furthermore, it takes time to construct a keyword list, and many school-age children are clearly not capable of generating effective keywords (Levin, Pressley, McCormick, Miller, & Shriberg, 1979).

In a recent study (Campos, González, & Amor, 2001), a new approach aimed at combining the advantages of experimenter-supply and subject-generation of keywords was tested. Specifically, keywords were generated previously by subjects of the same age and sociodemographic characteristics as the participants. To investigate the effectiveness of this strategy, three groups of participants were asked to learn a list of Latin words, (a) with subject-generation of keywords, (b) with experimenter-generated supplied keywords, or (c) with supplied keywords generated previously by subjects of the same age and sociodemographic characteristics. The results indicated that participants who used peer-generated keywords showed significantly better short- and long-term recall than did participants in the other two groups.

Mental images play an important role in many cognitive processes (memory, learning, reasoning, problem solving, etc.; Campos et al., 2001; Denis, 1979; Higbee, 1993; Paivio, 1971; Richardson, 1994). There have been numerous studies of relationships between imaging capacity (as measured by spatial tests) and memory, particularly in situations in which the item to be learnt is a concrete word (see Ernest, 1977, and Richardson, 1994, for reviews). On the basis of a number of reviews and experimental studies, Ernest (1991) suggested that imaging capacity —especially if measured by spatial tests— is an effective predictor of word recall. In a recent study, imaging capacity measured by spatial tests has been shown to affect both immediate and delayed recall of associated word pairs (Campos et al., 2001). Campos and Pérez (1997) found that imaging capacity as evaluated with a questionnaire had a significant influence on immediate recall of words from both a short list (16 word pairs) and a long list (32 pairs), but did not affect 1-week recall.

The use of bizarre images to aid recall is a truly ancient tradition. In the present context, many authors have suggested that mental images are more effective aide-mémoires when they are both bizarre and interactive. Experimental studies have shown that interaction aids memory (Bergfeld, Choate, & Kroll, 1982; Campos & Pérez, 1996, 1997; Carney & Levin, 1998). The use of drawings to present interactive images has previously been reported to be an effective learning aid (Atkinson et al., 1999; Tomas & Wang, 1996). However, the effect of bizarreness on memory has not been fully clarified experimentally, and depends on many variables, including type of task (recall of associated pairs, or free recall), other aspects of task design (pure lists or mixed lists), and the interval between learning and recall (Marchal & Nicolas, 2000).

Studies performed to investigate the influence of bizarreness on short-term recall of associated word pairs have suggested that bizarre images are no more effective than normal images (Andreoff & Yarmey, 1976; Bergfeld et al. 1982; Kroll, Schepeler, & Angin, 1986; O'Brien & Wolford, 1982; Riefer & Rouder, 1992; Webber & Marshall, 1978), or indeed that bizarre images are less effective than normal images (Campos & Pérez, 1997; Kroll, Jaeger, & Dornfest, 1992; Pra Baldi, De Beni, Cornoldi, & Cavedon, 1985; Riefer & LaMay, 1998; Wollen & Cox, 1981a, 1981b;). Bizarre images are rarely more effective than normal images, unless mixed lists are used and retention is assessed with a free-recall test (Marchal & Nicolas, 2000; McDaniel, DeLosh, & Merritt, 2000; Tess, Hutchinson, Treloar, & Jenkins, 1999).

The initial poor results obtained with the use of bizarre images as aids to immediate recall led researchers to examine the effectiveness of these images with longer retention intervals. On the grounds that bizarre images are more distinctive than normal images, many researchers have argued that they will be more effective as aids to long-term recall, as they are less likely to be subject to extra-experimental interference.

