Computers & Education xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com

Analyzing the influence of a visualization system on students' emotions: An empirical case study

Carmen Lacave[a,*], J. Ángel Velázquez-Iturbide[b], Maximiliano Paredes-Velasco[b], Ana I. Molina[a]

[a] Department of Information Technology and Systems; University of Castilla – La Mancha, Paseo de la Universidad, s/n, 13071, Ciudad Real, Spain
[b] Department of Languages and Computing Systems I; Rey Juan Carlos University, C/ Tulipán, s/n, 28933, Móstoles (Madrid), Spain

1. Introduction

Emotions are present in many aspects of our lives and they are a fundamental factor in education (Finch, Peacock, Lazdowski, & Hwang, 2015; Pekrun, 2014, p. 24). It is difficult to define 'emotion', but it can be understood as a complex set of interactions with subjective and objective factors mediated by a hormonal/neural system (Kleinginna & Kleinginna, 1981). Although emotions are a complex construct, they can be roughly classified into positive and negative emotions, as in the PANAS (Positive And Negative Affect Schedule) questionnaire (Watson, Clark, & Tellegen, 1998). Positive emotions are related to enthusiasm, activity, and alertness, whereas negative emotions are linked to anger, contempt, disgust, guilt, fear, and nervousness.

The role of emotions in learning has not been sufficiently studied so far. However, it has received increasing empirical and theoretical attention in recent years, suggesting that emotion plays both a positive and a negative role in the learning process (Ishkov & Magera, 2015; Martin, Hughes, & Richards, 2017; Rowe & Fitness, 2018). On the one hand, students who have positive emotional reactions to learning exhibit enhanced abilities to achieve successful outcomes, to develop higher problem-solving skills and are more engaged with the learning experience (Pekrun, Goetz, Titz, & Perry, 2002). Therefore, 'a goal of teaching [should be] to enhance the students' pleasant achievement outcomes' (Frenzel, Goetz, Ludtke, Pekrun, & Sutton, 2009). On the other hand, there is a debate about the effect of negative emotions on learning (Finch et al., 2015), with some authors arguing that negative emotions are a negative factor to avoid, and others arguing that negative emotions can increase student's motivation. In general, negative emotions are held to be detrimental to the pursuit of achievement goals, investment of effort, cognitive processes (such as attention and memory), motivation, self-regulation and self-efficacy.

The current study focuses on emotions in programming education. Given the abstract nature of programming, software visualization has been advocated as a promising technology to ease its learning process. However, the effect of software visualization on students' emotions has rarely been raised. Studies in the field of educational software visualization have focused on system usability (Urquiza- Fuentes & Velázquez-Iturbide, 2009) and on students' outcomes (Hundhausen, Dougkas, & Stasko, 2002; Urquiza- Fuentes & Velázquez-Iturbide, 2009). The most frequently advocated factor for educational success of visualizations has been students' engagement (Naps et al., 2003; Sorva, Karavirta, & Malmi, 2013). While Hundhausen et al.'s well-known meta-analysis considered general educational theories (Hundhausen et al., 2002), Hidalgo-Céspedes et al.'s (Hidalgo-Céspedes, Marín-Raventós, & Lara-Villagrán, 2016) have more recently surveyed fifteen specific learning principles. Most of their evaluation criteria addressed educa-

* Corresponding author.
E-mail addresses: carmen.lacave@uclm.es (C. Lacave); angel.velazquez@uclm.es (J.Ángel Velázquez-Iturbide); maximiliano.paredes@urjc.es (M. Paredes-Velasco); anaisabel.molina@uclm.es (A.I. Molina)

https://doi.org/10.1016/j.compedu.2020.103817
Received 11 May 2019; Received in revised form 20 January 2020; Accepted 26 January 2020
Available online xxx
0360-1315/© 2020.

tional effectiveness, with only two factors considering students' motivation and interest. However, in their review, no study had addressed motivation and only four had partially fostered students' interest. Only recently, some authors have addressed students' motivation (Velázquez-Iturbide, Hernán-Losada, & Paredes-Velasco, 2017). Finally, studies of professional visualization of complex software systems (Merino, Ghafari, Anslow, & Nierstrasz, 2018) have just focused on correctness and performance.

In this work we want to investigate the effect of algorithm visualization tools on the emotions perceived by students with two main purposes: firstly, to improve their motivation and performance; and secondly, to provide the educational community with an experience design which allows analyzing the variations in emotions when an algorithm visualization tool is used. Similar to other works (Uzun & Yildirim, 2018), we have not focused on specific types of emotions, but on their sign (either positive or negative). We conducted our research in the context of an algorithms course, focusing on the emotions experienced by students when they address complex topics. We selected backtracking because it is a difficult topic (Decker & Simkins, 2016) and students often build a number of misconceptions on the technique (Velázquez-Iturbide, 2019). To alleviate limitations of previous research (Bosch & D'Mello, 2017), our evaluation was conducted in a realistic learning scenario simultaneously in two different universities. We used an interactive visualization tool that supports a number of recursive, combinatorial algorithms, which are at the basis of the backtracking technique.

The main research question addressed was whether algorithm visualization tools impact positively on students' emotions while they learn complex algorithmic concepts, stating the following hypothesis:

H: Students' emotions while they learn complex algorithmic issues are positively influenced by the use of a visualization tool.

Considering that emotions can be classified as either positive or negative (Watson et al, 1998), hypothesis **H** can be decomposed into the following two constructs:

HP: Students' *positive emotions* while they learn complex algorithmic issues increase with the use of a visualization tool.
HN: Students' *negative emotions* while they learn complex algorithmic issues decrease with the use of a visualization tool.

Furthermore, in order to better understand the role of emotions in algorithm learning, we wanted to identify those (subjective and objective) factors that can have any influence on students' emotions, such as their subjective perception of their own knowledge, their real knowledge, their assessment of the tool, their subjective perception of the experience, the learning context and their personal profile (Davis, 1993; Frenzel, Taxer, Schwab, & Kuhbandner, 2018; Liaw, Huang, & Chen, 2007; Verdegem & De Marez, 2011). Consequently, the following additional hypotheses arise:

HEC: Students' educational context affects their emotions.
HPP: Students' personal profile affects their emotions.
HSP: Students' subjective perception of their own knowledge influences their emotions.
HPK: Students' previous knowledge of the subject influences their emotions.
HVB: Students' evaluation of the visualization tool influences their emotions.
HPA: Students' subjective perception of their own performance in the activity influences their emotions.

The contribution of the work is multiple. Firstly, we found evidence that negative emotions decreased after using the visualization tool, while positive emotions remained invariant. Secondly, we identified a factor that influenced students' emotions while they learnt a complex algorithmic design technique mediated by a visualization tool: the valuation of the visualization tool. Thirdly, the experience design is meticulously described so that it can be replicated in a different context or with a different visualization tool. In particular, two scales were developed and calibrated (about students' subjective perception of their knowledge of the topic and students' valuation of the visualization tool).

The article is organized as follows: in section 2, we describe in depth the quasi-experimental design carried out; section 3 explains how the research hypotheses has been performed; section 4 illustrates the replication of the experience; section 5 shows and discuss the results obtained and exposes the main threats to validity; and, finally, section 6 presents the main conclusions reached.

2. Background

A number of issues regarding emotions in education have been studied. Thus, it seems that motivation and emotions are correlated (Fanselow, 2018). Furthermore, emotions are one of the most common factors that influence students' success in learning new materials (Iskrenovic-Momcilovic, 2018). It has been found that emotions provide greater predictive value than cognition or motivation (Pekrun, Elliot, & Maier, 2006; Ruiz et al., 2016). Furthermore, emotions heavily depend on teachers' implication (Frenzel et al., 2018; Mainhard, Oudman, Hornstra, Bosker, & Goetz, 2018; Nadelson, Hardy, & Yang, 2015).

Various studies have addressed emotions in computer-supported education. Thus, there is evidence that emotions experienced by students during their interaction with the computer are very important in their learning process (Pekrun, 2014). It is a complex issue because emotions are related to different factors; for instance, computer science students suffer anxiety due to learning to program, computer usage, learning mathematics or tests (Nolan & Bergin, 2016). Not only emotions can influence students, but they also affect teachers. Thus, contradictory emotions or attitudes have been detected in educators regarding ownership of knowledge in open educational resources (Pirkkalainen, Pawlowski, & Pappa, 2017) or adoption of algorithm visualization systems (Ben-Bassat Levy & Ben-Ari, 2007).

Emotions have become a factor of research in multimedia learning, where different levels of emotional design in multimedia materials have been evaluated (Uzun & Yıldırım, 2018). An important factor to foster positive emotions with multimedia contents is to include videos with professional or funny contents (Adnan & Redzuan, 2017). In this sense, emotional design can play an important role (Mayer & Estrella, 2014; Plass, Heidig, Hayward, Homer, & Um, 2014; Stark, Brünken, & Park, 2018), since its objective is to create products and materials that generate positive emotions on the users (Norman, 2005). However, research is scarce, especially regarding negative emotions (Park, Knörzer, Plass, & Brünken, 2015).

A limitation of much of the available empirical evidence is that it has been undertaken with primary and secondary school students and therefore remains to be tested with undergraduate students (Rowe & Fitness, 2018). There are some studies concerning emotions in engineering education, such as the experience of using a competition-oriented pedagogical methodology (by means of gamification) with students of telecommunications engineering (Muñoz-Merino, Fernández, Muñoz-Organero, & Delgado, 2014). Moderate levels of motivation were found to be positively correlated with low levels of negative emotions. Moreover, inverse correlations were found between scores obtained in the game and negative emotions. Emotions also influence students' expectations about and attitude towards a subject (Nadelson et al., 2015). It has also been noticed that engineering students with higher level of positive emotions at the beginning of a course correlated negatively with their stress at the end of the course (Husman, Cheng, Puruhito, & Fishman, 2015).

Within computer science education, emotional response has been studied with respect to learning programming. Students' emotional reactions often are related to the frustration of dealing with the difficulties that are faced to solve programming problems. For example, students' emotions in their first programming session with Python were analyzed (Bosch, D'Mello, & Mills, 2013; Bosch & D'Mello, 2017) and it was concluded that they were mostly negative (confusion, frustration and boredom), with much less positive emotion (flow/engagement). In addition, emotions varied depending on instructional scaffolding (explanations and hints) and were linked to different programming activities (idling, constructing or running) (Bosch et al., 2013). The relation between students' positive and negative experiences and their subjective self-efficacy assessment based on interviews has also been studied (Kinnunen & Simon, 2012). Although positive and negative experiences usually occurred with their respective positive or negative self-efficacy assessments, it was found that some students who had a negative programming experience could maintain a positive self-efficacy judgement, while students who had a positive programming experience maintained a negative self-efficacy assessment. In a quantitative study (Lishinski, Yadav, & Enbody, 2017), students' emotional reactions were found to correlate with their performance on programming projects, with the direction of the correlation depending on the valence of the emotion.

