

Case study on the assessment of the affective dimension in virtual learning environments

Eva de Lera
Universitat Oberta de Catalunya
Avda. Tibidabo, 47
08035 Barcelona (Spain)
(+34) 932535700
edelera@uoc.edu

Carles Fernàndez
Universitat Oberta de Catalunya
Avda. Tibidabo, 47
08035 Barcelona (Spain)
(+34) 932535700
cfernandezba@uoc.edu

Magí Almirall
Universitat Oberta de Catalunya
Avda. Tibidabo, 47
08035 Barcelona (Spain)
(+34) 932535700
malmirall@uoc.edu

ABSTRACT

This paper presents a proposal for fully online or blended learning universities to assess the affective dimension in online learning design.

The paper focuses especially on affective assessment by presenting a study of emotional inference through the triangulation of facial expression interpretations, pupil size and students as self-evaluators.

Keywords

E-learning, affective computing, affective learning, pupil size, FACS, affective assessment

1. INTRODUCTION

Most of the universities that are currently offering online courses usually have some kind of system for course design assessment. Traditionally these evaluation tools have been based on a set of recognized indicators provided by experts in graphic design, instructional designers, technologists... Up to now, the evaluation of affect in online learning usually means the use of questionnaires where students respond to items like *'do you think the course is motivating enough?'* or *'has the course met your expectations?'*. Although questionnaires are a short and cost-effective way to gather data, we already know that users may answer just to offer an 'appropriate' opinion or are mediated by other variables. As Picard states, "self-report is colored by awareness of internal state, reflections on how such a report will be perceived, ability to articulate what one feels, etc." [1]. Also, the use of questionnaires requires the learning experience to be interrupted (or in any case, they cannot be used at the same time).

In fact, designers tend to implicitly propose a particular emotional scenario when they design virtual learning environments, but these decisions are generally not based on affective learning theories or findings.

In conclusion, at the present time the Universitat Oberta aims to introduce a new methodology that provide us with reliable data on the affective states of students in order to design engaging educational environments based on research outcomes.

2. INTRODUCING AFFECT IN ASSESSMENT

2.1 A reliable, feasible and cost-effective methodology

A method that requires adding more effort to any current assessment system needs to be reliable, feasible and cost- or time-effective. Affective evaluation must be integrated within the existing evaluation system and should not imply a considerable consumption of time so that institutions do not show resistance to it.

Our proposal consists of the use of three techniques: pupil size analysis, interpretation of facial and body expressions and students' self-assessment. All the data gathered through these techniques is compared to the real-time interaction of students.

In order to validate the combination (henceforth triangulation) of these techniques as a specific methodology, we planned a test with seven students interacting with the new mainpage of the virtual learning environment. Until now, after learners logged into the virtual campus they just found a mainpage with some news. The new mainpage is a personalized space based on widgets where students can choose the kind of information they would like to have when accessing the campus. It is very similar to the recognized and awarded *Igoogle* personal area and students are allowed to import external modules (gmail, google calendar, news and many other applications) as well as information from the university (forums, groups, news...), and personalize the design and position of these modules.

The sample was selected following heterogenic criteria: age, gender, studies and ICT skills were representative according to our target students at the university.

Although one of the objectives is to determine how affective the mainpage is, the essential focus of this research is to analyze the possibilities for inferring emotional data from pupil size, facial expressions and student self-assessment as a triangulation method.

2.2 Brief state of the art of the techniques used

The three techniques we have used have been tested as shown in the state of the art.

Pupil size has been widely studied by Picard and Hess (who coined the term in 1975), and accurately summarized by Timo

Partala [2] The latest research on pupilometry has confirmed that the increase of pupil size has a linear relationship with affective arousal. Arousal is significantly increased except with neutral stimuli. Nevertheless, other factors such as cognitive load and light reflex may have an effect and must be appropriately corrected in any research study.

One of the main advantages of this technique is that it does not require special sensors and measures involuntary responses from individuals, so that they cannot control their reactions.

Secondly, while pupilometry relies on arousal, FACS (Facial Action Code System) and other methods to analyze non-verbal information are related to valence. To date, there is some consensus on the existence of six basic emotions that can be codified according to FACS (anger, disgust, fear, joy, sorrow and surprise). The technique created by Ekman codes expressions as a combination of 44 facial movements called Action Units. [3]

Finally, concerning self-assessment; this technique has been partially criticized due to the fact that adults' self-report is colored by awareness of internal state, reflections on how such a report will be perceived, ability to articulate what one feels, and other factors [1] On the other hand, using self-assessment as a complementary method can bring interesting results. As an example of self-assessment interest, Sang-Hoon Jeong has developed a tool called VideoTAME [4] where subjects can view the recorded video of them performing tasks. This data can be useful as complementary information for emotion inference or confirmation of affective states.

