# Evaluating Software for Affective Education: A Case Study of Affective Heuristics

Reza GhasemAghaei, Ali Arya, Robert Biddle
School of Computer Science
Carleton University
Canada
Reza.GhasemAghaei@carleton.ca

**Abstract**: This paper addresses the use of a proposed heuristic method to evaluate affect in the user interface, educational design, and content of educational multimodal software system. A case study is presented where this method is used to evaluate a Museum Explorer system. This system was designed to support an educational objective of making the users more engaged and reflective, by supporting the use narrative to help the users tell a visual story. We recruited participants to apply the evaluation method to the Museum Explorer system. Our findings were that the evaluation method was effective, but we observed a number of effects that suggested necessary improvements. For example, the emphasis in the method had to be changed to avoid a focus on simple usability issues. Moreover, we found that the evaluators had to be clear on the affective teaching objectives and on how the software design was intended to support them. We conclude with a revised set of heuristics based on these findings.

## Introduction

Interactive narratives are a form of digital entertainment that allow users of a system to interact and change stories according to their own desires and participate in a collaborative experience (de Lima et al., 2011). In a multimodal learning environment, educational elements can be presented in more than one sensory mode, such as audio, music, text, picture, etc. to overcome the limitations of each modality alone. The combination of narrative, audio, picture and text supports proficient communicative behavior and leverages natural human capabilities to communicate and bring pattern recognition and classification methods (Turk, 2014).

Many multimodal learning environments have been developed, but there has been little work on evaluating their design (Kühnel, 2012). We have proposed a principled model for *affective* multimodal educational software (GhasemAghaei et al., 2015), and suggested some specific techniques for evaluating affect in software. In this paper we examine one of our evaluation techniques with a case study of a system for exploring a museum.

The rest of this paper is structured as follows. We first introduce our framework, and then the software that is the basis of our case study. We then explain our evaluation technique, describe our method, and present our findings.

## Overview of the MADE Framework

Our framework is based on *Bloom's taxonomy* (Bloom et al., 1956). We adapted the three domains (cognitive, affective and psychomotor) and we considered the *multiple sensory* and *quasi-sensory modalities*, and our particular focus is on the affective domain (GhasemAghaei et al., 2015). For the affective side we applied Kort et al.'s affective model (Kort et al., 2001). This highlights the relationships between emotion and learning, and was built on Russell's circumplex model. Our model focuses on what emotional state the learner is in, and from that, what help he/she needs, and attempts to link the emotion and the cognitive aspects of the learning process.

Figure 1 illustrates our MADE *(Multimodal Affect for Design and Evaluation)* Framework. In this framework the instructor's learning objective is shown as involving cognitive and affective aspects, and leads to linkages with sensory and quasi-sensory modalities. By quasi-sensory modalities we refer to concepts such as narrative and persuasion.