Experimental studies aimed at investigating the effect of bizarreness on long-term recall of associated word pairs presented in pure lists (as in the present study) have similarly produced contradictory results. Marshall, Nau, and Chandler (1980) found that bizarre images were significantly more effective than normal images. By contrast, Bergfeld et al. (1982) found that normal images were more effective. The apparent discrepancy between these two studies may reflect methodological differences: Marshall et al. (1980) examined 1-week recall of associated word pairs, whereas Bergfeld et al. (1982) examined 1-day recall of pairs of objects presented as line drawings. Despite the lack of any clear influence of image type, Marshall et al. (1980) suggest that a crucial factor may be the duration of the retention interval.

However, comparisons of recall over different retention intervals have likewise not revealed any clear pattern. Andreoff and Yarmey (1976) found that bizarre images significantly improved immediate recall, but did not find significant differences between bizarre and normal images in 1-day recall. O'Brien and Wolford (1982) found that bizarre and normal images were similarly effective for recall at 1 - 3 days, but that bizarre images were significantly more effective for recall at 5 - 7 days. Iaccino, Dvorak, and Coler (1989), in a study using both mixed and pure lists, also observed that bizarre images were more effective for recall at 7 days.

Other researchers have not found any significant differences in the effectiveness of bizarre and normal images, regardless of retention time. Kroll et al. (1986) and Campos and Pérez (1997) did not find significant differences between the two methods in 1-week recall. Similarly, Poon and Walsh-Sweeney (1981) did not detect significant differences, whether in young or older participants, in recall tested after 1 day, 1 week, and 1 month.

Mercer (1996), in a historical review of the effect of bizarre imagery on memory, concluded that between 1987 and the present researchers have done little more than replicate previous studies. The basic conclusions that can be drawn are that bizarre images are more effective in free recall, immediate recall (5 minutes or less) and mixed-list tests. The effects of bizarre images on delayed recall remain unclear.

In this study we aimed to evaluate whether subjects' imaging capacity, as assessed by a subjective test and a performance test, influences immediate recall and delayed recall of the L1 translation of Latin words learnt by image-reinforced keyword mnemotechnics. We also aimed to assess whether immediate and delayed recall were affected by the type of image used: normal images generated by the experimenter, bizarre images generated by the experimenter, normal images generated by peers, or bizarre images generated by peers. Finally, we aimed to assess whether there is any interaction, as regards effects on immediate and delayed recall, between imaging capacity and keyword generation method.

Method

Participants

The sample comprised 377 subjects (193 boys, 184 girls) aged between 12 and 16 years ($M = 13.3$, $SD = 1.14$). All participants were secondary school students ([1st] - [3rd] year of Spanish Secondary Compulsory Education). Four state schools of similar characteristics took part in the study. Each subject was randomly assigned to one of the experimental groups.