Emotions can have a strong impact on students' performance, as they can directly cause to fail an exam and even to drop out a course (Iskrenovic-Momcilovic, 2018). Another study on learning the C programming language in an e-learning system (Zhu, Zhang, Wang, Chen, & Zeng, 2014) found that positive emotions helped learners to attain higher achievements.

Human factors (such as personality, programming style or programming attitude) that play a positive or a negative role in relation to students' programming performance have also been identified (Li, 2017). There also is evidence that personality factors correlate with students' programming style (Karimia, Baraani-Dastjerdia, Ghasem-Aghaeea, & Wagner, 2016) and that the students' emotional state influences their relationship with programming styles (Cox & Fisher, 2009). Students with a visual learning style have been found to be more enthusiastic than those with a verbal style (Zhu et al., 2014).

Instructional materials and tools may also have an emotional impact on students. Thus, materials developed for learning object-oriented programming based on emotional design criteria decreased students' negative emotions, compared to those who used classical materials (Nurminen, 2016), and it took less time to complete programming tasks (Haaranen, Ihantola, Sorva, & Vihavainen, 2015).

3. Experimental design and method

This section describes the details of the evaluation designed to address our research questions. A quasi-experiment was designed, since it was intended that all the potential users of the visualization system were able to participate during the experience, without the need of a control group. Therefore, the rationale of our evaluation design is that, once the research hypothesis was determined, the variables representing participants' emotions were measured before and after the intervention, and analyzed using different statistical techniques, mainly *hypothesis contrast* and *correlation analysis*. Since one of the main limitations associated with quasi-experiments is related to the interpretation of the results, it is necessary to consider the possibility that the effects produced are due to other factors not taken into account (Cook & Campbell, 1986). Hence, we have tried to be aware of the variables that the quasi-experimental design is not able to control.

3.1. Participants and context

The experience was conducted in the academic year 2016/17 with participants from two universities: Rey Juan Carlos University (URJC) and the University of Castilla-La Mancha (UCLM). Both universities are intensive in teaching and research activities, with 46,000 students enrolled at the URJC and 30,000 enrolled at the UCLM. The two universities offer the Degree in Computer Science, which has around 800 students at the URJC and just under 700 at the UCLM.

This degree includes a compulsory course (6 ECTS) on algorithms, which is taught in the 4th semester (2nd semester of the 2nd year). Its contents are the same in both universities, being structured around complexity analysis and several algorithm design techniques, namely divide and conquer, greedy algorithms and backtracking.

Participants were those students enrolled at the algorithm course who agreed to participate, with a total of 36 students from the URJC and 30 from the UCLM. Only 11 participants were women, 4 enrolled in the UCLM and 7 in the URJC. Two instructors were responsible of conducting the experiment, one at each university.

In order to avoid biasing the results (McCambridge, Witton, & Elbourne, 2014) and to motivate students' participation (Shull, Singer, & Sjøberg, 2007), we explicitly stated at the beginning of the experience that information collected in the experience would be treated confidentially and used exclusively for the research study. After being informed, students gave their consent to use their data. Moreover, UCLM students were rewarded with additional points (0.05% of the total) on the course grade.

3.2. The visualization tool

We decided to conduct the research in the context of one of the algorithm design techniques addressed in the course. We guessed that the visualization system could contribute with the highest emotional impact if it was tightly coupled with a given design technique. As explained in the introduction, we selected backtracking because it is a difficult (and probably the most difficult) topic for students among the algorithm design techniques covered in the course. Anecdotally, instructors consider that it is the most difficult of such techniques taught in the course.

We searched and evaluated algorithm visualization systems that support the visualization of state spaces generated and traversed by backtracking algorithms. The best positioned candidates were AI-Search (McDonald & Ciesielski, 2002), SRec (Velázquez-Iturbide, Pérez-Carrasco, & Urquiza-Fuentes, 2008) and VisBack (Pérez-Mena, 2015). Finally, we selected VisBack because it is an available and active system, it provides visualizations that mirror textbooks illustrations, and it supports a range of interaction operations (which we identify below). VisBack is a visualization system aimed at understanding backtracking algorithms. VisBack displays the recursion tree for several built-in search processes that underlie many backtracking algorithms. In particular, VisBack illustrates Java implementations of algorithms for generating variations, permutations and combinations of elements, all of them with or without repetition.

Fig. 1 shows the appearance of the system user interface. The left-hand side panel shows the Java implementation of the algorithm and the right-hand side panel, the search tree generated during its execution with the entered input data. As for the graphic representation of the tree, each node shows the information associated with the corresponding recursive call. A coloring convention is used to differentiate between recursive calls where a complete solution is formed, the active recursive call and other recursive calls. Both panels are coordinated by highlighting the piece of code that best represents the active call.

VisBack supports several types of interaction with visualizations intended for educational use (Naps et al., 2003). It allows selecting the algorithm to display, entering input data and controlling the algorithm execution. Input parameters can be either typed using the keyboard or generated randomly. The user may also choose between performing the execution in one step (*normal execution*) and doing it in sequential steps (*step by step*). Moreover, he/she may advance in the execution either forwards or backwards. As a consequence, VisBack supports several forms of students' engagement. In terms of the 2DET engagement taxonomy (Sorva et al., 2013), VisBack supports Controlled Viewing in the direct engagement dimension, and Own Cases in the content ownership dimension. Finally, the user may customize certain elements of the graphical representations (color and size of some elements, and information displayed in each node of the tree).

3.3. Variables and instruments

The dependent variables of the quasi-experiment are related to the kind of emotions felt by students *before* and *after* using the visualization tool, differentiating between *positive* and *negative* emotions. Thus, we defined variables PRE_PE and PRE_NE to represent positive and negative emotions, respectively, previously to the use of VisBack. Analogously, POS_PE and POS_NE represent positive and negative emotions afterwards, respectively.

We have considered as independent variables those (objective and subjective) factors that might produce a change in students' emotions. We made an analysis of potential factors, taking into consideration that they could be extracted with accuracy in the context of an evaluation and that they were considered in the literature (Davis, 1993; Frenzel et al., 2018 ;). Although the list may be incomplete, the number and relevance of factors is not small:

- Personal and academic **profile**, considering age, gender, highest course in which the student is enrolled, and university to which he/she belongs. They are respectively represented by variables AGE, GENDER, COURSE and UNIV.
- Student's **subjective perception of his/her knowledge** on backtracking before the intervention, defined by variable GSP_BK.
- Student's performance (Iskrenovic-Momcilovic, 2018), represented by their **level of knowledge** on backtracking, both before and after using the visualization system, and defined by variables BK_PRE and BK_POS, respectively.
- **Acceptance** of the visualization system (Verdegem & De Marez, 2011), measured by participants' subjective assessment and represented by variable VB.
- Students' **attitude** (Liaw et al., 2007) during the realization of the experience, defined by variable PSA. It comprises the students' *performance subjective perception* and the *cognitive load* imposed by the task (Kinnunen & Simon, 2012; Plass, Moreno, & Brünken, 2010).

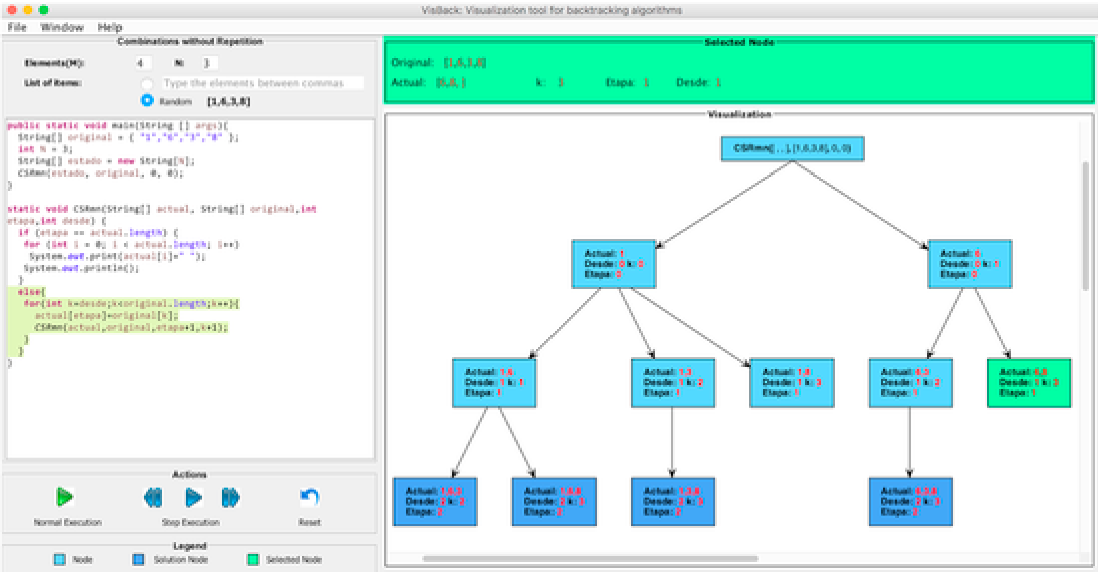

Fig. 1. VisBack visualization of 3-combinations without repetition of randomly-generated elements [1,6,3,8].

Table 1 contains the list of dependent and independent variables. All of them were measured by means of a number of questionnaires and tests. Every data-collection instrument used in the experience was delivered to students through Google Forms,[1] which enabled to conduct the experience almost in parallel in both universities. In addition, to motivate students' participation (Shull et al., 2007) and to avoid biases (McCambridge et al., 2014), each form began with two explanatory statements (see Appendix, Table 12).