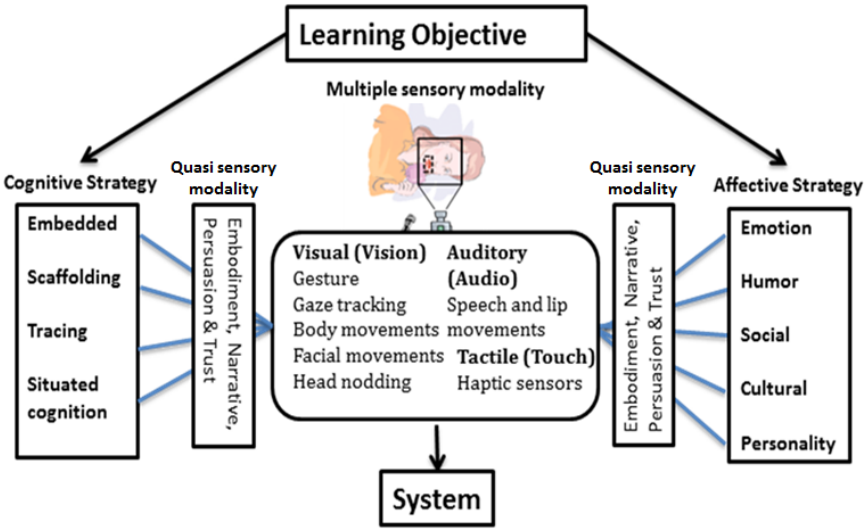


**Figure 1: The MADE Framework.**

## Case Study

Our case study is of a software system designed to apply narrative to help students visiting a museum. This system, Museum Explorer, was developed by Jesse Gerroir as part of a Master's thesis (Gerroir, 2015). This system was designed for making visits to museums better by supporting narrative visualizations including dramatic, categorical, sequential and slideshow approaches. Narrative should support continuity, storytelling and excitement (Gershon and Page, 2001). The system is designed to support an educational objective common in relationship to field trips. That is, to help students engage more with the subject matter by planning and reflection. This is a typical objective for teachers, because while students enjoy field trips, they are often unprepared and later they completely forget detail. The software designer's perspective on this issue was to build a system to help achieve these objectives using visual stories.

The Museum Explorer depicts exhibit areas from the Canadian Museum of Nature in Ottawa, Canada (see Figure 2). The design explores different visualizations to support narrative storytelling and comprehension. Gerroir evaluated the system and found the following results: the dramatic visualization ranked very highly for the majority of the ranked questions and seemed to be the one most liked overall. The categorical visualization ranked positively overall and was highest for learning. The sequential narrative visualization ranked neutral, being neither the highest nor lowest ranked visualization. And the slideshow ranked last in the qualitative satisfaction aspects.

We are using Gerroir's software as a case study for our MADE heuristic evaluation technique. We wished to explore how our MADE inspection technique might give insight about the design of the software and how it might be improved. In particular, we wish to explore affect of the narrative modality in the educational context. Gerroir gave four different narrative visualization styles for the museum experience, but did not specifically evaluate the objective of making the users more engaged and reflective, thinking about what they learned, or the user's affective experience.

## Exposition of MADE Heuristic Evaluation

The idea is to evaluate the effects of the modality that is being used in the Museum Explorer, narrative visualization. We are particularly interested in evaluating how effective the modality is in achieving the affective educational objectives. For evaluation of affect in software our proposed technique takes a standard method of evaluating usability (heuristic evaluation), and adapts it for this particular issue. Our modification is called *MADE Heuristic Evaluation*.

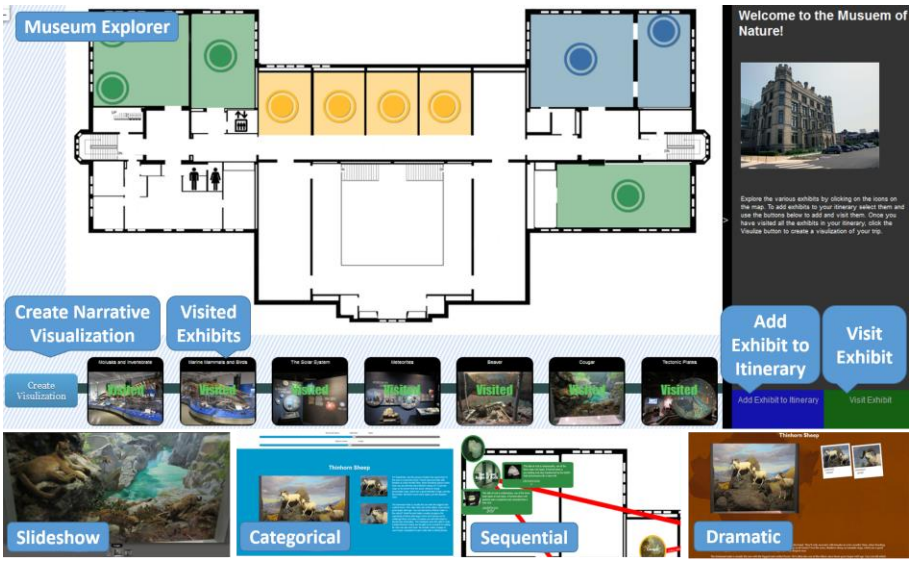


**Figure 2. The Museum Explorer with the four visualization styles.**

The MADE Heuristic Evaluation is what Nielsen and Molich (1990) term an *inspection* method, meaning the system is inspected by evaluators, rather than involving actual users. The method is therefore not as definitive, but the process is easier, cheaper, and shown to usually produce similar results. We follow the method of Nielsen and Molich, but with a new set of heuristics shown below. The proposed MADE heuristics themselves are adapted, with affect supported by the work of Norman (2005) and Kort et al. (2001) and multimodal design based on the work of Sankey (2007). The rational, and more detail, has been provided earlier (GhasemAghaei et al., 2015).