Instruments

To measure imaging capacity, we used the Spanish version of the Spatial Test of the Primary Mental Abilities Scale (ST-PMA; Thurstone & Thurstone, 1962) and the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). The ST-PMA measures the individual's capacity to imagine and conceive of objects in two or three dimensions, and is widely used as a measure of imaging capacity (Campos & González, 1994; Denis, 1979; Ernest, 1977; González, Campos, & Pérez, 1997). It comprises 20 items, each of which presents a flat geometric model and six similar figures: The participant must decide which of the figures corresponds to the model (rotations on the plane allowed, reflections not allowed). The mean test-retest reliability of the ST-PMA has been found to be .73 for Spanish subjects; score on this test shows a correlation of .42 ($p < .001$) with the Vividness of Visual Imagery Test (Campos, 1998).
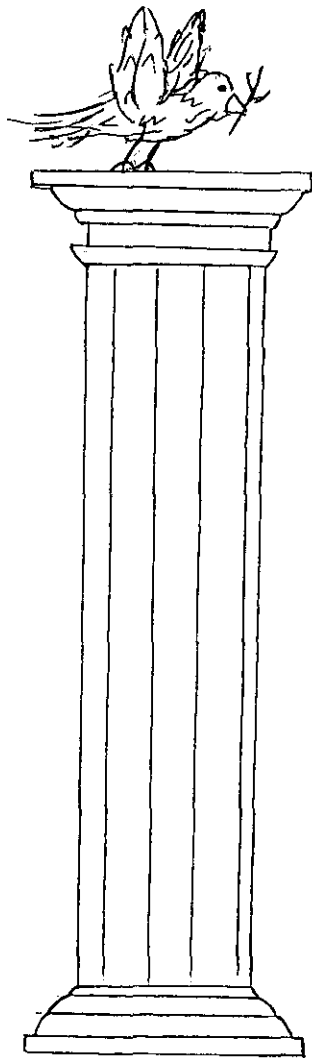
The VVIQ is made up of 16 items referring to different situations that the participant is asked to visualize, rating image vividness on a 5-point scale. The participant first completes the questionnaire with eyes open, and then again with eyes closed. A high score on the questionnaire indicates low imaging capacity. The VVIQ is the most widely used instrument for the measurement of imaging capacity (Campos, 1995; Campos, Chiva, & Moreau, 2000; McKelvie, 1995; Richardson, 1994). Campos, González, and Amor (2002) performed a factor analysis with the results obtained in a sample of Spanish subjects, and identified a single factor, namely vividness of visual imagery. The internal consistency of the questionnaire was high (Cronbach's a = .88). Campos, González, and Amor (2000) found a significant correlation ($r = .68$, $p < .001$) between score on this questionnaire and the Gordon Test of Visual Imagery Control (Richardson, 1969). We used two measures of imaging capacity because the relationship between the two, although statistically significant, is weak ($r = -.15$, $p < .01$).

Procedure

Construction of the Word Lists. The list of 30 Latin words and their translations to Spanish was drawn up as follows. First, we randomly selected 50 Spanish words from among those listed by Valle (1998) as being concrete and with an

imagery vividness score of more than 5 on Valle's scale. We then translated these words to Latin, and rejected cognates (i.e., words with similar pronunciation and/or spelling in the two languages). Latin was selected because, as pointed out by Atkinson (1975), it is well suited for studies of this type, and because none of our participants had any prior knowledge of it.

Having drawn up the list of Latin words and their Spanish equivalents, we then generated keywords for each Latin word (for subsequent use in the experimenter-supplied-keyword group). Care was taken to ensure that keywords were not phonetically similar to other words or keywords in the list. Of the Latin words, we randomly selected 30 (with Spanish equivalents and keywords).

The 30 Latin words were then presented to a group of 60 subjects (not forming part of the experimental group of 377 participants, but with very similar sociodemographic characteristics). These subjects were given a brief verbal explanation of the keyword technique, and were then each asked to choose keywords for each word. For each word, we then selected that keyword which had been chosen by most subjects. These were the "peer-generated keywords" given to Groups A and B.

We then prepared four sets of 30 drawings (one set for each experimental group), as follows: (A) normal ("non-bizarre") drawings relating each peer-generated keyword to the corresponding L1 target word (for an example, see Figure 1); (B) bizarre drawings relating each peer-generated keyword to the corresponding target word (for an example, see Figure 2); (C) normal drawings relating each experimenter-generated keyword to the corresponding target word; and (D) bizarre drawings relating each experimenter-generated keyword to the corresponding target word. The validity of the keywords and drawings was ensured by requiring the three researchers, all experts in mnemonics, to be in agreement. Normal images were images corresponding to reality and daily life, whereas bizarre images were clearly unreal, for example a cat with wings.