2.3 The methodology in the test design

Students have to carry out five tasks on the mainpage: read an email in their webmail, add a new module (widget), change the position of another module, access the forum of a subject and access the virtual library. All these tasks can be done directly from the new mainpage.

More important than the tasks themselves are the 'key events' that students need to perform in order to accomplish the task. For example, a key event in the task 'add a new module' is to find the right button that allows widgets to be added. Another key event for any of the tasks would be to know how to go back to the mainpage after the student uses webmail or the forums.

The key events are new stimuli within a 'steady/stable interaction', and represent opportunities for increase of arousal, especially when individuals cannot accomplish the event easily.

We have identified the following key events:

- To find the key buttons such as: 'add a module', 'personalize your mainpage', 'virtual library', 'webmail' and 'forums' (a total of 5 events)
- To be able to go back to the main page after a task
- To understand the concept of module as a box that can be dragged and dropped

Apart from key events, we also identified a group of events that require special attention by the user such as typing the password, waiting for the pages to load, reading instructions, accessing wrong pages... We call them 'secondary events'.

As stated before, our methodology is based on a combination of pupilometry, gesture expression heuristics and student self-assessment.

Each technique has its own objective. We used pupilometry to detect arousal increase, whether related to positive or negative affective states. Increased pupil size seems to indicate that 'something is happening internally' and may be a clear sign that impels us to also analyze the valence of the affective state through facial and gestural expressions. Moreover, triangulating pupil size with gestural heuristics and self-assessment should allow us to assess how precise pupil size is by comparing the average of increased pupil size.

In this sense, it is very interesting to analyze the correlation between increased pupil size and a body expression that confirms this physical response in arousal. We consider that pupil size is increased when size is 10% bigger than the average pupil size in the test for a particular subject.

Secondly, our gesture coding system is used to determine the valence of emotions, since pupilometry itself cannot do anything with valence. Our own technique, called *ten heuristics*, is based on observation and does not require extra implementation effort since most interface evaluations are conducted observing and recording the user as he or she interacts with the interface. In such a scenario, facial and body expressions are often observed and recorded, but generally not measured in a structured manner. The ten heuristic tool has been developed and adapted thanks to previous studies in which we tried to identify the most common expressions of students in virtual learning environments.

Finally, as for the third technique, the objective of self-assessment in our methodology is to confirm that both pupilometry and gestural interpretation can really infer affective states.

The triangulation between these three techniques has two objectives:

- To analyze the correlations between the data gathered with the techniques in order to assess the specific effectivity of each one
- To conclude whether the overall methodology is appropriate for affect measurement

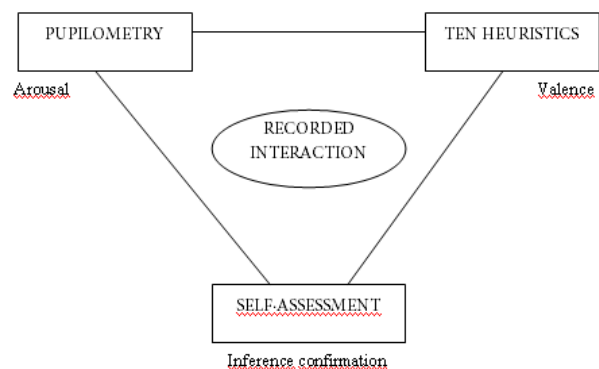


Figure 1. The methodology based on triangulation

2.4 Results and conclusions

As we stated before, due to the fact that we only tested a group of seven people, we cannot generalize conclusions. Nevertheless, there are a few results for reflection on the techniques we used:

Both pupilometry and 'ten heuristics' together were able to measure most of the key events and secondary events. We have analyzed the events where a key event could not be detected, the reason being that the student solved it without any problem, which means no arousal. The higher the level of success in a task, the less probable it is for arousal or non-verbal cues to appear.

These techniques have also showed that pupil size was increased or students reacted non-verbally to some events we had not considered as key or secondary.