The instruments used to measure the variables follow:

- Positive and negative emotions measured before (*PE_PRE*, *NE_PRE*) and after (*PE_POS*, *NE_POS*) using VisBack were gathered with the PANAS questionnaire (Watson, Clark, & Tellegen, 1988), where students rated, in a five-level Likert scale, their agreement or disagreement with ten positive and ten negative emotions (Appendix, Table 14). Although the table presents emotions classified according to their sign, the questionnaire provided to the students offered a random combination of the twenty items. Positive (*PE_PRE, PE_POS*) and negative (*NE_PRE, NE_POS*) emotions were computed as the sum of positive and negative answers, respectively.
- Students' personal and academic profile (*GENDER*, *AGE*, *COURSE*, *UNIV*) were gathered with an ad-hoc online survey.
- Students' global subjective perception of backtracking (*GSP_BK*) was computed as the sum of the answers provided to the *SSP (Students Subjective Perception) Scale* (see Appendix, Table 13). It initially consisted in ten items, rated in a five-point Likert scale.
- Students' objective *prior* and *posterior* knowledge of backtracking (*BK_PRE, BK_POS*) were defined as the number of correct answers in a quiz which consisted in 11 questions about a combinatorial problem. The problem used to measure BK_PRE asked the different routes which could be generated among several cities (see Appendix, Table 12), while to measure BK_POS, students had to obtain the different honor rolls that could be formed in a literary contest where three prizes are awarded and to which 4 writers have presented.
- Students' acceptance of VisBack was measured with the VB Scale (Appendix, Table 15), developed to subjectively assess VisBack as a tool for learning backtracking. In addition, the PSA questionnaire (Appendix, Table 16) was created to measure students' subjective overall perception of their performance. It is composed of five items extracted from several validated instruments (Liaw, 2008). Answers in the PSA and VB Scales were defined on a five-level Likert scale, representing the respondents' level of agreement or disagreement with each statement. The PSA and VB variables were computed as the sum of the answers to all the items.

Since SSP and VB Scales were specifically defined for this research experience, section 3.5 presents the analyses of their validity and reliability, which consist of various types of statistical analyses (Lacave, Molina, & Redondo, 2018) .

[1] https://www.google.com/forms/about/.

Table 1
Summary of dependent and independent variables, their names and values.

Student Aspect	*Variable*	*Name*
Emotions before using the visualization tool	Previous positive emotions	PE_PRE
	Previous negative emotions	NE_PRE
Emotions after using the visualization tool	Posterior positive emotions	PE_POS
	Posterior negative emotions	NE_POS
Profile	Gender	GENDER
	Age	AGE
	Highest enrolled course	COURSE
	University	UNIV
Previous knowledge on backtracking	Perceived knowledge level	GSP_BK
	Previous test results	BK_PRE
Posterior knowledge on backtracking	Posterior test results	BK_POS
Visualization tool	Acceptance of the visualization tool	VB
Attitude	Subjective perception of performance and cognitive load	PSA

3.4. Phases

The quasi-experiment was divided into four phases, conducted in three sessions of 90 min each:

- **Phase 1 (phase of *preparation*).** The instructors presented in the classroom the foundations of the algorithmic scheme of backtracking. They were illustrated with some classic examples, such as the n-queens problem (Golomb & Baumert, 1965) or the sum-of-subsets problem (Horowitz & Sahni, 1974). This phase took place in a full 90-min session.
- **Phase 2 (phase of *pre-tests*)**. The second phase took the first 50 min of a second session. The participants filled several *pre-tests* to gather data about those independent variables (Table 1) corresponding to the students' profile (GENDER, AGE, COURSE, UNIV), *and* their level of knowledge of backtracking, both subjectively (GSP_BK) and objectively (BK_PRE). Finally, students filled an emotional test to measure their sensations before using the visualization system (PE_PRE, NE_PRE).
- **Phase 3 (phase of *development*)**. In the following 10 min of the second session, the instructors presented the visualization system VisBack. Then, the participants had to perform two similar activities, each one with students interacting with a different combinatorial problem using VisBack and answering a number of questions. In terms of Bloom's taxonomy, the questions involved different understanding cognitive processes. The first problem involved permutations, and asked to obtain all the ways in which three persons could sit in a row of three seats (see the Appendix, Table 11). The second problem involved combinations, and asked to obtain the possible offensive triads that a coach coud form with 5 strikers. The students had the last 30 min of session 2 and the first 30 min of session 3 to complete these tasks.
- **Phase 4 (phase of *post-tests*)**. This phase was conducted in the remaining 60 min of the third session. Participants were asked to fill out *post-tests* which involved the variables (see Table 1) corresponding to students' posterior knowledge of backtracking (BK_POS), acceptance of VisBack (VB), and the perceived subjective attitude during the realization of the experience (PSA). Finally, students again completed the same emotional questionnaire than in phase 2 in order to measure positive and negative emotions after using VisBack (PE_POS, NE_POS).

To summarize, Fig. 2 illustrates graphically the process followed in this study.

3.5. Validity and reliability of the SSP and VB scales

In this section, we summarize the calibration and validation of the SSP and VB scales. Statistical calculations were performed with the program *IBM SPSS Statistics* Version 24.

Following the recommendations of previous researchers (Zamanzadeh et al., 2014), the panel of judges to validate the *content* of the SSP and VB scales consisted of six experts in the field of backtracking, who are faculty members at three state universities. Given that all of them agreed to keep the original dimensions and items, as well as the Likert-type scale, it was not considered necessary to calculate the content validity index.

Regarding the SSP Scale, the *Cronbach alpha value* was 0.879 (Appendix, Table 17) although it increases to 0.906 if the first three items are removed. Moreover, the homogeneity index is greater than 0.6 for each of the remaining items.

After removing the first three items of the SSP Scale, an *exploratory factor analysis* was performed to determine the number of dimensions of the reduced survey (*KMO* $= 0.832$ (>0.5), p-value of *Bartlett's test of sphericity* $= 0$). The extraction was performed for the two factors whose eigenvalues were greater than 1 (Kaiser, 1960), which explains more than 80% of the data variance, as Table 2 shows. It also describes the amount of variance explained by each item after applying an *oblique rotation* method (*normal Oblimin*).

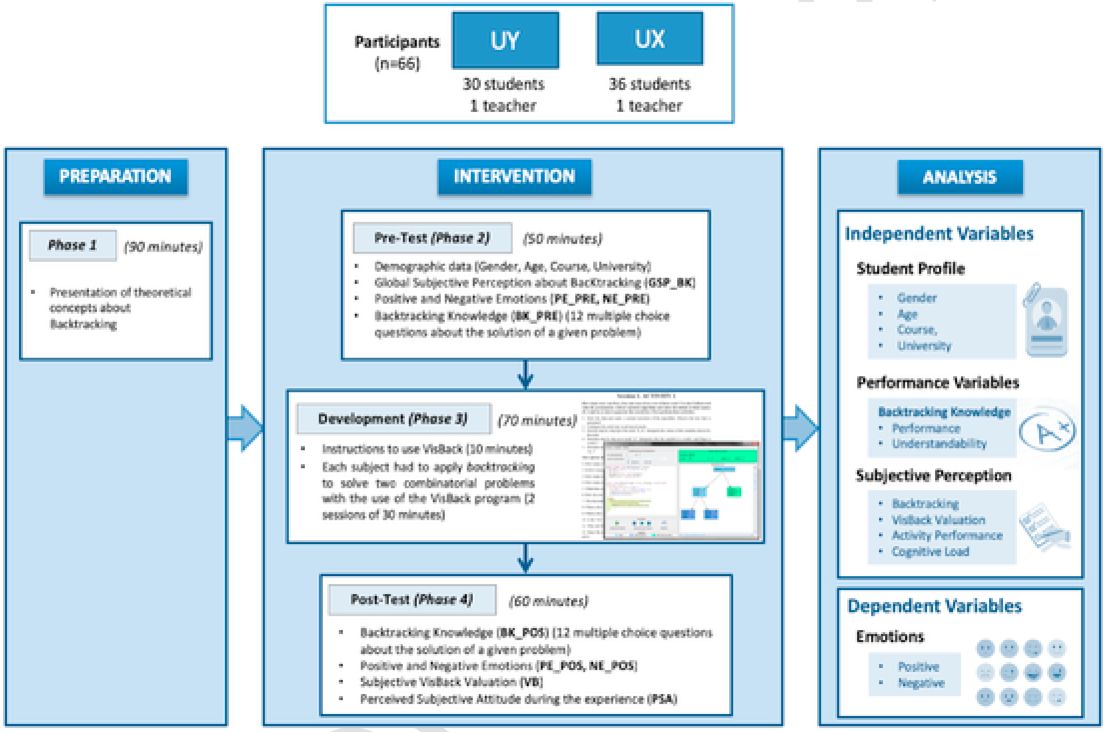

Fig. 2. Quasi-experiment design.

Table 2
Variance explained by eigenvalues and factor matrix of the reduced SSP Scale (Highest values are marked in bold)

Factor	1	2
Variance	4.494	1.126
% of Variance	64.206	16.087
B1	-.106	**.946**
B2	.066	**.888**
B3	.140	**.820**
B4	**.934**	-.042
B5	**.905**	.031
B6	**.907**	-.054
B7	**.744**	.144

This result suggests that items B1 to B7 can be better classified into two dimensions: one related to prior knowledge of general trees and the other one involving the theoretical foundations of backtracking.

In the case of the VB Scale, its *Cronbach's alpha* was 0.817, increasing to 0.845 should item V1 be removed. Therefore, data provided by items V2 to V5 yield higher internal consistency than the original five items.

Therefore, the results provided by the seven items of the SSP* Scale (Table 3) and the four items of the VB* Scale (Table 4), rated in a five-point Likert scale, are more consistent for the study sample than those of the original surveys (SSP and VB Scales).

4. Hypotheses analysis

In this section, we first present the analysis of the hypotheses based on (subjective and objective) factors associated to students' learning process, namely based on students' profile (HPP), educational context (HEC), students' subjective perception of their knowledge (HSP), students' previous knowledge of backtracking (HPK), students' opinion about the VisBack visualization tool (HVB) and students' subjective perception of the performance of the activity (HPA). Then, the major research hypotheses, related to the increase in positive emotions and the decrease in negative ones (hypotheses HP and HN, respectively), are analyzed.