COLUMBA (columna)

paloma

Figure 1. An example of a normal interactive drawing.
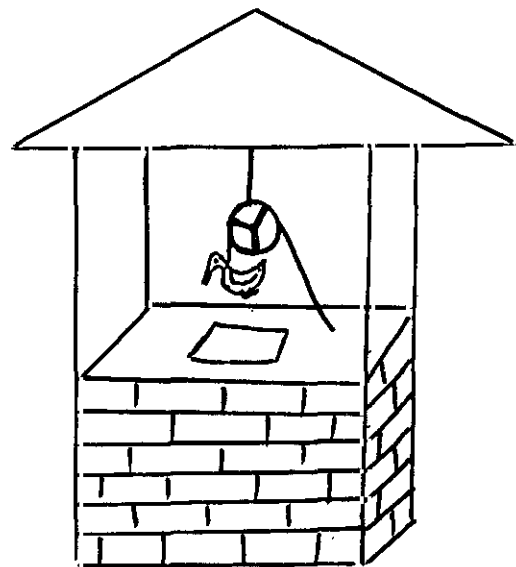
PETEUS (patos)

pozo.

Figure 2. An example of a bizarre interactive drawing.

Recall Tests. The group of participants was divided into four groups. All participants were presented with the 30 words on a half-A4 looseleaf binder (one per page): Each page bore a drawing, and at the bottom of the page the Latin (L2) word, its translation to Spanish (L1), and the keyword. All participants were told that they should learn the meaning of each Latin word by the keyword method. The participants in Group A were given keywords generated by other subjects of similar characteristics ("peer-generated keywords") and normal drawings. The participants in Group B were given peer-generated keywords and bizarre drawings. The participants in Group C were given experimenter-generated keywords and normal drawings. The participants in Group D were given experimenter-generated keywords and bizarre drawings.

The experiment was carried out in two classroom sessions, one week apart. On the first day, participants received the corresponding word list, a recall test (immediate recall), and the ST-PMA. On the second day, one week later, participants again received the recall test (1-week recall), and the VVIQ. In all cases, instructions were given both verbally and in writing. For memorization of the word list, participants were allowed 15 minutes; each of the recall tests lasted 3 minutes. Subsequently, participants were classified as strong or weak imagers, depending upon whether their ST-PMA score or VVIQ score (as appropriate) was above or below the mean for the group. The experiment was performed in the participants' normal classroom in groups of about 22 individuals. All participants were volunteers.

As each word list comprised 30 items, mean time per item was 30 s. In Mercer's (1996) review, 30-35 s is the maximum time given in studies of this type, and is sufficient for good results. For example, Imai and Richman (1991) compared common sentences with atypical and illogical sentences. The illogical sentences were no better recalled than the common sentences after a processing time of 7 s, but were better recalled after 35 s. Cornoldi, Cavedon, De Beni and Pra Baldi (1988) found that bizarre images gave better recall than normal images with presentation intervals of 30 s. Mercer (1996) suggests that there may be a need to perform further studies using longer presentation intervals, so that subjects have more time to create bizarre images.

Results

We first analyzed the data by 3-factor multivariate analysis of variance (MANOVA), with dependent variables immediate recall and 1-week recall, and factors ST-PMA imaging capacity (2 levels: high or low), VVIQ imaging capacity (2 levels: high or low) and mnemotechnic method (4 levels: peer-generated or experimenter-generated keyword, normal or bizarre drawing). Mean immediate and 1-week recalls in each group (± standard deviations) are shown in Table 1.

The factor ST-PMA imaging capacity had a statistically significant effect on recall, $F(2, 362) = 7.18$, $p < .001$, Wilks's $\Lambda = .96$. Subsequent univariate analyses revealed that PMA imaging capacity was a significant source of variance in both immediate recall, $F(1, 362) = 14.39$, $p < .001$, and 1-week recall, $F(1, 362) = 8.31$, $p < .01$. Participants with strong imaging capacity showed better immediate and 1-week recall than did participants with low imaging capacity.