**Table 1: The Modified MADE Heuristics.**

| Affective Design | The MADE Heuristics |
|---|---|
| **Visceral** | H1: Only include design elements and modalities that support the learning objective |
| | H2: Provide help and documentation only where it is likely to be needed |
| **Behavioral** | H3: Maintain visibility of progress achievements and challenges |
| | H4: Allow the user freedom to explore but also to return to the previous step |
| | H5: Avoid or prevent actions with no learning value |
| | H6: Visualize options clearly to support exploration |
| | H7: Tailor frequent actions considering different skill-levels of learners |
| | H8: Challenge learners and provide constructive feedback if they fail |
| **Reflective** | H9: Match the learners world view in language and multimodal support |
| | H10: Maintain interface consistency to support task continuity |

To examine the technique we conducted a study where we recruited participants to evaluate the Museum Explorer system. We conducted the study in our research lab. We recruited only people with Human-Computer Interaction (HCI) evaluation experience, but not members of our own research group. A big screen was utilized

to display the Museum Explorer and other computers were provided for exploring the software. Handouts were provided with the steps of the heuristic evaluation. The study was approved by our research ethics board.

We took a qualitative approach, and both audio recorded and took notes for later detail analysis to identify issues. The sessions involved participants acting as evaluators and two facilitators representing the instructor and the software developer.

The evaluators applied the proposed inspection method to evaluate the Museum Explorer software. We as the facilitators first gave an introduction to the software. The evaluators individually went through the MADE heuristics (Table 1) and found the interface elements violated, the design problems, gave a solution and rated the severity, following Nielsen's process.

We monitored the evaluators and collected data to see how the evaluation methods worked, trying to improve the inspection technique with every evaluation session. This was an iterative process; we were recoding, going back and discussing what they have learned from the observations; we modified the proposed heuristic evaluation with evaluators' feedback and reshaped the proposed heuristics during the study.

Figure 2, shown earlier, illustrated an example screen of the Museum Explorer. This figure also shows the visualization selections with the choices shown along the bottom. The left visualization is the *Slideshow*, which is meant to display pictures in a way that is commonly used as a presentation. The next visualization is the *Categorical*, where information is arranged by the topic of the locations. The one after is the *Sequential*, which focused around giving a sense of time and place. The last visualization is the *Dramatic*, to give a sense of the user's personal experience. Figure 3 shows the first part of the visualization: view your visit. It includes the map and a timeline of exhibits visited. The map is static and simply displays the exhibits the user visited in a colour coded fashion. And the timeline shows the time user spent in a particular exhibit.