4.1. External factors that might affect students' emotions

The analysis of the factors that could influence emotions was carried out by studying the linear correlations of every variable with all the others. Such correlations allow us investigating the degree of variation of all the variables to deduce whether there is any relationship between them. Thus, we first examined which variables follow a normal distribution with the *Kolmogorov-Smirnov test* (using the *Lillefors significance corrections*). Only variables PSA, VB, PE_PRE and PE_POS had a normal distribution (*p-value* = 0.69, 0.2, 0.2 and 0.2 respectively). Since most variables do not fit a normal distribution and all of them are ordinal variables, *nonparametric bivariate Kendall's Tau-b correlation coefficients* were calculated for the variables involved in the study. The choice of this coefficient is due to the fact that it considers ties. All items from *GT1* to *BK4* are positively correlated with each other and with the students' subjective perception of their knowledge of backtracking (GSP_BK). This means that students' subjective perception about backtracking is similar. For reasons of space and clarity, the rows and columns corresponding to those variables are not shown so as to focus on the most relevant results associated with emotions. Something similar occurs with the variables VB1 to VB4, which are positively correlated with each other (their columns are not shown for simplicity) and with the global valuation of VisBack (VB). This means that the students assigned a similar valuation of each item related to VisBack. Thus, we will consider only global variables GSP_BK and VB for simplicity of further analysis. Table 5 shows the results obtained for the other variables in which the coefficients representing significant correlations with at least a 95% confidence level are highlighted in bold.

Concerning emotions, it seems that they are slightly influenced by students' global subjective perception of backtracking (*GSP_BK*), by the personal assessment of the tool (*VB*) and by their university (UNIV):

- The positive correlation during the pretest (*NE_PRE)* between the subjective perception of backtracking (*GSP_BK)* and negative emotions shows that students' negative emotions during the pre-test phase become more pronounced as their subjective perception of backtracking increases. This may be due to a believe that the activity has no interest for them.
- The positive correlation during the post-test (*PE_POS)* between the global subjective perception of backtracking (*GSP_BK*) and positive emotions indicates that positive emotions in the post-test increase as students' global perception of backtracking does.

Table 3
SSP* Scale for students' subjective perception of their knowledge of backtracking.

General Trees
GT1. I know the theoretical concepts underlying general trees
GT2. I know the depth traversal algorithms for general trees
GT3. I understand the depth traversal algorithms for general trees
Backtracking
BK1. I know the theoretical concepts underlying backtracking algorithms
BK2. I know how backtracking algorithms work
BK3. I know the code of backtracking algorithms
BK4. I understand how the backtracking search tree is generated

Table 4
VB* Scale for evaluating the VisBack tool.

VisBack Evaluation
VB1. VisBack helps me in understanding backtracking
VB2. I think VisBack is useful to study backtracking
VB3. I will use VisBack to study backtracking
VB4. I will recommend VisBack to study backtracking

Table 5
Tau-b coefficients for each pair of variables. Those which are significant are marked in bold (*: p < 0.05; **: p ≤ 0.01).

	GSP_BK	BK_PRE	BK_POS	PSA	VB	PE_PRE	NE_PRE	PE_POS	NE_POS	GENDER	AGE	COURSE	UNI
GSP_BK	1	.154	-.121	.158	.011	-.044	**.185***	**.266****	-.171	-.086	.074	-.062	**.374****
BK_PRE		1	-.015	.164	.166	-.151	-.027	-.012	-.090	.037	-.093	.043	.154
BK_POS			1	-.094	.048	-.014	-.016	-.035	.066	-.067	-.066	.136	**-.329****
PSA				1	.109	-.129	.020	.102	-.133	-.069	.089	.167	.142
VB					1	-.155	-.122	.169	-.014	.169	.112	**.207***	-.048
PE_PRE						1	-.166	.029	-.019	.034	.167	.040	.139
NE_PRE							1	-.103	.164	.073	-.190	-.114	-.029
PE_POS								1	.011	.005	.054	.070	**.212***
NE_POS									1	.124	-.057	.084	**-.399****
GENDER										1	.056	**.246***	-.082
AGE											1	**.511****	.163
COURSE												1	-.095
UNI													1

- Positive (and negative) students' emotions during the post-test are positively (and negatively) correlated with university (*UNIV*), which suggests that the educational context, specially the teacher, is a factor which could affect the emotions, as it already has been proved (Mainhard et al., 2018).

Significant correlations detected have low values (most of them are less than 0.4). Therefore, it was decided to analyze the differences between emotions, considering each possible independent variable as a factor of change. For this purpose, data were compared separately for each university and a hypothesis contrast for independent samples was carried out, considering the homogeneity of the samples as the null hypothesis. The consequent analyses are described for each research hypothesis. Testing multiple hypotheses increases the probability of a rare event and therefore increases the probability of incorrectly rejecting a null hypothesis. Bonferroni's correction (Bonferroni, 1936) compensates for that increase by testing each individual hypothesis at a significant α/m level where (1-α) is the desired general confidence level and m is the number of hypotheses.

4.1.1. HEC: Students' educational context (university and course) affects their emotions

In this case, the hypothesis includes two different aspects, university and course. Therefore, we performed two independent hypothesis contrasts, one for each aspect. This meant that when considering the university as a factor of change, the Mann-Whitney test assumed as the null hypotheses the following ones:

- H01UNIV: *There were no significant differences at a 95% confidence level between positive emotions felt by students of U*RJC *before using VisBack and by students of U*CLM.
- H02UNIV: *There were no significant differences at a 95% confidence level between negative emotions felt by students of U*RJC *before using VisBack and by students of U*CLM.
- H03UNIV: *There were no significant differences at a 95% confidence level between positive emotions felt by students of U*RJC *after using VisBack and by students of U*CLM.
- H04UNIV**:** *There were no significant differences at a 95% confidence level between negative emotions felt by students of U*RJC *after using VisBack and by students of U*CLM.

Table 6 summarizes the descriptive statistics obtained, as well as the value of the statistics obtained in the contrast and its *p-value* (highlighting in bold those which are significant at 95% confidence). Since there are 4 tests and the desired value for α is 0.05, the confidence for each test must be of 98.75%. Due to the dependence of the p-values on the sample size, and following experts' recommendations (Bowman, 2017), the *effect size* of the non-parametric test (Tomczak & Tomczak, 2014) was also calculated (Lenhard & Lenhard, 2016) to measure the magnitude of the findings.

Table 6
Differences in emotions by university (Hypothesis contrast results for HEC. Significant differences are marked in bold)

Null Hypothesis *(Variable)*	UCLM (N = 30)			URJC (N = 36)			U-Mann Whitney (p-value)	Effect size (r)
	Mean	*Median*	*Standard Deviation*	*Mean*	*Median*	*Standard Deviation*		
H01UNIV *(PE_PRE)*	34.07	35	6.25	32.14	32.5	7.72	1.335 (0.182)	
H02UNIV *(NE_PRE)*	17.97	17	6.10	19.14	17	8.08	-.277 (0.781)	
H03UNIV *(PE_POS)*	32.77	32.5	6.91	30.50	30	6.68	2.044 (0.041)	
H04UNIV *(NE_POS)*	**13.07**	12	3.48	**19.47**	18.5	7.45	**−3.781 (0.001)**	**0.46**

The results illustrate that hypothesis H01UNIV, H02UNIV and H03UNIV can be assumed but H04UNIV should be rejected. The assumption of hypotheses H01UNIV, H02UNIV and H03UNIV means that the emotions felt in the pretest phase and the positive emotions felt in the posttest are similar in students from both universities. The rejection of hypothesis H04UNIV implies that, during the posttest phase, the UCLM students experienced significantly different negative emotions than URJC students (see H04UNIV row). In particular, UCLM students experienced fewer negative emotions than URJC students. Considering the effect size interpretation in the educational context (Cohen, 1988; Hattie, 2009), the rejection of hypothesis H04UNIV could be generalizable.

The analysis which considered the course as a differentiating factor defined four similar null hypotheses (H01COURSE, H02-COURSE, H03COURSE, H04COURSE) and provided no significant differences between emotions (values are not presented for clarity and space), justifying the absence of correlation between the course and emotions.

Therefore, considering these results and taking into account that hypothesis HEC includes university and course as possible factors that might affect emotions, hypothesis **HEC** should be rejected.

4.1.2. HPP: Students' personal profile (gender and age) affects their emotions

Two similar analysis were done considering gender and age, respectively, as the differentiating factor.

In the former case, the null hypotheses H01GENDER, H02GENDER, H03GENDER, H04GENDER represent the absence of significant differences at a 95% level between positive and negative emotions, before and after using VisBack, respectively, among male and female students. The results of the contrast test (whose values are not shown for clarity and space) present no significant differences at 95% confidence level (all p-values were greater than 0.0125, analyzed in a similar way as in section 4.1.1). Therefore, hypotheses H01GENDER, H02GENDER, H03GENDER, H04GENDER cannot be rejected, meaning that gender is not a factor affecting emotions.

On the other hand, the contrast test of null hypotheses H01AGE, H02AGE, H03AGE, H04AGE was carried out with the nonparametric *Kruskal-Wallis test*, the generalization of the Mann-Whitney *U* test, because the AGE variable has more than two values. The results have not revealed significant differences at 95% confidence level. Therefore, hypotheses H01AGE, H02AGE, H03AGE, H04AGE cannot be rejected, meaning that age is not a factor affecting emotions.

Therefore, given that neither gender nor age affects emotions, the hypothesis **HPP** should be rejected.

4.1.3. HSP: Students' subjective perception of their subject knowledge influences their emotions; HPK: students' previous knowledge of the subject; HVB: students' evaluation of the visualization tool influences their emotions; and HPA: students' subjective perception of their performance in the activity influences their emotions

We performed the analyses of emotions considering students' subjective perception of their knowledge of backtracking (GSP_BK), students' previous knowledge of backtracking (BK_PRE), students' evaluation of the visualization tool VisBack (VB) and students' subjective perception of their performance (PSA) as a differentiating factor, respectively. We proceeded as above: first, assuming homogeneity of the samples as the null hypothesis in each case, and afterwards applying the nonparametric *Kruskal-Wallis test*, the generalization of the Mann-Whitney *U* test (since the variables GSP_BK, BK_PRE, VB and PSA have more than two values). The p-values obtained for every test that compares emotions considering GSP_BK, BK_PRE, VB and PSA as an independent factor of change, were greater than 0.0125, showing no significant differences among emotions in each case. These results indicate that neither the students' subjective perception of their knowledge nor their assessment performance in the activity exerted any influence on students' emotions during the experience. Consequently, hypotheses **HSP**, **HPK**, **HVB** and **HPA** should be rejected.