The factor VVIQ imaging capacity had a statistically significant effect on recall, $F(2, 362) = 3.31$, $p < .05$, Wilks's $\Lambda = .98$. Subsequent univariate analyses revealed that VVIQ imaging capacity was a significant source of variance in both

Table 1
Means and Standard Deviations of Immediate and 1-Week Recall in Participants as a Function of PMA Score, VVIQ Score, and Experimental Group

Group	Recall					
	Inmediate		One-week		Total	
	M	SD	M	SD	M	SD
High ST-PMA	20.10	6.24	14.22	6.21	17.58	5.87
Low ST-PMA	17.30	7.04	12.08	6.40	15.37	6.44
High VVIQ	16.76	7.27	11.97	6.46	15.33	6.66
Low VVIQ	19.63	6.45	13.89	6.23	17.17	5.88
Normal drawings (experimenter)	17.36	6.63	12.48	6.27	15.06	5.81
Normal drawings (peer)	17.14	7.19	12.38	6.16	14.73	6.32
Bizarre drawings (experimenter)	19.26	7.25	13.14	6.48	16.38	6.40
Bizarre drawings (peer)	20.01	6.46	14.27	6.52	17.16	6.16
Total	18.46	6.96	13.12	6.39	15.88	6.22

Note. ST-PMA = Spatial Test of the Primary Mental Abilities Scale (Thurstone & Thurstone, 1962). VVIQ = Vividness of Visual Imagery Questionnaire (Marks, 1973). High VVIQ scores indicate low imaging capacity, and vice versa.

immediate recall, $F(1, 362) = 5.84$, $p < .01$, and 1-week recall, $F(1, 362) = 5.89$, $p < .01$. Again, participants with strong imaging capacity showed better immediate and 1-week recall than did participants with low imaging capacity.

The factor mnemotechnic method also had a statistically significant effect on recall, $F(6, 722) = 2.26$, $p < .05$, Wilks's $\Lambda = .96$. Subsequent univariate analyses revealed that mnemotechnic method was a significant source of variance in immediate recall, $F(1, 362) = 3.40$, $p < .01$, but not in 1-week recall, $F(3, 363) = 1.38$, $p > .05$. The Student-Newman-Keuls test indicated that subjects who used bizarre drawings of peer-generated keywords achieved better immediate recall than subjects who used normal images, independently of whether the keywords were experimenter- or peer-generated. Recall did not differ significantly between subjects who used bizarre images of experimenter-generated keywords and subjects who used bizarre images of peer-generated keywords. Likewise, recall did not differ significantly between subjects who used bizarre images and subjects who used normal images.

Lastly, we investigated possible between-variable interactions, and found that there was no interaction between the factors ST-PMA imaging capacity and mnemotechnic method, $F(6, 722) = .83$, $p > .05$, Wilks's $\Lambda = .44$). Subsequent univariate analyses revealed that the interaction between these variables did not influence either immediate recall, $F(1, 362) = .83$, $p > .05$, or one-week recall, $F(1, 362) = .70$, $p > .05$. There was likewise no significant interaction between VVIQ imaging capacity and mnemotechnic method, $F(6, 722) = .98$, $p > .05$, Wilks's $\Lambda = .98$), either in immediate recall, $F(1, 362) = 1.68$, $p > .05$, or one-week recall, $F(1, 362) = 1.48$, $p > .05$. There was similarly no significant interaction between ST-PMA imaging capacity and VVIQ imaging capacity, $F(2, 361) = 10$, $p > .05$, Wilks's $\Lambda = .99$, either in immediate recall, $F(1, 362) = .18$, $p > .05$, or in one-week recall, $F(1, 362) = .07$, $p > .05$. Finally, there was no significant three-way interaction between the three variables (ST-PMA imaging capacity, VVIQ imaging capacity, mnemotechnic method), $F(6, 722) = 1.57$, $p > .05$, Wilks's $\Lambda = .97$, either in immediate recall, $F(1, 362) = 1.19$, $p > .05$, or in one-week recall, $F(1, 362) = 2.49$, $p > .05$.

