

Getting Under the Skin of Learners: Tools for Evaluating Emotional Experience

Madeline ALSMEYER¹, Rose LUCKIN², Judith GOOD¹

¹*IDEAS lab, School of Science and Technology, University of Sussex, UK*

²*London Knowledge Lab, Institute of Education, University of London, UK*

Abstract: This paper reports a study that evaluated a methodology and tool for eliciting the emotional experience of language learning. It has looked at the relative rate of change of different emotions over the course of a lesson and also the differences between self-report and observer reports of changes in emotion. Implications for the design of affectively intelligent systems are discussed.

Keywords: Emotion and affect, language learning, classroom-based experimentation, qualitative, case study, think aloud

1. Introduction

Successful language learning depends on the willingness of the learner to interact with other members of a learning group, or a community of native speakers. Learning which depends on such sociality is undoubtedly governed by a learner's affective state. The work reported in this paper is concerned with clarifying the affective needs of such learners, in support of the design of affectively intelligent systems [1-5].

Our research is concerned with how a learner's affective state is related to the context in which learning occurs. Can a particular context be enhanced to support the emotional experience of learning? Are there ways in which learners (and others in the context) can adapt the learning context to better support their affective needs?

This research focus is grounded in sociocultural theories of learning [6, 7] and a belief that learning within an individual is the result of that individual's interactions with others within a particular cultural context. Learning is therefore dependent upon the nature of the context and what it affords learners in terms of physical tools, people and the location itself [8, 9]. If we accept that a learner's affective state (at least, in part), is also dependent upon her social interactions then both cognitive and affective states can be influenced by what a context is able to facilitate.

To evaluate the effectiveness of an affective learning environment, reliable data about the emotional experiences of the participants are needed. Such data can take the form of biophysical sensor recordings, interaction logs [10] and self-report [11].

The aim of this study was to evaluate a methodology and a set of tools for eliciting self-reports of emotional experience. This paper begins with an exploration of the tools and methodologies currently adopted in analysing students' emotional experiences. This is followed by a description of the tool and methodology developed by the authors of this paper. Lastly we describe a study carried out to evaluate the tool and methodology and present the preliminary results of this study.

2. Measuring Emotional Experience

Relatively few studies have attempted to uncover students' emotional experiences of learning. Those that have [12-16] have made two clear methodological decisions. The first relates to when to collect a participant's self-report: throughout the academic experience, directly afterwards or using stimulated recall, i.e, using a device to provoke recollection of an event? The second relates to the appropriate method for collecting self-reports. There are three main methods: free response approach, the discrete emotion response approach and the dimensional emotion response approach.

2.1. Free Response Approach

This approach asks participants to describe how they felt during an academic experience in their own words. The verbalisations are then clarified through participant-researcher discussion until a consensus is reached.

This approach allows participants to describe their emotional experiences in a personally meaningful way, and is generally believed to result in specific and accurate self reports of emotional experience [11]. It relies however on participants being comfortable and capable of describing their emotions and upon the researcher's skills in conducting the interviews. Furthermore, any clarification discussion with the participants requires them to feel comfortable enough to defend or elaborate on their verbalisation. This may render the technique more appropriate for more mature participants, or those with high emotional intelligence [17].

2.2. Dimensional Emotions Approach

This approach [18] attempts to build a structured description of an emotion within a dimensional space, using axes (which the researchers believe adequately differentiate one emotion from all others). Such studies require participants to locate their emotion in the space constructed by the two or three axes.

This approach revolutionised the way in which researchers talk about emotion, but it suffers from a number of limitations. For example, describing an emotional experience using dimension axes such as valence, arousal and tension is an abstract task, as participants are unlikely to break down their own emotional experiences in this way when they are reflecting on them personally. Furthermore, it is difficult to judge whether participants who chose the same dimensional space actually felt the same emotion, as it is possible, for two completely separate emotions to occupy roughly the same dimensional space, (for example, fear and anger share the same region of a two-dimensional space, constructed by the axes valence and arousal) [19].

2.3. Discrete Emotions Approach

This approach requires participants to describe the way they are feeling using a given list of words. Participants either rate their experienced emotions along a scale, e.g. a Likert-style scale or a verbal scale (strongly – not at all), or respond to questions or statements, which summarise a particular emotional state without directly mentioning the emotions. Methods vary from participants reporting only their most pertinent emotions, to reporting a degree of strength for all the emotions listed by the researcher.