The most important consequence of *the analysis of the external factors that might affect the students' emotions is* that the factors related to students' personal profile, their educational context, the students' subjective perception of their knowledge, their previous knowledge, their opinion about the visualization tool and students' subjective perception of their performance in the activity did not affect their emotions before and after the use of the visualization tool.

4.2. Variations in emotions before and after using the visualization tool

Having confirmed that the potential factors that could affect students' emotions did actually not influence them, we analyzed the main research hypotheses HP and HN, which state that positive emotions increase, and negative emotions decrease, respectively, when using a visualization tool.

Table 7 shows the descriptive statistics for variables PE_PRE, PE_POS, NE_PRE, NE_POS, representing the corresponding variables. An interesting fact shown by the table is that positive emotions are experienced at a greater extent than negative emotions both in the pretest and the posttest, as reported elsewhere (Watson et al., 1988).

4.2.1. HP: Students' positive emotions while learning complex algorithmic issues increase with the use of a visualization tool

Given the means of variables representing positive emotions (see Table 7), it seems that positive emotions decrease after using the visualization tool, contrary to our hypothesis HP. To measure the difference between positive emotions at the beginning and at the end of the experience, we used a *Student's T-test for related samples*, since their distributions were adjusted to normal. The results revealed at a 95% level of confidence that there were no significant differences between PE_PRE and PE_POS ($Sig = 0.217 > 0.05$). This means that the positive emotions felt by students are similar before and after using VisBack. Thus, the evidence allows rejecting hypothesis **HP.**

Table 7
Descriptive statistics for emotions.

Emotions	Variable	Mean	Std. Deviation
Positive	PE_PRE	33.02	7.109
	PE_POS	31.53	6.830
Negative	NE_PRE	18.61	7.219
	NE_POS	16.56	6.755

4.2.2. HN: Students' negative emotions while learning complex algorithmic issues decrease with the use of a visualization tool

Note that *NE_PRE* and *NE_POS* measure negative emotions before and after using VisBack, respectively. As both variables did not fit a normal distribution, the hypothesis contrast for negative emotions was performed using the nonparametric *Wilcoxon-signed range test for related samples.* The results revealed at a 95% level of confidence ($Sig = 0.043 < 0.05$) that the decrease of negative emotions experienced after using VisBack is significant.

It has been confirmed that students' negative emotions during the posttest phase were not influenced by any other factors (see Section 4.1), therefore it can be assumed that the variation in negative emotions is due to the use of VisBack, which does not allow us rejecting hypothesis **HN**. However, the effect size (Morris, 2008) is 0.313, which leads to think that the impact of the finding is small (Cohen, 1988).

4.3. Valuation of the visualization tool

The assessment of VisBack was quite good (see Table 8). Adding the two positive values in the Likert scale, 81% of the participants consider that Visback can assist them in understanding backtracking, 77% consider it useful to study, 66% would recommend it to study, and 56% would use it to study.

Moreover, the positive correlation between *VB* and *COURSE* (0.207*, see Table 5) allows us deducing that students in higher courses value the tool more than students in lower courses.

5. Experience replication

We replicated the experience in the academic year 2017/18 in order to confirm the results shown in the previous section. The number of voluntary participants was 94: 24 students from UCLM and 70 from URJC. The method and instruments were the same (see Sections 3.3 and 3.4) with the only difference that students participating from both universities obtained a small reward in their final grades in the form of additional points (0.05% of the total).

5.1. Instrument calibration

Given that the sample size was greater than in the former evaluation, we decided to calibrate again the instruments to obtain data. For the reliability analysis, the Cronbach alpha value was 0.833. and it increased to 0.875 if items R1, R2 and R3 were removed (Appendix, Table 18). Moreover, the homogeneity index was greater than 0.5 for every item. Therefore, it can be assumed that the resulting survey which contains items B1 to B7 has higher internal consistency than the original one, confirming the results of the former experience.

As in the SSP Scale calibration of the former experience, an exploratory factor analysis was applied ($KMO = .771 > 0.5$; p-value of Bartlett's test of sphericity $= 0$). The extraction was performed for the two factors whose eigenvalues were greater than 1, which explains more than 77% of the variance of the data (77.695). Table 19 in the Appendix shows the amount of variance explained by each item after applying the *normal Oblimin* rotation method.

Table 8
Frequencies and percentages of the items of the VB* Scale.

	Likert scores				
	1	2	3	4	5
VB1. VisBack helps me in understanding backtracking	0	2 (3%)	10 (15%)	30 (45%)	24 (36%)
VB2. I think VisBack is useful to study backtracking	0	2 (3%)	13 (20%)	25 (38%)	26 (39%)
VB3. I will use VisBack to study backtracking	2 (3%)	6 (9%)	21 (32%)	14 (21%)	23 (35%)
VB4. I will recommend VisBack to study backtracking	2 (3%)	4 (6%)	16 (25%)	24 (37%)	19 (29%)

These results are the same as those obtained in the original experience, confirming that:

- Students' subjective perception of the foundations of recursion does not assist them to analyze more advanced topics involving recursion, such as backtracking.
- B1 to B7 can be better classified into two dimensions: one related to prior knowledge of general trees and the other involving foundations of backtracking.

Regarding the VB Scale, the *KMO* value was 0.829 (>0.5) for the VB Scale and the significance of *Bartlett's test of sphericity* was 0. The principal factor analysis confirmed that one single factor explains 70.06% of the total variance. The reliability analysis of this scale provides a Cronbach's alpha value of 0.890, which increases to 0.917 after removing items VB1 and VB2. In this case, a posterior confirmatory analysis (with a KMO value of 0.815) provides an explanation for 85% of the total variance.

5.2. VisBack valuation

The assessment of VisBack again was quite good, as illustrated by Table 9. Thus, 78% of the students consider it useful to study backtracking, 62% would use it to study and 63% would recommend it.

5.3. Hypotheses analysis

A correlation analysis of the variables associated with students' learning process, which can influence their emotions during backtracking learning using VisBack, indicates that values given by students to their global subjective perception of backtracking (*GSP_BK*) and to their VisBack valuation (*VB*) are akin to those given in the former experience. Table 10 shows the results obtained for the other variables and the coefficients that represent significant correlations with a 95% confidence level (highlighted in bold). In this case, only the correlation between university (*UNIV*) and subjective perception of backtracking (*GSP_BK*) remain similar. However, some interesting new relations appear, such as those related to emotions:

- Students with high positive emotions (*PE_PRE*) and low negative emotions (*NE_PRE*) during the pretest, and higher grades in the backtracking knowledge posttest (*BK_POS*), have experienced lower negative emotions during the posttest (*NE_POS*).
- Positive emotions are negatively correlated with negative emotions, both in the pretest and in the posttest. This means that the higher the positive emotions, the lower the negative emotions, and vice versa.
- Emotions are positively correlated between phases, which implies that the higher the emotions in the pretest, the higher the emotions in the posttest, and vice versa.
- Positive emotions in the posttest (*PE_POS*) are positively correlated with VisBack valuation (*VB*), which seems to indicate that the visualization tool contributes to the well-being of students during backtracking learning.

Since the significant correlation coefficients are low again (almost all of them are less than 0.3), the study of the factors that could affect emotions was performed following the same procedures as in the original experience and all significant results were confirmed, except that no difference in negative emotions was found when the university is considered as a factor of change.

Table 9
Frequencies and percentages of the items of the reduced VB Scale.

	Likert scores				
	1	2	3	4	5
VB2. I think VisBack is useful to study backtracking	3 (3%)	2 (2%)	16 (17%)	26 (28%)	47 (50%)
VB3. I will use VisBack to study backtracking	4 (4%)	3 (3%)	29 (31%)	25 (27%)	33 (35%)
VB4. I will recommend VisBack to study backtracking	5 (5%)	6 (6%)	24 (25%)	27 (29%)	32 (34%)

Table 10
Tau-b coefficients for each pair of variables in the replicated experience.Those which are significant are marked in bold.

	BK_PRE	BK_POS	PSA	VB	PE_PRE	NE_PRE	PE_POS	NE_POS	UNIV
GSP_BK	**.176**	.117	-.098	-.050	.114	.009	.116	.021	**-.326**
BK_PRE	1.000	.058	-.062	-.013	.076	.075	-.004	.069	-.023
BK_POS		1.000	-.082	.099	**.166**	**-.240**	.113	**-.163**	.079
PSA			1.000	-.026	.048	.003	.039	-.049	.132
VB				1.000	.101	-.047	**.232**	-.094	**.256**
PE_PRE					1.000	**-.288**	**.288**	**-.280**	.088
NE_PRE						1.000	-.137	**.585**	-.053
PE_POS							1.000	**-.185**	.062
NE_POS								1.000	.057

Variations of students' emotions before and after using VisBack are similar to the original experience: positive emotions remain without significant changes ($p = 0.481 > 0.05$), which means that hypothesis **HP** can be rejected, while negative emotions in the posttest are significantly lower ($p = 0.028 < 0.05$) than those in the pre-test, which does not allow us rejecting hypothesis **HN**. In this case, the effect size for the change in negative emotions is small (0.178).

6. Results and discussion

This section describes and discusses the main findings, their implications, and threats to validity that might influence them.

6.1. Results

Let us remind that we proposed eight different (sub)hypotheses at the introduction section. Six of them were rejected at a 95% of confidence level by the results shown in the previous section. In particular, students' positive emotions were not increased as a consequence of using the visualization system (hypothesis HP). Furthermore, students' emotions were not influenced by their personal profile (HPP), their subjective perception of their knowledge on the topic (HSP), their previous knowledge of the subject (HPK), their subjective perception of their performance in the activity (HPA), or their valuation of the visualization system (HVB). However, one null hypothesis could not be rejected, namely that **the use of the visualization system decreases students' negative emotions** (HN).

We tried to create a regression model incorporating independent variables to measure how much variance of the dependent variable (*NE_POS*) can explained by each factor. However, no model was obtained, probably due to the low correlation between the corresponding variables.

Mixed results were obtained for the eighth hypothesis, which claimed that students' educational context affects their emotions (HEC). The first experiment yielded that the UCLM students experienced during the posttest phase more positive emotions and less negative emotions than URJC students. However, such a difference disappeared in the replication of the experiment. We guess that differences measured in the first evaluation may be due to the distinctive fact that UCLM students received a small reward. During the replication, students of both universities obtained a small reward for their participation and no differences in negative emotions were found when considering the university as a factor of change. This result suggests that emotional differences disappear if the external conditions at both universities are the same, i.e. hypothesis HEC can be rejected. However, it deserves further analysis.