Discussion

A previous study, Campos et al., 2001 observed that peer-generated keywords were significantly more effective than subject- or experimenter-generated keywords. In this study, we used interactive images and did not include a subject-generated-keyword group, because it would be unreasonable to expect participants to draw their own interactive images in a short time. Neither immediate nor 1-week recall differed significantly between the peer-generated-keyword groups and the experimenter-generated-keyword groups. This result in

not in accordance with the one obtained by Campos et al. (2001), although it should be stressed that in the latter study, drawings were not used. As pointed out by Atkinson et al. (1999), Hwang et al. (1999) and Thomas and Wang (1996), drawings improve learning effectiveness, so the use of drawings in the present study may account for the lack of significant differences between the two groups.

When keywords are generated by peers of the intended learners, we can expect those keywords to be very similar to those that would be generated by the learners themselves, in theory offering many of the putative advantages of subject-generation of keywords, and at the same time avoiding many of the disadvantages of experimenter-generation (see Levin et al., 1979; Thomas & Wang, 1996; Wang et al., 1992). Despite the fact that in the present study, we did not find any significant differences between the experimenter-generated-keyword groups and the peer-generated-keyword groups, we consider that further studies are necessary to investigate differences between the two methods, with consideration not only of the factors considered in the present study, but also of the similarity of the peer- and experimenter-generated keywords.

Participants who used bizarre images of peer-generated keywords showed significantly better immediate recall than participants who used normal images, whether of peer- or experimenter-generated keywords. This is probably because bizarre images highlight distinctiveness more effectively than normal images (Howe, Courage, Vernescu, & Hunt, 2000; Tess et al., 1999), which helps individuals to recognize drawings, and, through them, the keywords and their meanings. This process is probably not so marked when bizarre drawings are generated by the experimenter, or when normal images are used. Delayed recall did not differ significantly between subjects who used normal and bizarre images. These results are in agreement with those of previous studies of delayed recall (Campos & Pérez, 1997; Kroll et al., 1986; Poon & Walsh-Sweney, 1981). Mercer (1996), in a review of pre-1995 studies of the effect of bizarreness on recall, concluded that bizarreness has no clear effect on long-term recall.

Participants' imaging capacity, whether measured by the VVIQ or the ST-PMA, had significant effects on both immediate and 1-week recall. This finding is in line with previous studies, particularly those that have evaluated imaging capacity using spatial tests (Campos & Pérez, 1997; Campos et al., 2001; Ernest, 1977, 1991).

The results of the present study thus indicate that imaging capacity influences both immediate and 1-week recall using the interactive image-keyword mnemotechnic method. However, the results as regards effects of image type (normal or bizarre) and method of keyword generation (experimenter or peers) on immediate recall are unclear, suggesting a need for further studies of this aspect. We believe peer generation of keywords may be a useful approach but further studies are required to identify those procedures and conditions that maximize method efficiency.

References

Andreoff, G.R., & Yarmey, A.D. (1976). Bizarre imagery and associative learning: A confirmation. *Perceptual and Motor Skills, 43*, 143-148.

Atkinson, R.C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*, 821-828.

Atkinson, R. K., Levin, J. R., Kiewra, K. A., Meyers, T., Kim, S., Atkinson, L. A., Renhandya, W. A., & Hwang, Y. (1999). Matrix and mnemonic text-processing adjuncts: comparing and combining their components. *Journal of Educational Psychology, 91*, 342-357.

Bergfeld, V.A., Choate, L.S., & Kroll, N.E. (1982). The effect of bizarre imagery on memory as a function of delay: Reconfirmation of the interaction effect. *Journal of Mental Imagery, 6*, 141-158.

Campos, A. (1995). Twenty-two years of the VVIQ. *Journal of Mental Imagery, 19*, 129-131.

Campos, A. (1998). A measure of visual imaging capacity: A preliminary study. *Perceptual and Motor Skills, 87*, 1012-1014.