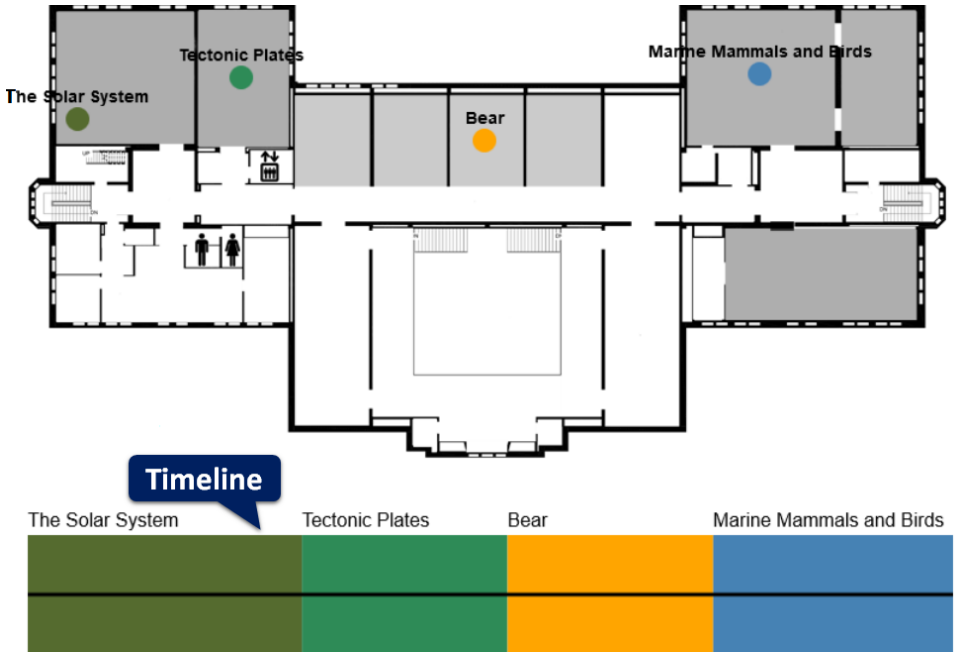


**Figure 3: The map and the timeline.**

Our particular study was about evaluating the software to see how well the proposed inspection method worked. In the evaluation sessions we collected a list of prioritized design issues and possible solutions.

As a teacher, before you take the students and go to the museum, you help them learn about the museum and plan the visit, which means when you are there, you think more deeply about what is there. Therefore, you plan an itinerary and add exhibits to the itinerary, as seen in Figure 2. When you come back to the school, you do various activities based on what you learned. The Museum Explorer is a system to support this process.

As we mentioned earlier, in the MADE inspection technique we are considering multimodal and affective aspects. For instance, an evaluator may think that when the software did nothing, it should have done something extra, and there should be some sort of multimodal support more than what is there. The multiple sensory modalities we are considering are such as music, sound or movies, and the quasi-sensory modalities are

narrative or storytelling (does it help a story?), or persuasion (was it persuading about the right or wrong thing?). Below we discuss our findings using the MADE Heuristics Evaluation.

## MADE Heuristics

In the study participants read the heuristics, explored, and then reflected on the software. They then wrote comments about the system filling in a table (see Table 2) with respect to the different heuristics. For heuristic severity we adapted Nielsen's severity ratings for usability problems (Nielsen, 1995), changing their emphasis to reflect learning impact.

**Table 2: The MADE Heuristic Evaluation Table.**

| MADE Heuristic(s) Involved | Interface Element | Usability Problem | Solution | Severity (0 to 4) |
|---|---|---|---|---|
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

For each heuristic, we show below participant comments that reflect on the effectiveness of the heuristics. Overall, participants were able to understand the heuristics and apply them sensibly. Each of the ten heuristics led to a useful comment at least once. The heuristics invited more reflection on modality as well as affect. However, some participants commented more than expected about simple usability concerns; see H2 and H6 as noted below. Of course, poor usability will inevitably have an effect on learning.

**Table 3: MADE Heuristics result, participant comment, and severity.**

| **Heuristics** | **Participant comment** | **Severity** |
|---|---|---|
| H1: Only include design elements and modalities that support the learning objective | *P3: The interaction flow does not make creating the visualization the focus for the learner, e.g. the activity should be something like "let's make a slideshow"* | 3 |
| H2: Provide help and documentation only where it is likely to be needed | *P4: The* objective *of adding exhibits and then visiting them is unclear, as a user I would simply click on the circles and try to explore directly. This is for the main page. Visualizations are clear in this regard [Note the usability emphasis.]* | 1 |
| H3: Maintain visibility of progress achievements and challenges | *P4:* Challenges *and achievements aren't highlighted in all visualizations, some of them are such as the dramatic visualization, while others just display what has been done. It tells you what you found, and it tells you are missing something which you have to go back and explore again* | 2 |
| H4: Allow the user freedom to explore but also to return to the previous step | *P1:* Should *allow the user to generate a more personal story of their experience at the museum including photos they have taken, comments, etc.* | 2 |
| H5: Avoid or prevent actions with no learning value | *P3: The* map *metaphor doesn't help the students engage with the topics* | 3 |
| H6: Visualize options clearly to support exploration (somethings this was misinterpreted as leading to usability and not educational aspects) | *P3: All of the buttons on the page are the same size. Similar shape and size visually groups these buttons together, but they have very different functions [Note the usability emphasis.]*<br><br>*P3: The information sidebar is not well-integrated into the rest of the interface. It updates when you click on the timeline or the map [Note the usability emphasis.]* | 3<br><br>1 |