Apart from the analysis of hypotheses, we gathered some data which deserve a commentary. Firstly, **positive emotions were experienced at a greater extent than negative emotions** during the pretest and the posttest. Secondly, the **visualization system was highly appreciated by students**. They considered VisBack useful and claimed that they would recommend its use. However, the students themselves were not as enthusiastic about using it in the future.

With the aim of measuring some aspects of the research, two scales (SSP and VB) were developed and calibrated. Consequently, reduced versions of both scales were developed (SSP* and VB*). Replication of the experience with a larger sample size led to the same reduction in the SSP Scale and to a very similar reduction in the VB scale. Although this is a side finding of our research, both calibrations of the SSP scale induce to think that **students' subjective perception of the foundations of recursion does not provide information to analyze more advanced topics involving recursion, such as backtracking**.

6.2. Discussion

Some of our findings are aligned with the results of other studies on algorithm visualization. For instance, VisBack high acceptance is not too surprising, as visualizations and animations typically are appealing to students (according to 'informal evaluations' reported by Urquiza-Fuentes & Velázquez-Iturbide, 2009). Less expected was the relative interest of students in using again VisBack for their studies. This could be interpreted as students' lack of initiative for self-study. However, the phenomenon of low adoption of visualization systems also has been observed in teachers (Ben-Bassat Levy & Ben-Ari, 2007).

A relevant finding of our study is that negative emotions are reduced (hypothesis HN) but positive emotions are not increased (rejection of hypothesis HP), which are consistent with findings regarding emotional design (Nurminen, 2016). This confirmation suggests that software visualization should not only be used for its cognitive effect but also for its emotional effects, especially to reduce negative emotions in first year students ;.

Some consequences of the validation process of the SSP questionnaire may be relevant for future studies. Such validation process confirmed the results obtained in previous experiences (Lacave, Molina, & Redondo, 2018), stating that students' subjective perception of the foundations of recursion (a generic topic) does not provide information to analyze more specific and complex topics involving recursion, such as backtracking (a complex topic). This may be because either these questions are too generic, or the students have a misperception of what they are actually being asked, or they do not answer with complete sincerity (King & Bruner, 2000). Therefore, it is not worth including questions on generalities or foundations of a topic in the future design of questionnaires on more complex and specific topics.

One result can be interesting from a methodological point of view. As noted above, the results obtained for the HEC hypothesis (students' educational context affects their emotions) are problematic. During the first experience, UCLM students experienced more positive emotions and less negative emotions than students from the URJC. We thought that it could be due to something in the context that may cause students from one university to experience more negative emotions because both universities have different environments, rules, policies, instructors, and student populations. However, the hypothesis contrasts made in sections 4.1.1 to 4.1.3 discarded that the context or the other controlled factors of our experience influenced emotions. For that reason, we wondered whether the difference might be due to the reward received by UCLM students. In fact, during the experience replication, students of both universities got a small reward for their participation, and in that case, there was no difference in their negative emotions considering university as a factor of change. This result suggests that rewards reduce negative emotions, but it should be further investigated. Some authors have claimed that rewards may cause a feeling of manipulation (Domínguez et al., 2013) or may reduce students' motivation (Kyewski & Krämer, 2018).

It is generally assumed that the larger the effect size, the smaller the sample size needed to detect the population occurrence of a phenomenon and, hence, to be generalized. In our case, the effect sizes obtained in the replica are smaller than in the original experience, despite the fact that the sample was somewhat larger. This is worth studying in the future, analyzing the student's emotions in more personal scenarios, even outside the classroom. Therefore, the results should be taken with caution, and our work should be considered as a pilot experience.

We obtained a few results which are in contradiction with other authors. In particular, we rejected the hypothesis that students' personal profile affects their emotions (HPP), contrary to previous studies (Rowe & Fitness, 2018). Moreover, positive emotions were experienced at a greater extent than negative emotions, contrary to previous findings (Atiq, 2018; Bosch, D'Mello, & Mills, 2013). Unfortunately, the activities used in the experiments of Bosch et al. (2013) did not involve interacting with program visualizations, which is supported by VisBack. Moreover, our students were in their second year of their undergraduate studies, rather than in their first year, as in Bosch et al.'s work. These issues require further inquiry.

6.3. Implications

Our findings open opportunities to new lines of research. Firstly, most research on algorithm visualization has been focused on students' engagement with visualizations (Naps et al., 2003) and, more generally, with the factors that influence students' learning performance when they use visualizations (Hundhausen et al., 2002; Urquiza-Fuentes & Velázquez-Iturbide, 2009). The graphical design of algorithm visualizations has traditionally been neglected, but it could be interesting to make a stronger connection between algorithm visualization and emotional design (Nurminen, 2016).

Secondly, alternative uses of visualization could be explored to increase positive emotions. Bosch et al. (Bosch et al., 2013) reported that positive emotions of flow/engagement were more frequent when construction tasks were undertaken. Therefore, using visualizations for constructive programming tasks (e.g. Velázquez-Iturbide & Pérez-Carrasco, 2016) has potential as an instructional alternative for increasing positive emotions.

Thirdly, emotions and motivation are linked (Pekrun and Perry, 2014). However, up to our knowledge, there are no studies analyzing the mutual relation of both factors in algorithm visualization. Note that motivation has hardly been studied in algorithm visualization (Velázquez-Iturbide et al., 2017), but there are no studies on emotions.

Finally, the overall analysis presented is informative and useful, but from an educational perspective, a fine-tuned understanding of emotions may be more impactful. Specific emotions should be measured, at least those most common emotions. This probably implies, for a future study, extending the range of tasks performed by students (Bosch & D'Mello, 2017; Bosch, D'Mello, & Mills, 2013) .

6.4. Threats to validity

The study presents several threats to validity (Shadish, Cook, & Campbell, 2002) that might have influenced our results.

- *Statistical Conclusion Validity.* We have tried to enhance our results by the proper application of the statistical tests. In fact, besides descriptive results, we have checked whether distributions adjusted to the normal to perform the subsequently parametric or non-parametric analysis. Moreover, we have provided participants with non-dichotomic variables to avoid the restriction of range and we have tried to identify different sources of emotions variability. However, some uncontrolled extraneous factors, such as personal feelings or circumstances, the day of the week or the time of day, the weather, etc. that could affect students' emotions have been beyond our reach.
- *Internal Validity.* Although quasi-experiments avoid most of the threats to internal validity that arise in other kind of experiments (Bärnighausen et al., 2017), our work has some limitations that must be considered. For example, the students might not have been properly motivated (Fittkau, Krause, & Hasselbrin, 2017) because they received a reward in the form of extra points which are not required to pass the subject and such reward was not the same for both set of students. Moreover, the time given to students to acquire fluency with the application and its learning capabilities may have been short. Finally, the theoretical concepts have been explained to students by different instructors and not at the same moment.
- *Construct Validity.* Our research questions may not provide complete coverage of emotions because we only measured positive vs. negative emotions, and not specific emotions, e.g. flow/engagement, confusion, frustration or boredom (Bosch et al., 2013, 2017). Moreover, the list of factors used to check alternative explanations was intended to be exhaustive, but additional factors could also be considered (e.g. visual style or general programming experience). Finally, our instrumentation was based on questionnaires, but emotions are a complex phenomenon; therefore, the use of physiological (Villanueva, Raikes, Ruben, Schaefer, & Günther, 2014) or even multi-modal methods (Atiq, 2018; Feidakis, Daradoumis, & Caballé, 2011) could give additional information.

- *External Validity*. Since in a quasi-experimental design the groups are pre-formed, the researcher does not have certainty that the sample is representative of the generality. Moreover, it does not represent the distribution of students by gender, although it is quite close, since the number of women is considerably lower than that of men in computing studies. Besides, data collected in a class may not be generalized to other educational scenarios, for instance studying at home, or also whether they study individually or in group. Moreover, although we have considered a generic research hypothesis, its empirical approach has forced us to focus on a concrete context and with a specific tool. In spite of the fact that by repeating the experience the main findings were confirmed, which gives them added value, the results are not generalizable for all university students. Therefore, the reproduction of the case study in other contexts and with other visualization tools remains open as an important line of future work. On the other hand, the effect size of the findings is small, which suggests that since the statistical procedures are adequate, a more detailed study is required.

7. Conclusions

The article has empirically addressed whether the use of an algorithm visualization tool may affect students' emotions while they learn programming. A quasi-experience was designed to analyze variations in positive and negative emotions experienced by undergraduate students while they learn backtracking with the assistance of the algorithm visualization system VisBack. We noted that the interpretation of quasi-experimental results should be considered with caution, as changes produced might be influenced by other objective and subjective factors. Consequently, we also considered several potentially relevant factors: the educational context, student's personal profile, student's previous knowledge, student's subjective perception of his/her own knowledge, student's assessment of the tool, and the cognitive load associated with the activity. The evaluation was conducted and later replicated to obtain higher evidence of the results.

The most relevant findings for programming education are that the use of the visualization system decreases students' negative emotions, and that students experienced more positive than negative emotions. Moreover, two scales were developed and calibrated (resulting in a reduced version for each one) to measure two factors (namely, students' subjective perception of their knowledge, and students' acceptance of the visualization system). Finally, the validation of the SSP Scale induces us to think that students' subjective perception of the foundations of recursion does not provide information to analyze more advanced forms of recursion, such as backtracking.

Some instructional implications can be straightforwardly deduced from the results of our study. Algorithm visualizations should be fostered in programming courses given that they are highly accepted tools, produce more positive than negative emotions in students, and reduce their negative emotions. These impacts can be especially valuable in the first courses of programming (Bosch et al., 2013; Bosch & D'Mello, 2017).

More specific recommendations deserve additional research. Past research on the impact of algorithm visualization on students' performance noted that positive effects only were achieved under certain conditions, because students' instructional use of visualizations was identified to be more relevant than the contents of the visualizations (Hundhausen, Douglas & Stasko, 2002). Similarly, we cannot claim that program visualizations always have a positive impact on students' emotions. Future research should hopefully shed light on the impact of visualizations on students' emotions in different educational settings.

Credit author statement

Carmen Lacave: Conceptualization, Investigation, Formal analysis, Validation, Writing - Original Draft, **J. Angel Velázquez-Iturbide:** Investigation, Writing - Review & Editing, Funding acquisition, **Maximiliano Paredes-Velasco:** Conceptualization, Methodology, Resources, Validation, **Ana I. Molina:** Methodology, Formal analysis, Validation.