Campos, A., Chiva, M., & Moreau, M. (2000). Alexithymia and mental imagery. *Personality and Individual Differences, 29*, 787-791.

Campos, A., & González, M.A. (1994). Viveza de las imágenes mentales y rendimiento académico en estudiantes de Bellas Artes, Ciencias y Letras. *Revista Latinoamericana de Psicología, 26*, 69-81.

Campos, A., González, M. A., & Amor, A. (2000). Rendimiento académico en tecnología: el valor de las imágenes mentales. In M.A. Santos (Ed.), *A educación en perspectiva* (pp. 379-387). Santiago de Compostela, Spain: Universidad de Santiago de Compostela.

Campos, A., González, M.A., & Amor, A. (2001). *Different strategies for keyword generation*. Manuscript submitted for publication.

Campos, A., González, M. A., & Amor, A. (2002). The Spanish version of the Vividness of Visual Imagery Questionnaire (VVIQ): Factor structure and internal consistency reliability. *Psychological Reports, 90*, 503-506.

Campos, A., & Pérez, M.J. (1996). La estrategia de imágenes mentales extrañas como ayuda al recuerdo de pares asociados. *Apuntes de Psicología, 48*, 41-50.

Campos, A., & Pérez, M.J. (1997). Mnemonic images and associated pair recall. *Journal of Mental Imagery, 21*, 73-82.

Carney, R. N., & Levin, J. R. (1998). Do mnemonic memories fade as time goes by? Here's looking anew! *Contemporary Educational Psychology, 23*, 276-297.

Cornoldi, C., Cavedon, A., De Beni, R., & Pra Baldi, A. (1988). The influence of the nature of material and of mental operations on the occurrence of the bizarreness effect. *Quarterly Journal of Experimental Psychology, 40A*, 73-85.

Denis, M. (1979). *Les images mentales*. Paris: Puf.

Ernest, C.H. (1977). Imagery ability and cognition: A critical review. *Journal of Mental Imagery, 2*, 181-216.

Ernest, C.H. (1991). Ability differences and prose learning. *Intelligence, 15*, 455-477.

González, M.A., Campos, A., & Pérez, M.J. (1997). Mental imagery and creative thinking. *Journal of Psychology, 131*, 365-370.

Hall, J.W. (1988). On the utility of the keyword mnemonic for vocabulary learning. *Journal of Educational Psychology, 80*, 554-562.

Hall, J.W., Wilson, K.P., & Patterson, R.J. (1981). Mnemotechnics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology, 73*, 345-357.

Higbee, K.L. (1993). *Your memory*. New York: Paragon House.

Howe, M.L., Courage, M.L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology, 36*, 778-792.

Hwang, Y., Renandya, W.A., Levin, J.R., Levin, M.E., Glasman, L.D., & Carney, R.N. (1999). A pictorial mnemonic numeric system for improving students' factual memory. *Journal of Mental Imagery, 23*, 45-69.

Iaccino, J.F., Dvorak, E., & Coler, M. (1989). Effects of bizarre imagery on the long-term retention of paired associates embedded within variable contexts. *Bulletin of the Psychonomic Society, 27*, 114-116.

Imai, S., & Richman, C.L. (1991). Is the bizarreness effect a special case of sentence reorganization? *Bulletin of the Psychonomic Society, 29*, 429-432.

King-Sears, M.E., Mercer, C.D., & Sindelar, P.T. (1992). Toward independence with keyword mnemonics: A strategy for science vocabulary instruction. *Remedial and Special Education, 13*, 22-33.

Kroll, N.E.A., Jaeger, G., & Dornfest, R. (1992). Metamemory for the bizarre. *Journal of Mental Imagery, 16*, 173-190.

Kroll, N.E.A., Schepeler, E.M., & Angin, K.T. (1986). Bizarre imagery: The misremembered mnemonic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 42-53.