This method suffers from various drawbacks, for example, by specifying only a limited set of words, the researcher may be restricting what the participant can report as experienced with a resultant loss of data. Furthermore, different studies using this method have tended to use emotional descriptive words which are applicable to their study or participant group, meaning that results cannot be compared across studies. Lastly, using a set of descriptive emotional words may prompt the participant to remember something more about their experience, but may also lead the participant to falsely report an emotion in order to please the researcher.

3. An Evaluative Emotion Reporting Tool and Methodology

The tools and methodologies described in this paper are intended to streamline the process of collecting self-reports of emotional experience within a real world educational context. The tools developed must be able to collect data quickly, easily and as reliably as possible. Stimulated recall was chosen as it allows self-reports on emotional experience to be collected in a way which does not interfere with the participants' learning within lessons, and it provides flexibility with respect to when the subsequent interviews can be carried out.

In conjunction with the stimulated recall approach, the study used a discrete emotions approach to structure the collection of self-reports. This approach was chosen since the participants' age range (13 – 18 years) meant the participants might lack the vocabulary or confidence necessary to adequately report their emotional experience using a free-response approach. It was also our concern that the level of abstraction required for the dimensional approach would be unsuitable for our participants.

We view this study as a starting point of a user centred design process for the development of an intervening technology. As such, we need to have an understanding of the types of emotional language and labels that are appropriate for use within this context. Those adopted for use in this study were based on Pekrun's work [13] which found that the common emotions experienced during academic interactions were: enjoyment, hope, pride, relief, anger, anxiety, shame and boredom.

The first phase of the tool aims to analyse the emotional state of the learner at the beginning of the learning experience, since without this knowledge it is difficult to contextualise the resulting reported experience. The tool is made up of a number of sliders which are moveable along an axis of "not at all X" where X is the emotion being evaluated to "extremely X". We chose to use semantically labelled sliders, since results from pilot studies suggested participants found Likert scales quite restrictive.

The second phase of the tool is used during the stimulated recall interview to collect each participant's self report of emotional experience. The tool simply asks the learner to report how they remember feeling at a certain point in their academic experience in comparison with how they remember feeling at the beginning of the experience by selecting either "More X" (where X is the emotion currently being investigated), "The same" or "Less X". Qualitative terms have been used since we are predominantly interested in how the emotional experience *changes* with time, and collecting this information doesn't necessitate we know the exact quantities of change, just the direction of change. Furthermore, the interviews generally took place between 1 day and a week after the learning experience itself, we therefore believed it would be less suitable to ask our participants for precise quantities of change.

4. Methodology

4.1. The Study

The aim of this short study was to begin the evaluation of an emotion reporting tool and methodology, based on how easy the tool was for the participants to use, how flexibly the tool could fit into a real learning context and finally, how reliable the tool was in terms of collecting participants' emotional experiences of learning.

4.2. Participants

The study was carried out in a British school, with 3 separate German language classes in two separate school years, year 9 (13 – 14 year olds) and year 12 (16 – 17 year olds). A total of 22 students in 11 pairs took part. 7 pairs of students were selected from the year 9 classes to participate in the study. The year 12 class is significantly smaller in size which meant all 8 students (i.e. 4 pairs) within the class took part.

4.3. Procedure

The study took place over the space of a month in the autumn/winter term. One pair of students were video recorded per classroom session. Typically the pairs of students chosen were classroom neighbours, or shared the same table. Each member of the pair was required to complete the first part of the emotion reporter tool at the beginning of the classroom session. Throughout the lesson, two pieces of footage were recorded for each student. One recording was made as they participated in a verbal activity (generally with the teacher) in front of the whole class and the second recording was made as they participated in a small group activity. The participants completed the stimulated recall interview individually within one week of the classroom activity. During this interview the participants were shown the clips of themselves participating in class. Each video clip was paused once every 30 seconds at which point the participant was asked to complete phase 2 of the emotion reporting tool (I.E reporting for each emotion whether they recalled feeling more X, less X or the same X, where X is the emotion being reported on). After reporting on their emotional experience, a semi-structured interview was conducted in order to ascertain whether any emotions had been experienced by the students which hadn't been enquired about in the emotion reporting tool and in order to capture thoughts, feelings and suggested changes with respect to the use of the tool and methodology.

Phase 2 of the emotion reporting tool was used again at a later date by an observer. During this phase of the study, the observer watched each piece of footage and independently gave their perception of the emotional change experienced by each of the participants at 30-second intervals, again systematically over each emotion.