Acknowledgments

This work was supported by research grants iProg of MINECO (TIN2015-66731-C2-1-R and TIN2015-66731-C2-2-R) and eMadrid of CAM (P2018/TCS-4307). We thank the participants of this study and the anonymous reviewers for their useful suggestions.

APPENDIX.

Table 11
First activity performed by students with the VisBack tool

ACTIVITY 1
In how many ways can Paco, José and Ana sit in a row of three seats? Use the *permutations without repetition* algorithm provided in the VisBack system to generate a search tree with the initials of their names as input data (P, J and A). First, perform these activities: 1 Enter input data and run an execution of the algorithm. Analyze the generated tree. 2 Configure the type of the tree nodes to advanced mode. 3 Run the algorithm step by step up to node 'P, A', focusing on the values of the variables shown in each node. 4 Run the algorithm step by step up to node 'A'. Think why the value of variable *k* is 2 and the value of *step* is 0. 5 Execute step by step until the end; think why node 'A,P' is generated before node 'A,J'.

Now answer to the following questions. Run the algorithm backward or step forward, as needed:

a How many solutions are generated?
b In how many steps do you get the solution(s)?
c How many nodes are generated before obtaining the first solution?
d How many nodes are generated to obtain all the solutions?
e What does each stage represent?
f How do you detect if a node is a solution?
g The descendants of a node are generated by adding ….
h What is the value of variable *stage* when node 'P, A' is created?
i What is the value of variable *k* at node 'J, A, P'?
j Is the 'JA' node a descendant of the 'J' node or of the 'A' node? Why?
k Why the 'P,P' node isn't created?
lDraw the tree generated with input data [P, J, A, R]. You don't have to draw it in one piece.

Table 12
. Pretest of students' level of knowledge of backtracking

Thank you for participating in this experience, which seeks to know your knowledge of backtracking.
Remind that we are not evaluating you for the course grade, and that the information collected in this test will be treated confidentially and used exclusively for this study.

PRETEST OF BACKTRACKING KNOWLEDGE

A seller wants to visit N cities without going twice through any city, so he/she wants to know how many valid routes exist when he/she starts from a particular city.

1 Which algorithm would you use to solve the problem?

a.Combinations without repetition:

```
static void CSRmn(String[] current, String[] original, int step, int from) {
  if (step == current.length) {
   for (int i = 0; i < current.length; i++)
    System.out.print(current[i]+" ");
   System.out.println();
  }
  else{
   for(int k=from;k<original.length;k++){
     current[step]=original[k];
     CSRmn(current,original,step+1,k+1);
   }
  }
}
```

b.Permutations without repetition:

```
static boolean isValid(String[] current, int step, String new){
  boolean valid = true;
  for (int n = 0; n < step && valid; n++)
   if (current[n].equals(new))
    valid = false;
  return valid;
}

static void PSRmn(String[] current, String[] original,int step) {
  if (step == current.length) {
   for (int i = 0; i < current.length; i++)
    System.out.print(current[i]+" ");
   System.out.println();
  }
  else{
   for(int k=0;k<original.length;k++){
    if(isValid(current,step,original[k])){
      current[step]=original[k];
      PSRmn(current,original,step+1);
    }
   }
  }
}
```

c.Variations with repetition:

```
static void VCRmn(String[] current, String[] original, int step) {
  if (step == current.length) {
   for (int i = 0; i < current.length; i++)
    System.out.print(actual[i]+" ");
   System.out.println();
  }
  else{
   for(int k=0;k<original.length;k++){
     current[step]=original[k];
     VCRmn(current,original,step+1);
   }
  }
}
```

d.Any of the three algorithms **can solve the problem**

Given the algorithm you selected and the following Java statements: String[ ] original = {‘R’, ‘M’, ‘C’}; step = 0; from = 0; **draw on a piece of paper the search tree generated when the algorithm selected (PSRmn, CSRmn or VCRmn) is run, and answer the following questions:**	
2 How many solutions can be generated?	a.3
	b.6
	c. 9
	d.12
3 In how many steps the solution(s) is (are) obtained, provided the external call is the first step?	a.0
	b.1
	c.3
	d.4
4 How many nodes must be generated before obtaining the first solution (including the solution node)?	a.0
	b.4
	c. 6
	d. 16
5 How many nodes must be generated to obtain all the solutions?	a.1
	b.4
	c.6
	d.16
6 Which is the first solution generated?	a.‘C,R,M’
	b.‘M,R,C’
	c.‘R,M,C’
	d. ‘C,M,R’
7 At depth level 1 of the tree (when variable *step* = 0) the following three nodes are generated in this order:	a.‘C’, ‘M’ y ‘R’
	b.‘R’, ‘M’ y ‘C’
	c.‘M’, ‘R’ y ‘C’
	d.‘R’, ‘C’ y ‘M’
8 The node ‘M,R’ is generated as a completion of the node:	a.‘R’
	b.‘M’
	c.‘C’
	d. That node is not generated
9 When variable *step* = 2, the node is:	a.a leaf
	b.the parent of a leaf node
	c.a child of the root
	d.the tree root
10 What does each stage represent?	a.The addition of a valid city to the solution
	b.The addition of any city to the solution
	c.A three-city solution
	d. That no more completions can be generated
11 How are the descendants of a node generated?	a.By adding a city
	b.By adding a city not previously included in the solution
	c.By adding all cities
	d.By adding all cities not previously included in the solution

Table 13
The SSP scale to measure students' subjective perception of backtracking

Recursion
R1. I know the theoretical concepts underlying recursion
R2. I know how recursion works
R3. I understand how recursion works
Backtracking
B1. I know the theoretical concepts underlying general trees
B2. I know the depth traversal algorithms for general trees
B3. I understand the depth traversal algorithms for general trees
B4. I know the theoretical concepts underlying backtracking algorithms
B5. I know how backtracking algorithms work
B6. I know the code of backtracking algorithms
B7. I understand how a backtracking search tree is generated

Table 14
The PANAS emotional scale

Positive Emotions	Negative Emotions
Attentive	Hostile
Active	Irritable
Alert	Ashamed
Excited	Guilty
Enthusiastic	Distressed
Determined	Upset
Inspired	Scared
Proud	Afraid
Interested	Anxious
Strong	Nervous

Table 15
The VB scale to evaluate the VisBack tool

VisBack
VB1. I think VisBack is easy to use
VB2. VisBack helps me in understanding backtracking
VB3. I think VisBack is useful to study backtracking
VB4. I will use VisBack to study backtracking
VB5. I will recommend VisBack to study backtracking

Table 16
The PSA survey to measure students' subjective performance of the experience

PSA
PSA1. I think the activity was difficult
PSA2. I was very concentrated during the activity
PSA3. I had to make a big effort to perform the experience
PSA4. The realization of the activity was difficult for me
PSA5. I think I did the activity well

Table 17
Homogeneity index and Cronbach alpha value of each item of the SSP Scale if deleted (Those higher than the previous Cronbach's Alpha are marked in bold)

Item	Homogeneity Index	Cronbach's Alpha if Item Deleted
R1	.386	**.881**
R2	.398	**.880**
R3	.409	**.880**
B1	.657	.864
B2	.769	.854
B3	.770	.853
B4	.618	.866
B5	.717	.859
B6	.628	.866
B7	.648	.864

Table 18
Homogeneity index and Cronbach alpha value of each item of the SSP Scale (Those higher than the previous Cronbach's Alpha are marked in bold).

	Original SSP scale		After removing items R2 and R3		After removing item R1	
Cronbach's alpha	*0.833*		*0.860*		*0.875*	
Item	Hom. index	C-α if deleted	Hom. index	C-α if deleted	Hom. index	C-α if deleted
R1	.426	.827	**.277**	**.875**		
R2	**.272**	**.838**				

R3	**.212**	**.844**				
B1	.662	.804	.612	.843	.571	.867
B2	.655	.804	.661	.837	.636	.860
B3	.706	.799	.683	.835	.663	.856
B4	.606	.809	.680	.835	.695	.852
B5	.575	.813	.672	.836	.704	.851
B6	.514	.819	.597	.845	.628	.861
B7	.569	.814	.663	.837	.696	.852

Table 19
Factor matrix of the reduced SSP Scale after rotation (The highest values for each factor are marked in bold).

Item	F1	F2
B1	.013	**.830**
B2	-.026	**.947**
B3	.025	**.915**
B4	**.889**	-.005
B5	**.965**	-.087
B6	**.815**	.001
B7	**.763**	.136

References

Adnan, H., & Redzuan, F. (2017). Evaluating students' emotional response in video-based learning using Kansei Engineering. *Proceedings of the 4th International Conference on User Science and Engineering*, 237–242.

Atiq, Z. (2018). Emotions experienced by first-year engineering students during programming tasks. Proceedings of the 2018 ACM conference on international computing education research: Vol. 2018 (pp. 258–259). ICER.

Bärnighausen, T., Tugwell, P., Røttingen, J. A., Shemilt, I., Rockers, P., & Geldsetzer, P., et al. (2017). Quasi-experimental study designs series—paper 4: Uses and value. *Journal of Clinical Epidemiology*, 89, 21–29.

Ben-Bassat Levy, R., & Ben-Ari, M. (2007). We work so hard and they don't use it: Acceptance of software tools by teachers. *ACM SIGCSE Bulletin*, 39(3), 246–250.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3–62.

Bosch, N., & D'Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, 27(1), 181–206.

Bosch, N., D'Mello, S., & Mills, C. (2013). What emotions do novices experience during their first computer programming learning session? In Lane, H. C., Yacef, K., Mostow, J., & Pavlik, P. (Eds.), *Artificial intelligence in education* (pp. 11–20). Berlin, Heidelberg: Springer.

Bowman, N. D. (2017). The importance of effect size reporting in communication research reports. *Communication Research Reports*, 34(3), 187–190.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2. Auflage). Hillsdale, NJ: Erlbaum.

Cook, T., & Campbell, D. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141–180.

Cox, A., & Fisher, M. (2009). Programming style: Influences, factors, and elements. *2009 second international conferences on advances in computer-human interactions* (pp. 82–89).

Davis, F. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), 475–487.

Decker, A., & Simkins, D. (2016). Uncovering difficulties in learning for the intermediate programmer. *Proceedings of the 2016 IEEE frontiers in education conference (FIE), Erie, PA, USA* (pp. 1–8).