Levin, J.R., Pressley, M., McCormick, C.B., Miller, G E., & Shriberg, L.K. (1979). Assessing the classroom potential of the keyword method. *Journal of Educational Psychology, 71*, 583-594.

Marchal, A., & Nicolas, S. (2000). Is the picture-bizarreness effect a generation effect? *Psychological Reports, 87*, 331-340.

Marks, D. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

Marshall, P.H., Nau, K.L., & Chandler, C.K. (1980). A functional analysis of common and bizarre visual mediators. *Bulletin of the Psychonomic Society, 15*, 375-377.

McDaniel, M.A., DeLosh, E.L., & Merritt, P.S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1045-1056.

McGivern, J.E., & Levin, J.R. (1983). The keyword method and children's vocabulary learning: An interaction with vocabulary knowledge. *Contemporary Educational Psychology, 8*, 46-54.

McKelvie, S.J. (1995).The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery, 19*, 1-106.

Mercer, C. (1996). The bizarre imagery. *Journal of Mental Imagery, 20*, 141-152.

O'Brien, E.J., & Wolford, C.L.R. (1982). Effect of delay in testing on retention of plausible versus bizarre mental images. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 148-152.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Poon, L.W., & Walsh-Sweeney, L. (1981). Effects of bizarre and interacting imagery on learning and retrieval of the aged. *Experimental Aging Research, 7*, 65-70.

Pra Baldi, A., De Beni, R., Cornoldi, C., & Cavedon, A. (1985). Some conditions for the occurrence of the bizarreness effect in free recall. *British Journal of Psychology, 76*, 427-436.

Pressley, M., Levin, J.R., & Delaney, H.D. (1982). The mnemonic keyword method. *Review of Educational Research, 52*, 61-91.

Pressley, M., Levin, J.R., & McDaniel, M.A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary instruction* (pp. 107-127). Hillsdale, NJ: Erlbaum.

Pressley, M., Levin, J.R., Nakamura, G.V., Hope, D.J., Bispo, J.G., & Toye, A.R. (1980). The keyword method of foreign vocabulary learning: An investigation of its generalizability. *Journal of Applied Psychology, 65*, 635-642.

Richardson, A. (1969). *Mental imagery*. London: Routledge & Kegan Paul.

Richardson, A. (1994). Individual differences in imaging. Amityville, NY: Baywood.

Riefer, D.M., & LaMay, M.L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review, 5*, 312-317.

Riefer, D.M., & Rouder, J.N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition, 20*, 601-611.

Tess, D.E., Hutchinson, R.L., Treloar, J.H., & Jenkins, C.M. (1999). Bizarre imagery and distinctiveness: Implications for the classroom. *Journal of Mental Imagery, 23*, 153-170.

Thomas, M.H., & Wang, A.Y. (1996). Learning by the keyword mnemonic: Looking for long-term benefits. *Journal of Experimental Psychology: Applied, 2*, 330-342.

Thurstone, L.L., & Thurstone, T.G. (1962). *Primary mental abilities*. Chicago, IL: Science Research Associates. [Spanish translation: *Aptitudes Mentales Primarias*. Madrid: TEA Ediciones (1989).]

Valle, F. (1998). *Normas de imaginabilidad*. Oviedo, Spain: University of Oviedo.

Wang, A.Y., Thomas, M.H., & Ouellette, J.A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology, 84*, 520-528.

Webber, S.M., & Marshall, P.H. (1978). Bizarreness effects in imagery as a function of processing level and delay. *Journal of Mental Imagery, 2*, 191-300.

Wollen, K.A., & Cox, S.D. (1981a). The bizarreness effect in a multitrial intentional learning task. *Bulletin of the Psychonomic Society, 18*, 296-298.

Wollen, K.A., & Cox, S.D. (1981b). Sentence cuing and the effectiveness of bizarre imagery. *Journal of Experimental Psychology: Human, Learning, and Memory, 7*, 386-392.