Lastly, the video data was coded by a researcher along four dimensions. These were: changes in facial expression (frowning, smiling, breaking eye contact, as well as touching the face or hair), changes in body language / posture (sitting up, slouching, fidgeting), changes in vocalisations (intonation, speed of speech, fillers, long pauses) and lastly noting what was occurring within the context (such as, what the task was, who the participant was working with and in front of).

5. Results

The data presented is based on the 22 transcribed interviews. We provide a general description of the results space and in addition a detailed case study of one participant's complete data set. This analysis begins to reveal the reliability of the self-reports collected, as well as some of the inherent challenges and limitations of this approach.

5.1. *How Do We Get under the Skin of Learners?*

Three findings emerged from the 22 transcribed interviews. First, a number of the emotional terms used within the tool were inappropriate for use within the particular study context. For example, from a total of 22 participants, 12 reported that they never experienced anger at school and 6 commented that they never felt hopeless during the school day. In addition, the term "motivation" caused some confusion for the year 9 students, 6 of whom asked to have it explained to them.

Second, the interview data suggests, when given the choice, that the majority (well over half of those asked) would prefer to discuss and reflect on their emotional experience of learning with a group of friends, rather than at home on their own, with their teacher, or an independent adult. However, three of these participants believed completing this activity with their friends might render their reports less reliable.

Lastly, phase two of the tool requires the participants to compare the way they remember feeling during a certain segment of the video with how they remember feeling at the outset of the learning experience. Although most of the students believed the video helped them remember fairly accurately how they were feeling during the learning experience, fewer felt confident they could accurately remember how they were feeling at the outset of the lesson (prior to the video data being shot). Furthermore, some students reported feeling quite restricted by using the terms *More X*, *Less X* and *The same* when discussing their emotional experience. Their preferred means of discussing their comparative emotional experiences varied from either preferring to report on experience using a Likert scale type instrument, or introducing more specific qualitative terms into the second phase of the tool, such as "A Little More X", "A Lot More X" and so on (suggested solely by Yr 9 participants).

5.2. *The Emotional Experience of Foreign Language Learning: Data Trends*

One aim of this study was to develop an understanding of how a learner's emotional experience changes over time. From the data collected with our tool, it is possible to infer that emotions such as anxiety and embarrassment are liable to change significantly throughout an academic experience ($\chi^2 = 37.4$, $p = .000$, $\chi^2 = 34.9$, $p = .000$ respectively). The emotions most subject to large swings of change (i.e from More X to Less X within a 30 second segment) are relief and anxiety. The data also suggests that emotions such as boredom, hopelessness and anger are least likely to change, therefore if a learner starts a session feeling bored, the learning context is rarely able to alter this assessment.

5.2.1. *A Case Study: Clare, a 17 Year Old Female A Level Student*

The self-report data collected using our tool with Clare has been analysed and compared with two additional perspectives: that of the observer and that of the

researcher's video coding. This enables us to triangulate data sources to estimate the reliability of looking beyond self-report data collection for emotional experiences and provides evaluative data about the reliability of the data collected with the tool.

The observer's recorded beliefs about Clare's emotion change over each segment of the video data is presented in the middle section of Figure 1 and 2 alongside Clare's self-report data. The video data, coded by the researcher using four separate dimensions: changes in facial expression, changes in body language / posture, changes in vocalisations and lastly contextual changes is presented in the top section of Figure 1 and 2. The bottom section of Figure 1 and 2 depicts the emotional state as reported by Clare at the outset of the experience, using phase 1 of the emotion reporting tool.

The key difference between these two experiences can be found in Clare's success in her interactions. In Experience 1, we see Clare interacting successfully, albeit with the odd mistake, in contrast during Experience 2 she is unable to answer the question posed by the teacher and must watch whilst another student correctly answers the question. This perceived success can be correlated with a marked difference in visible emotional reaction, which is perhaps one reason why the tool appears more reliable at capturing her emotional experience in Experience 2. In other words, where there is an overt emotional reaction, the tool and methodology is capable of showing reliability.

However, what is perhaps more interesting are the less expected reported experiences. One would expect that as a student interacts successfully, and without many visible signs of discomfort, their confidence and pride and possibly their motivation would increase and their anxiety decrease. However, Clare reported feeling less confident, less proud and less motivated after a 30 second period of successful interaction with the teacher. Furthermore, if we compare Clare's reported emotional experience of being unable to answer the teacher's questions, with what Clare reports when she makes a small mistake (comparison of the Experience 1 at 30 sec, and Experience 2 at 30 sec), we see the reported emotional experience is similar.