| | | |
|---|---|---|
| H7: Tailor frequent actions considering different skill-levels of learners | *P4: Does not specify tasks for different learning levels i.e. all task provided are meant for one level and advanced users can be limited by this* | 3 |
| H8: Challenge learners and provide constructive feedback if they fail | *P4:* There *is no [challenge in the] task itinerary, it's just adding exhibits to a list* | 3 |
| H9: Match the learners world view in language and multimodal support | *P1: No dynamic visual representation of dynamic objects like animals, big bang*<br><br>*P4: There is an emphasis on images, and a lack of use of multimedia such as audio, movies, videos, etc., which is less like the real world where multimedia is employed to help students learn. This applies to the dramatic, categorical and possibly sequential* | 4<br><br>2 |
| H10: Maintain interface consistency to support task continuity. | *P4: The* timeline *isn't available in each exhibit which affects the visit exhibit element. Having the timeline would allow for navigation between previously visited exhibits and allows for more freedom to explore. The timeline has been employed in other places such as the visualizations so it is not consistent with the remainder of the user interface [Comment also related to H4.]* | 2 |

## Discussion

This paper described an inspection method used to evaluate a Museum Explorer system. Instead of evaluating the usability we were trying to evaluate the effectiveness of the affective modality in this educational software. The system helps the students to make a visual story of their trip to the museum, so the modality that we were interested in was the narrative aspect.

The participants, using our inspection method, found a number of issues in the system that might lead to significant improvements. For example: "the *map metaphor doesn't help the students engage with the topics"*. They also identified many places where multimodality (animation, music, video) would have been beneficial, but was not provided. The participants also found the system surprising for having no dynamic visual representation of dynamic objects like animals; being passive and even boring (just adding picture and comments, but no humor, no music, no video), and did not bring the engagement expected from narrative. They also found inconsistencies (e.g. the timeline) and lack of continuity in the interface, which also detracted from engagement. Overall, the participants also thought the design would not be motivating for users. They commented that creating a visual narrative should be exciting like generating a more personal story of their experience at the museum.

Our primary purpose in this case study was not to evaluate the Museum Explorer system, but rather to test and refine our MADE inspection techniques. Overall, we were pleased and feel the points identified above would lead to significant improvements in the Museum Explorer. However, as our sessions progressed, we discovered several important ways to improve our inspection methods.

During the first observation sessions we noticed that participants interacting with the system during the walkthrough focused a lot on usability issues. This resulted in a substantial loss of emphases on modality and learning. For instance, one participant commented: "*The buttons for add exhibit to itinerary and visit exhibit do not look buttons, but create visualization does look like a button, which might confuse the user.*" (P1).

In typical usability inspection methods the goal is clear: usability. In our MADE method, however, the issues are complex. In particular, we assume that there is a specific educational objective and a modality with some known advantages is being used to help support that objective. We found that in order for participants to use our methods, we needed to state the objectives and the modality advantages more clearly and explicitly. For example, in the case of the Museum Explorer system, the objective was to make leaners more engaged and reflective (think about what they learned) and the strategy was narrative, to help the learners tell a visual story. We found that the more explicitly we stated these objectives, the better our participants would understand and usefully apply the inspection methods.

Therefore, we modified the inspection method, and came up with the heuristics described in Table 4. The general structure is similar, but it has a stronger focus on the particular learning strategies and modalities

involved. It also emphasizes Kort et al.'s affective model (2001), depicted in Figure 4. This identifies four phases of learning and the affective character of each. The first phase is encouraging exploration with positive affect. The second phase introduces challenges, and negative affect is expected. The third phase is to support overcoming challenges and reduce the negative affect, and the fourth phase is to affirm learning and restore positive affect.

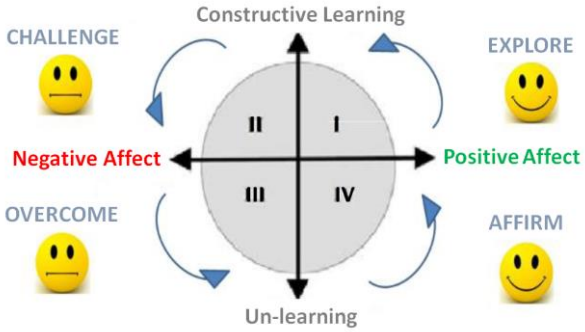


**Figure 4: The affective model of education from Kort et al.**

We improved the affective heuristics and adopted it in later evaluation sessions. Table 4 shows our final version of the affective heuristics.

