

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221875385>

Teaching Students About Facial Composites Using the FACES Software

Article in *Teaching of Psychology* · March 2012

DOI: 10.1177/0098628312437701

CITATION

1

READS

405

5 authors, including:



Jaume Masip

Universidad de Salamanca

108 PUBLICATIONS 2,316 CITATIONS

[SEE PROFILE](#)



Eugenio Garrido

Universidad de Salamanca

69 PUBLICATIONS 1,355 CITATIONS

[SEE PROFILE](#)



Carmen Herrero

Kintsugi Therapeutics

62 PUBLICATIONS 2,320 CITATIONS

[SEE PROFILE](#)



Ana M Ullán

Universidad de Salamanca

56 PUBLICATIONS 603 CITATIONS

[SEE PROFILE](#)

Running head: TEACHING ABOUT FACIAL COMPOSITES

Teaching Students about Facial Composites Using the FACES Software

Jaume Masip, Eugenio Garrido, Carmen Herrero, Ana M. Ullán, and Jorge Conde

University of Salamanca, Spain

Teaching of Psychology, 39 (2), 137-141.
Copyright: The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0098628312437701
<http://top.sagepub.com>

1. The teaching activity described in this report was supported by the University of Salamanca, Convocatoria de Ayudas para la Innovación Docente 2009/2010 (Ref. ID9/155). The authors are grateful to two anonymous reviewers for their helpful comments.
2. Send correspondence to Jaume Masip, Department of Social Psychology and Anthropology, University of Salamanca, Facultad de Psicología, Avda. de la Merced, 109-131, 37005 Salamanca, Spain; e-mail: jmasip@usal.es

Abstract

An active learning exercise was carried out in an eyewitness psychology course in which students first built up a facial composite of a famous person using the FACES software. Then the students had to name the person depicted in each composite. The results of this exercise were then described by the instructor during a theoretical lecture about facial composites. The students experienced for themselves how difficult it is to build and identify facial composites of familiar faces. Pre-post analyses showed that the exercise was effective in changing students' initially optimistic beliefs about the utility of facial composites.

Teaching Students about Facial Composites Using the FACES Software

Drawing on the early work of education reformists such as Dewey (1938) and Lewin (1942), a number of alternative approaches to traditional teaching methods emerged during the late 20th century (Barrows & Tamblyn, 1980; Bruner, 1961; Kolb, 1984; Tai, 1985). What these approaches have in common is that they all are forms of active learning (Bonwell & Eison, 1991) in which the students are strongly involved in the learning process and experience for themselves the phenomena that are being taught. Our purpose was to create an active learning exercise in which students had to build a facial composite using a specific computer program, and then had to try to identify the individuals depicted in the composites of their peers. The exercise can be used in forensic psychology, applied cognitive psychology, or policing courses. Although its core element was the experiential one, the exercise was strongly structured and also involved direct instruction. In this report, we describe this exercise and provide empirical evidence of its effectiveness. The purposes of this exercise are to make students aware of the limited utility of facial composite systems, and to illustrate certain findings of psychological research on facial composites.

Background

When creating the likeness of a criminal suspect from a witness's description, police may resort to facial composite systems (see Davies & Valentine, 2006, or Wells & Hasel, 2007, for overviews). A distinction must be made between mechanical and software systems (e.g., Davies & Valentine, 2006). The former contain a series of facial features either sketched on acetates that can be superimposed to build the suspect's face (Identikit) or photographed on cards that can be assembled together in a jigsaw-like fashion. Modern software systems (such as Mac-a-Mug Pro, E-Fit, or FACES) also contain individual facial features, but they are

displayed on a computer screen and take advantage of sophisticated computer graphics utilities to build the facial composite.

The utility of mechanical systems is poor (Davies & Valentine, 2006). One possible reason for this is that the range and representativeness of the facial features contained in these systems is limited. This makes it literally impossible to build certain faces. This drawback has been addressed by the software systems; however, their utility is also poor. It has been argued that faces are normally encoded and retrieved as a whole (Tanaka & Farah, 2003), but building a facial composite requires witnesses to recall the faces feature-by-feature. This is difficult for witnesses because it is at odds with both the witness' default mode of functioning and the way the target face was encoded.