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63, 380–392.

Fanselow, M. S. (2018). Emotion, motivation and function. *Current Opinion in Behavioral Sciences*, 19, 105–109.

Feidakis, M., Daradoumis, T., & Caballé, S. (2011). Emotion measurement in intelligent tutoring systems: What, when and how to measure. *2011 third IEEE international conference on intelligent networking and collaborative systems* (pp. 807–8011).

Finch, D., Peacock, M., Lazdowski, D., & Hwang, M. (2015). Managing emotions: A case study exploring the relationship between experiential learning, emotions, and student performance. *International Journal of Management in Education*, 13, 23–36.

Fittkau, F., Krause, A., & Hasselbrin, W. (2017). Software landscape and application visualization for system comprehension with ExplorViz. *Information and Software Technology*, 87, 259–277.

Frenzel, A. C., Goetz, T., Ludtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, 101(3), 705–716.

Frenzel, A. C., Taxer, J. L., Schwab, C., & Kuhbandner, C. (2018). Independent and joint effects of teacher enthusiasm and motivation on student motivation and experiences: A field experiment. *Motivation and Emotion*, 1(3), 3.

Golomb, S. W., & Baumert, L. D. (1965). Backtrack programming. *Journal of the ACM*, 12, 516–524.

Haaranen, L., Ihantola, P., Sorva, J., & Vihavainen, A. (2015). In search of the emotional design effect in programming. Proceedings of the 37th international conference on software engineering -: Vol. 2 (pp. 428–434). Piscataway: IEEE Press.

Hattie, J. (2009). Visible learning. London: Routledge.

Hidalgo-Céspedes, J., Marín-Raventós, G., & Lara-Villagrán, V. (2016). Learning principles in program visualizations: A systematic literature review. *Proceedings of the 2016 IEEE frontiers in education conference (FIE)*. doi:10.1109/FIE.2016.7757692.

Horowitz, E., & Sahni, S. (1974). Computing partitions with applications to the knapsack problem. *Journal of the ACM*, 21(2), 277–292.

Hundhausen, C., Dougkas, S., & Stasko, I. (2002). A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, 13, 259–290.

Husman, J., Cheng, K. C., Puruhito, K., & Fishman, E. J. (2015). Understanding engineering students stress and emotions during an introductory engineering course. *Proceedings of the 2015 ASEE annual conference and exposition. Seattle, Washington.*

Ishkov, A., & Magera, T. (2015). Emotional competency in an engineering university: Yes or no? *Procedia Engineering*, 117, 148–153.

Iskrenovic-Momcilovic, O. (2018). Learning a programming language. *International Journal of Electrical Engineering Education*, 55(4), 324–333.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151.

Karimia, Z., Baraani-Dastjerdia, A., Ghasem-Aghaeea, N., & Wagner, S. (2016). Links between the personalities, styles and performance in computer programming. *Journal of Systems and Software*, 111, 228–241.

King, M., & Bruner, G. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing*, 17(2), 79–103.

Kinnunen, P., & Simon, B. (2012). My program is ok – am I? Computing freshmen's experiences of doing programming assignments. *Computer Science Education*, 22(1), 1–28.

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5, 345–379.

Kyewski, E., & Krämer, N. C. (2018). To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education*, 118, 25–37.

Lacave, C., Molina, A. I., & Redondo, M. A. (2018). A preliminary instrument for measuring students' subjective perceptions of difficulties in learning recursion. *IEEE Transactions on Education*, 61(2), 119–126. doi:10.1109/TE.2017.2758346.

Lenhard, W., & Lenhard, A. (2016). Computation of effect sizes. Available at: https://www.psychometrica.de/effect_size.html.%20Bibergau ((Germany): Psychometrica).

Li, X. (2017). The influence of human factors on programming performance: Personality, programming styles and programming attitudes PhD Thesis. Alabama, USA: Auburn University.

Liaw, S. (2008). Investigating students' perceived satisfaction, behavioral intention and effectiveness of e-learning: A case study of the blackboard system. *Computers & Education*, 51(2), 864–873.

Liaw, S. S., Huang, H. M., & Chen, G. D. (2007). An activity-theoretical approach to investigate learners' factors toward e-learning systems. *Computers in Human Behavior*, 23(4), 1906–1920.

Lishinski, A., Yadav, A., & Enbody, R. (2017). Students' emotional reactions to programming projects in introduction to programming: Measurement approach and influence on learning outcomes. *Proceedings of the 2017 ACM conference on international computing education research, ICER'2017* (pp. 30–38).

Mainhard, T., Oudman, S., Hornstra, L., Bosker, R. J., & Goetz, T. (2018). Student emotions in class: The relative importance of teachers and their interpersonal relations with students. *Learning and Instruction*, 53, 109–119.

Martin, C., Hughes, J., & Richards, J. (2017). Learning experiences in programming: The motivating effect of a physical interface. *Proceedings of the 9th international conference on computer supported education, SCITEPRESS* (pp. 162–172).

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, 33, 12–18. doi:10.1016/j.learninstruc.2014.02.004.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, 67(3), 267–277.

McDonald, P., & Ciesielski, V. (2002). Design and Evaluation of an Algorithm Animation of State Space Search Methods. *Computer Science Education*, 12(4), 301–324. doi:10.1076/csed.12.4.301.8622.

Merino, L., Ghafari, M., Anslow, C., & Nierstrasz, O. (2018). A systematic literature review of software visualization evaluation. *Journal of Systems and Software*, 144, 165–180.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386.

Muñoz-Merino, P. J., Fernández, M., Muñoz-Organero, M., & Delgado, C. (2014). Motivation and emotions in competition systems for education: An empirical study. *IEEE Transactions on Education*, 57, 182–187.

Nadelson, L., Hardy, K. K., & Yang, D. (2015). I like therefore I learn! Engineering student motivation to learn in their least and most favorite courses. ASEEE annual conference & exposition, Seattle, WA, USA: Vol. 26. 870.1-26,870.12.

Naps, T., Roessling, G., Almstrum, V., Dann, W., Fleischer, R., & Hundhausen, C., et al. (2003). Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bulletin*, 35(2), 131–152.

Nolan, K., & Bergin, S. (2016). The role of anxiety when learning to program: A systematic review of the literature. *Proceedings of the 2016 Koli calling conference* (pp. 61–70).

Norman, D. (2005). Emotional design: Why we love (or hate) everyday things. New York: Basic Books.

Nurminen, M. I. (2016). Effects of emotional design and goal orientation on students' emotions and learning outcomes in university programming education Master Thesis. Finland: Tampere University of Technology.

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education*, 86, 30–42.

Pekrun, R. (2014). Emotions and learning. Educational practices series-24. Geneva, Switzerland: UNESCO International Bureau of Education.

Pekrun, R., Elliot, A., & Maier, M. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychologist*, 98(3), 583–597.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In Pekrun, R., & Linnenbrink-Garcia, L. (Eds.), *International handbook of emotions in education* (pp. 120–141). New York: Routledge.

Pérez-Mena, J. (2015). VisBack: Herramienta para la visualización de algoritmos de backtracking Master Thesis. Spain: University of Castilla-La Mancha.

Pirkkalainen, H., Pawlowski, J. M., & Pappa, D. (2017). Educators' open educational collaboration online: The dilemma of emotional ownership. *Computers & Education*, 106, 119–136.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, 29, 128–140.

Plass, J., Moreno, R., & Brünken, R. (2010). Cognitive load theory. Cambridge University Press.

Rowe, A., & Fitness, J. (2018). Understanding the role of negative emotions in adult learning and achievement: A social functional perspective. *Behavioral Sciences*, 8(2), 27.

Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernández-Castro, I., & Duval, E. (2016). Supporting learning by considering emotions: Tracking and visualisation. A case study. Proceedings of the sixth international conference on learning analytics & knowledge - LAK: '16 (pp. 254–263).

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Shull, F., Singer, J., & Sjøberg, D. I. (Eds.). (2007). *Guide to advanced empirical software engineering*. Springer Science & Business Media.

Sorva, J., Karavirta, V., & Malmi, L. (2013). A review of generic program visualization systems for introductory programming education. *ACM Transactions on Computing Education*, 13(4). article 15.

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers and Education*, 120, 185–196.

Tomczak, M. T., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *TRENDS in Sport Sciences*, 1(21), 19–25.

Urquiza-Fuentes, J., & Velázquez-Iturbide, J. (2009). A survey of successful evaluations of program visualization and algorithm animation systems. *ACM Transactions on Computing Education*, 9(2), 9.

Uzun, A. M., & Yıldırım, Z. (2018). Exploring the effect of using different levels of emotional design features in multimedia science learning. *Computers & Education*, 119, 112–128.

Velázquez-Iturbide, J. (2019). Students' misconceptions of optimization algorithms. *Proceedings of the 24th annual conference on innovation and technology in computer science education, ITiCSE'19* (pp. 464–470).

Velázquez-Iturbide, J., Hernán-Losada, I., & Paredes-Velasco, M. (2017). Evaluating the effect of program visualization on student motivation. *IEEE Transactions on Education*, 60(3), 238–245.

Velázquez-Iturbide, J., & Pérez-Carrasco, A. (2016). Systematic development of dynamic programming algorithms assisted by interactive visualization. *Proceedings of the 21st annual conference on innovation and technology in computer science education. ITiCSE'16* (pp. 71–76).

Velázquez-Iturbide, J., Pérez-Carrasco, A., & Urquiza-Fuentes, J. (2008). SRec: An animation system of recursion for algorithm courses. *Proceedings of the 13th annual conference on innovation and technology in computer science education, ITiCSE* (Vol. 2008, pp. 225–229).
Verdegem, P., & De Marez, L. (2011). Rethinking determinants of ICT acceptance: Towards an integrated and comprehensive overview. *Technovation*, 31(8), 411–423.
Villanueva, I., Raikes, A., Ruben, N., Schaefer, S., & Günther, J. (2014). The use of physiological tools to identify changes in affective responses for graduate students recently admitted into a scientific discipline. *Proceedings of the 2014 IEEE frontiers in education conference* (pp. 1–5). FIE).
Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.
Zamanzadeh, V., Rassoul, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, 1(3), 163–171.
Zhu, H., Zhang, X., Wang, X., Chen, Y., & Zeng, B. (2014). A case study of learning action and emotion from a perspective of learning analytics. *IEEE 17th International Conference on Computational Science and Engineering*, 420–424.