**Table 4: The final version of the MADE Heuristics.**

| Affective Design | The MADE Heuristics (Refined Version) |
|---|---|
| **Visceral** | H1: Design elements and modalities should support the affective learning strategy |
| | H2: Ensure help and documentation is provided where needed but does not distract from affective learning strategy |
| **Behavioral** | H3: Maintain visibility of progress, affirming challenges already overcome, and those remaining |
| | H4: Allow the user freedom to explore but also to return to the previous step |
| | H5: Avoid or prevent actions with neither feedback to help overcome, nor affirmation when success |
| | H6: Visualize options clearly to encourage exploration |
| | H7: Tailor actions to be encouraging at first and efficient later, while learners are attempting to overcome challenges |
| | H8: Challenge learners and provide constructive feedback if they fail, and affirming success when they succeed |
| **Reflective** | H9: Match the learners world view in affective strategy and multimodal support |
| | H10: Maintain interface cohesion to support affective strategy |

## Conclusions

This paper focused on an inspection evaluation technique for evaluating affect in educational software. We presented a case study where participants used the technique to evaluate a Museum Explorer system. The system featured narrative and visual stories to support greater engagement and reflection. Our inspection technique was a heuristic evaluation based on a technique from usability evaluation but adapted for multimodal affective systems.

In our case study participants were able to apply the inspection technique and make useful comments that would significantly improve the system. However, by observing the participants and by qualitative analysis of our observations and transcripts of participant think-aloud comments, we were able to identify several ways to improve our inspection technique. In particular, we identified a need to de-emphasize ordinary usability, state explicitly the educational strategies and modality advantages involved, and align with the model of Kort et al. (2001) on the emotional trajectory of learning.

To summarize, our MADE inspection technique is an easy way to help evaluators for quickly assessing affect in multimodal educational environments. In future work, we suggest conducting further case studies using our refined techniques and especially finding case studies that present significant learning challenges.

## Acknowledgments

We thank the participants, HotSoft lab colleagues, and acknowledge an NSERC Discovery Grant.

## References

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (Eds.), (1956). *Taxonomy of Educational Objectives – The Classification of Educational Goals – Handbook 1: Cognitive Domain*. London, WI: Longmans, Green & Co. Ltd.

de Lima, E. S., Feijó, B., Barbosa, S., da Silva, F. G., Furtado, A. L., Ciarlini, A. E. M., & Pozzer, C. T. (2011). *Multimodal, multi-user and adaptive interaction for interactive storytelling applications*. In: Games and Digital Entertainment, IEEE SBGAMES. (pp.206-214).

Gerroir, J. (2015). *Constructing Visual Narratives of Museum Experiences*. Master's thesis, Applied Science in Human Computer Interaction, School of Information Technology, Carleton University, Ottawa, Canada.

Gershon, N. & Page, W. (2001). *What storytelling can do for information visualization*. Communications of the ACM 44(8), (pp.31-37).

GhasemAghaei, R., Arya, Ali., & Biddle, R. (2015). *The MADE framework: Multimodal software for affective education*, In Ed-Media: World Conference on Educational Multimedia, Hypermedia, and Telecommunications, Montreal, Canada, Association for the Advancement of Computing in Education.

GhasemAghaei, R., Arya, Ali., & Biddle, R. (2015). *Multimodal software for affective education: UI evaluation*, In Ed-Media: World Conference on Educational Multimedia, Hypermedia, and Telecommunications, Montreal, Canada, Association for the Advancement of Computing in Education.

Kort, B., Reilly, R., & Picard, R. W. (2001). *An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion*, ICALT ’01.

Kühnel, C. (2012). *Evaluating multimodal systems*, In: Quantifying Quality Aspects of Multimodal Interactive Systems, (pp. 13-21). Springer.

Nielsen, J. & Molich, R. (1990). *Heuristic evaluation of user interfaces*, Proc. ACM SIGCHI, Human Factors in Computing Systems, CHI '90, New York, NY.

Nielsen, J. (1995). *How to Conduct a Heuristic Evaluation*, http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation.

Norman, D. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things*, Published by Basic Books.

Sankey M. D. (2007). *How to Develop 15 Multimodal Design Heuristics in 3 Easy (Not) Lessons*, Intl. Journal of Pedagogies and Learning.

Turk, M. (2014). *Multimodal Interaction: A review*, Pattern Recognition Letters, 36, (pp.189-195).