Nonetheless, the police often use current software systems to construct facial composites. The media (by only reporting successful cases) and movies often convey the idea that facial composite software systems are nearly infallible; this may generate misconceptions in the public. Building a composite is difficult and time-consuming, the degree of likeness with the target face tends to be low, and the recognition rate is poor (Davies & Valentine, 2006). The present exercise required students to build a facial composite of a celebrity and then try to identify the composites created by their peers. Recognition rates were later presented to the students during a lecture. We used the versatile FACES software, marketed by the US firm IQ Biometrix, Inc. (<http://www.facesid.com>). According to IQ Biometrix, FACES is being utilized by “thousands of police agencies worldwide—including the CIA, FBI and the US Military” (http://www.iqbiometrix.com/products_faces_40.html). Specifically, we used the education modality of the current version of the program (FACES 4.0 EDU).

Predictions

During the recognition session, students first named the famous person depicted in each composite (Naming Task), rated the extent to which each celebrity was famous (Famousness-Rating Task), and then chose the name of the celebrity depicted in each composite from a list (Recognition Task). We sought to replicate previous empirical findings with our students as participants. The students would later receive feedback. This experience was expected to make them aware of the limitations of current facial composite software systems. Based mostly on the results of experiments using similar facial composite programs, we posed the following hypotheses:

Hypothesis 1. Accuracy rates in the naming task will be low (see reviews by Davies & Valentine, 2006; Wells & Hasel, 2007; Wells, Memon, & Penrod, 2006).

Hypothesis 2. The reason for the low accuracy rate will not be that the stimulus persons are not famous. Poor accuracy rates have been reported of composites of celebrities, and people perform poorly at naming people they know well from composites (Davies & Oldman, 1999; Frowd et al., 2005; Kovera, Penrod, Pappas, & Thill, 1997). Thus, although we expected the stimulus persons to be rated as famous (Hypothesis 2), we also expected the correlation between famousness ratings and identification to be non-significant (as found by Frowd et al., 2005).

Hypothesis 3. The students will show an exaggerated confidence in their ratings. This prediction was based on evidence from a variety of domains showing that people tend to be overconfident in their judgments (Hoffrage, 2004).

Hypothesis 4. In the recognition task, accuracy rates will be significantly greater than in the naming task.

Hypothesis 5. In the recognition task, accuracy rates will be significantly greater than chance.

Hypotheses 4 and 5 are based on studies (e.g., Frowd et al., 2005) showing that in naming tasks, where the number of possible targets is virtually infinite, performance is lower than in tasks in which the number of possible target faces is reduced (e.g., tasks in which the composite is compared with an array of faces containing the target face). In the latter case, performance is expected to be greater than chance (e.g., Frowd et al., 2005).

Hypotheses 4 and 5 were posed to show that the reason behind poor accuracy in the naming task would not be poor likeness alone (in this respect, see the reviews by Davies & Valentine, 2006; Wells & Hasel, 2007; Wells et al., 2006), as there might be multiple causes of error in the identification of a person depicted in a composite picture. Performance is normally greater in recognition tests than in recall tests like our naming task. However, if the likeness were null (no resemblance at all between the target person and the facial composite), then accuracy would be extremely poor both in recall and in recognition tests. Support for Hypotheses 4 and 5 would indicate that the composites did resemble the target faces to some extent.

Hypothesis 6. Most importantly for the purposes of the classroom exercise, we also predicted that the students would have more accurate views about the usefulness of facial composites after the exercise than before it.

Classroom Exercise

Method

Participants

Participants were undergraduate students enrolled in an Eyewitness Psychology course at the University of Salamanca. Although the number of students attending each session varied

(see below), the data were collected in Session 3 from 32 participants (29 women and three men; M age = 23 years, SD = 4.28).

Procedure

We bought twenty copies of the FACES program and installed them on the computers of a computer room in the University of Salamanca (Spain). The present exercise was structured in four sessions.