We have also compared whether pupil size or ten heuristics were more reactive to key and secondary events. Pupil size responded to 74% of these events while ten heuristics did so in 65%.

One of the most important questions for our proposed methodology is the correlation between positive responses in pupil size and positive responses in heuristics. For most of the students (five) only 40% of the events with increased pupil size also showed any gestural reaction. The two others showed over 80% of correlation. Our conclusion is that pupil size does not seem to be a very good predictor if we only want to use it to infer or to state that the student is 'feeling something'. The size of pupils does not seem to be enough to assess arousal in our particular environment or conditions. Also, almost 40% of the events with increased pupil size did not show any particular facial or body expression.

Very interestingly, the correlation between high pupil size and specific gesture (average of 47.8%) is higher than the correlation between normal pupil size-specific gesture (20.7%) and high pupil size-no gestures (31.1%). This data may show that affective states are more commonly accompanied by the presence of both high pupil size and non-verbal expression, though evidence is not great enough to focus on pupilometry alone.

The analysis of gestural expressions by itself is richer and allows a good determination of valence. Most of the key events showed a particular expression, and the emotions inferred through these were confirmed through students' self-assessment.

Through students' self-assessment we have particularly checked whether the expressive reactions in Ten Heuristics can be inferred as specific emotions. Some of the heuristics were frequently repeated in the pilot test, such as frowning, compressing the lips, expressing vocally or hands touching the face.

The amount of non-verbal expression depends on each individual, and is probably the most important factor for heuristics performance. There was a relationship between success in tasks and amount of expressions (more difficulties, more presence of body expressions), but there was an individual who had important problems in some of the tasks but did not show many expressions. In her self-assessment she let us know she is not very expressive.

Globally, students' self-assessment confirmed more than 96% of the events detected by both pupilometry and heuristics. Almost 100% of the heuristics were confirmed and more than 90% of the events detected through increased pupil size.

Through the qualitative analysis of each specific case study we have identified other variables that have a particular impact on affect and the techniques we have selected to measure:

1. Pupil size tends to be higher during approximately the first minute due to some kind of tension
2. Sustained concentration makes pupil size increase but is not comparable to pupil size increase in presentation of new interfaces, errors produced or difficulties in performance.
3. Concerning patterns of timeline analysis, pupil size is reactive usually 500 ms. before the key event and it keeps increasing depending on how well the user finds a solution but generally maintains the maximum size during 2 seconds.
4. Pupil size is not especially high when users speak, probably because talking is a way to decrease arousal.
5. We have analyzed the highest moments of increased pupil size but could not find specific body expressions related to high arousal.
6. The two most successful users did not show high increases in pupil size although they also had to solve new problems. Overall confidence may be the explanation.

In conclusion, the methodology of triangulation we have proposed for data collection and affect inference is useful, but we cannot rely on one technique alone, especially on pupilometry, since it showed that some key events and some difficulties experienced by users did not increase pupil size (less than half of events). Nevertheless, it is convenient to collect pupil size data since it can make us focus on events where subjects did not express anything.

Concerning other considerations of implementation such as feasibility, usability or time or cost-effectiveness, we can argue that these techniques do not require a big investment and the way they gather and export data is very useful for research. Pupilometry only requires an eyetracker with infrared rays, facial expression monitoring is carried out with a simple webcam, and for students' self-assessment we only need learners. These technologies are quite affordable.

With the appropriate guidelines for integrating this methodology in course assessment, staff should not notice a significant increase in work, or they should at least feel a balance between time investment and higher quality in course assessment and design.

3. REFERENCES

- [1] Picard RW. et al., 2004. Affective learning – a manifesto. *BT Technology Journal* (22), 255
- [2] Partala, T. and Surakka, V., 2002. Pupil size variation as an indicator of affective processing. *International Journal of Human-Computer Studies*, 59 (1-2), 185-198
- [3] Michel, P., and El Kaliouby, R., 2003. Real time facial expression recognition in video using support vector machines. In proceedings ICMI International Conference on Multimodal Interfaces 2003 (Vancouver, Canada, 2003) ACM Press, New York, NY 258-263. DOI=<http://doi.acm.org/10.1145/958432.958479>
- [4] Sang-Hoon, J., 2007. Suggestion of Methods for Understanding User's Emotional Changes While Using a Product. In *Human Interface*, M.J. Smith, G. Salvendy, Eds. HCII 2007, Springer-Verlag Berlin Heidelberg, 59-67