Session 1: Training. The purpose of this session was to make students familiar with the FACES program. Two separate groups of students attended this session at different times. The instructor introduced the FACES program to them, explained how it worked, invited them to practice freely, and answered all of their questions. Then the students played a game that comes with the program and consists of re-building a composite shown only for a few seconds. Finally, each student was asked by one of her peers to make a specific change in a composite. This session took around 50 min.

Session 2: Composite building. Twenty-seven of the students who had attended Session 1 also attended Session 2. The instructor gave a card to each student. Each card had the name of a famous person written on it. Fourteen female names and 13 male names were distributed to the students. Their task was to make a composite of the person whose name was on the card. They had around 50 min. to perform this task. Afterwards, each student saved the composite in jpg format (with the famous person's name as filename) and e-mailed it to the instructor. The purposes of this session were: (a) to make the students aware of how difficult it is to build a facial composite and (b) to obtain the stimuli for Session 3.

Before Session 3, three PowerPoint presentations were prepared. The *composite presentation* contained all of the facial composites in random order. There was one composite per slide. The *composite-and-name presentation* also contained one composite per slide, but

the male composites (i.e., those depicting a famous male) were accompanied with a written list of the names (in random order) of all the famous males whose composites were included in the presentation. The same was true for the female composites. Finally, in the *composite-and-picture presentation*, each slide contained both the composite and a photograph of the stimulus person.

Session 3: Assessment. Thirty-two students attended Session 3, in which four tasks were performed.

Naming Task. The composite presentation was shown to the students. For each facial composite, they had to indicate in a questionnaire (a) who that famous person was (write a name or write “I don’t know”), and (b) if they wrote a name, they also had to indicate their confidence on a 1 (*not confident at all*) to 5 (*fully confident*) scale. The students were instructed not to rate their own composites, but to write instead that this was his/her composite. The completed questionnaires were collected before proceeding to the next task.

Famousness-Rating Task. The students were asked to complete another questionnaire with the names of all the celebrities whose composites were in the presentations. For each name the students had to rate how famous that person was on a 1 (*little known*) to 5 (*very famous*) scale. The completed questionnaires were collected before proceeding to the next task.

Recognition Task: This task was analogous to the naming task, but the composite-and-name presentation was shown instead of the composite presentation. Therefore, this time the students had a list of names to choose from.

Feedback: After collecting the questionnaires, the composite-and-picture presentation was shown to the students. To dynamize the session, the students were asked to indicate by

raising their hands whether they had been able to recognize each person in the naming and recognition tasks.

This session took approximately 55 min. Its purpose was to show the students how difficult it is to identify a familiar face in a composite.

Session 4: Explanation. The results of the above exercise were included in the theoretical lecture about facial composites. Several mechanical (Identikit and Photofit) and software (Mac-a-Mug, E-fit, and FACES) systems were described, and empirical results regarding their utility were provided. When describing the FACES program, the empirical results of the above study (whose participants were the students themselves) were presented.

Results and Discussion

As predicted in Hypothesis 1, accuracy rates in the naming task were low ($M = 15.26\%$; range across raters: 0 - 33%; across stimuli: 0 - 63%). Fifty-six percent of the composites were recognized by fewer than 10% of the students. The reason for the low accuracy was not that the stimulus persons were not famous enough because (a) in support of Hypothesis 2, the stimulus persons were very famous ($M = 3.88$ on the 1-to-5 scale; this was significantly above the middle point on the scale, $t(31) = 7.03, p < .001, 25th\ percentile = 3.49, Mdn = 4.09$); and (b) as predicted, in the recognition task the composites of more famous individuals were not better identified than the composites of less famous individuals, $r(27) = .28, p = .156$.

However, unexpectedly, this latter correlation was significant (and positive) in the naming task, $r(N = 27) = .40, p = .042$.

The prediction that the students would show an exaggerated confidence in their ratings (Hypothesis 3) was not supported. Average confidence scores ($M = 2.58$ for the naming task, and $M = 3.27$ for the recognition ask) were transformed from the 1-5 scale to a 0-100 scale ($M = 39.41$, and $M = 56.78$, respectively), and then were compared with the overall accuracy rate

for those composites for which the students had named someone (because confidence was only measured when the students had named a person). The accuracy rates were low because the students did not name anyone, but when they named someone they were more likely to be accurate than inaccurate (for the naming task, $M = 60.95\%$; for the recognition task, $M = 67.05\%$). The confidence-accuracy comparison showed that the students were *underconfident* (for the naming task, $t [31] = 4.39, p < .001, d = 0.93$; for the recognition task, $t [31] = 3.13, p = .004, d = 0.57$). Apparently, the students' discouragement in being unable to identify most stimulus persons decreased their confidence even when their judgments were accurate. The confidence-accuracy correlation was not significant in the naming task, $r (N = 32) = .28, p = .116$, but it was significant in the recognition task, $r (N = 32) = .47, p = .007$.

As predicted in Hypotheses 4 and 5, in the recognition task the accuracy rate ($M = 52.99\%$) was significantly greater than in the naming task ($M = 15.26\%$; $t [31] = -14.76, p < .001, d = -2.22$) and significantly greater than chance (for females, $M = 53.27\%$, $t (31) = 15.42, p < .001, d = 2.73$; for males, $M = 52.74\%$, $t (31) = 14.25, p < .001, d = 2.52$).¹ It is noteworthy, however, that in the recognition task the participants still failed to identify the stimulus person roughly one half of the time.

The higher accuracy in the recognition task (relative to the naming task) suggests that the composites did resemble the stimulus persons to some extent. However, even though the students tried to make a good composite, their success was limited. One reason for this may be that faces are encoded, processed, and stored *holistically*, whereas building a facial composite requires retrieving and combining *individual* features.² Currently, new software systems are being developed and tested that account for the way humans process faces. These systems create consecutive generations of faces using genetic algorithms. The eyewitness selects the faces in each generation that most closely resemble the culprit, and then a new set of faces

(generation) is bred from these faces. The process goes on until all the faces in a generation are equally similar to the culprit (Frowd, Hancock, & Carson, 2004; Gibson, Pallares, & Solomon, 2003). This was also explained to the students in Session 4.

Assessment

Method

Participants

Participants were 52 students (46 women and 5 men; one participant did not report gender) enrolled in the Eyewitness Psychology course at the University of Salamanca who had participated in at least one session of the exercise.

Procedure

A questionnaire was designed and administered to the students the first day of the semester and then the day of the exam (two weeks after Session 4). The students were asked to estimate (1) the percentage of police cases in which facial composites are used that are solved because of the facial composites (0-to-100 scale), (2) the time it takes building a composite of a familiar person (e.g., a famous individual) (less than 5 min, 5-to-15 min, 15-to-30 min, 30-to-60 min, or more than 60 min), (3) how difficult it is to make a facial composite of a familiar face (0-to-10 scale), (4) how similar the facial composite of a familiar face will be to the target face (0-to-10 scale), (5) how likely it is that a familiar person will be identified from a facial composite (0-to-10 scale), and (6) whether more famous individuals would be better identified from facial composites than less famous individuals (*No* vs. *Yes*).

Results

We compared the participants' end-of-semester ratings with their initial ratings (Table 1). Regarding Questions 1, 3, 4, and 5, it is readily apparent that the exercise had the desired effect of substituting the students' misconceptions with more accurate information. In

response to Question 2, at the beginning of the semester there was no significant trend among the students to select any specific category, but at the end of the semester fewer students than expected selected the less-than-15-min. category and more students than expected selected the more-than-60-min. category. As for Question 6, at the beginning of the semester, 90% of the students indicated that more famous individuals would be better recognized than less famous ones. This coincided with the results of the naming task, but not with those of the recognition task. As a stronger emphasis was placed during Session 4 on the results of the naming task (which was more ecologically valid than the recognition task), many students did not change their view. Overall, the exercise was quite effective, as predicted in Hypothesis 6.

Conclusions

By participating in this activity, the students received first-hand experience with facial composites and saw for themselves what their utility is. This exercise can be used in forensic psychology, applied cognitive psychology or policing courses. Additions and variations can be made. For example, one sample of students can make the composites from memory and another sample while watching a picture of the stimulus person. Also, one group of students can assist the instructor in preparing the materials, introducing the data in the computer, and running the analyses; these students would improve their experimental research skills.

References

- Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*. New York: Springer.
- Bonwell, C., & Eison, J. (1991). *Active learning: Creating excitement in the classroom*. *AEHE-ERIC Higher Education Report No.1*. Washington, DC: Jossey-Bass.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.
- Davies, G., & Oldman, H. (1999). The impact of character attribution on composite production: A real world effect? *Current Psychology*, 18, 128-139.
- Davies, G. M., & Valentine, T. (2006), Facial composites: Forensic utility and psychological research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology* (vol. 2, pp. 59-96). Mahwah, NJ: Erlbaum.
- Dewey, J. (1938). *Experience and education*. New York: Collier Books.
- Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Psychology*, 1, 1–21.
- Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. (2005). Contemporary composite techniques: The impact of forensically-relevant target delay. *Legal and Criminological Psychology*, 10, 63-81.
- Gibson, S., Pallares, A., & Solomon, C. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey & J. A. Bangham (Eds.), *Proceedings of the British Machine Vision Conference* (pp. 221-230). London: BMVA.
- Hoffrage, U. (2004). Overconfidence. In Pohl, R. F. (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 235-254). New York: Psychology Press.

- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.
- Kovera, M. B., Penrod, S., Pappas, C., & Thill, D. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82*, 235-246.
- Lewin, K. (1942). Field theory and learning. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers* (pp. 212-230). London: Social Science Paperbacks.
- Tai, V. (1985). Update on participative training methods. *Journal of European Industrial Training, 9*(7), 13-16.
- Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 53-74). Oxford: Oxford University Press.
- Wells, G. L., & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science, 16*, 6-10.
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence. Improving its probative value. *Psychological Science in the Public Interest, 7*, 45-75.

Footnotes

¹ Chance was 1/14 for the composites of female individuals, as there were 14 such composites in the composite-and-name presentation and each of them was accompanied by 14 names, and 1/13 for the composites of male individuals.

² One reviewer noted that we did not directly test the notion that faces are processed holistically. Indeed we did not, as the main purpose of this exercise was just providing the students with a classroom demonstration. Others have directly tested this idea (see Tanaka & Farah, 2003). We only claim that our findings are consistent with this notion.

Table 1

Analyses of the Pedagogical Effectiveness of the Exercise

Question	Beginning of Semester	End of Semester	Contrasts
1. Percentage of cases solved	47.88%	24.04%	$t(51) = 8.88, p < .001, d = 1.46$
2. Time			
< 15 min. ^a	13.47%	0.00%	Beginning of semester: $\chi^2(3, N = 52) = 4.31, p = .230$
15-30 min.	32.69%	19.23%	End of semester: $\chi^2(3, N = 52) = 26.00, p < .001$
30-60 min.	28.85%	32.69%	
> 60 min.	25.00%	48.08%	
3. Difficulty	6.48	7.40	$t(51) = -2.74, p = .008, d = -0.48$
4. Likeness	6.42	4.94	$t(51) = 4.75, p < .001, d = 0.83$
5. Identification	6.81	5.19	$t(51) = 5.10, p < .001, d = 0.86$
6. The more famous the better identification?			
No	9.62%	28.85%	Beginning of semester: $\chi^2(1, N = 52) = 33.92, p < .001$
Yes	90.38%	71.15%	End of semester: $\chi^2(1, N = 52) = 9.31, p = .002$

^a Because of the small frequency for the less-than-5-min. category ($n = 2$ [3.85%] at the beginning of the semester; $n = 0$ at the end of the semester), the less-than-5-min. category was pooled with the 5-to-15-min. category.