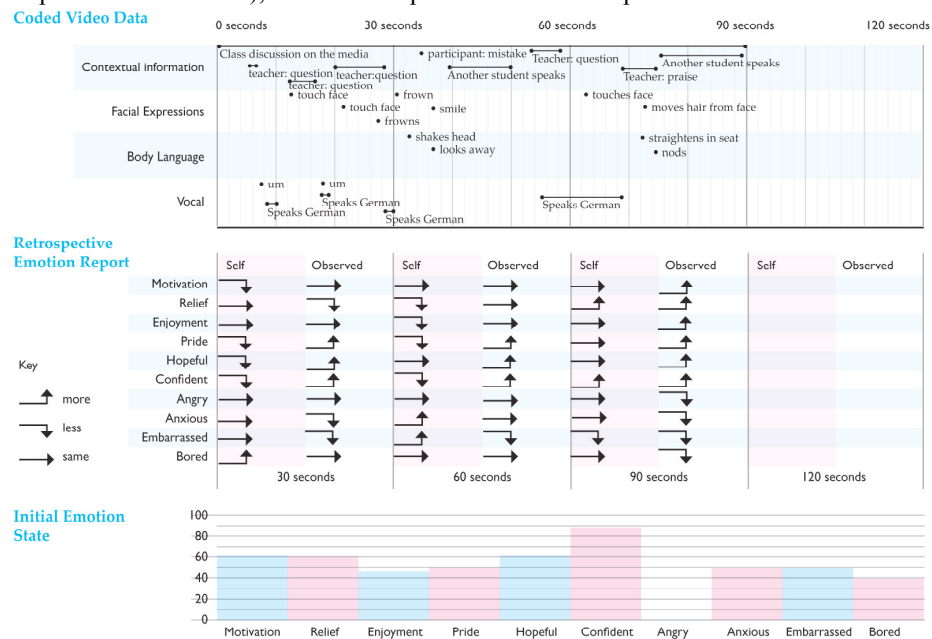


Figure 1. A graph comparing Clare's reported experience and the observed experience for Experience 1.

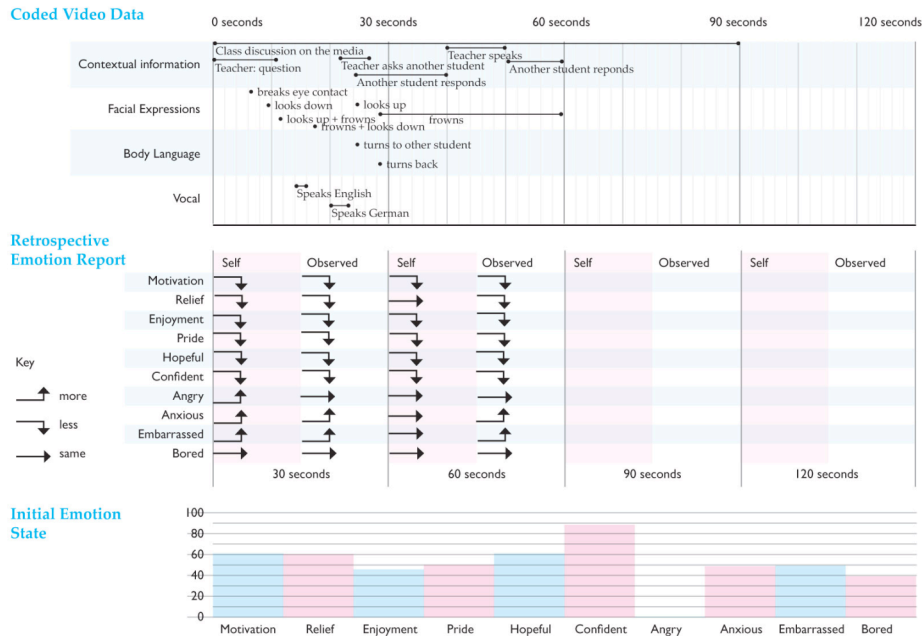


Figure 2. A graph comparing Clare's reported experience and the observed experience for Experience 2.

6. Discussion and Conclusion

This small study has shown there are clear and abundant discrepancies between what an observer believes the emotional reaction of a learner to be and what a learner reports their emotional reaction to be. In this case, who can we say is correct when it comes to interpreting a learner's emotional reaction and whose judgement should we rely on when adapting our educational interactions to the affective state of the learner? Clearly one contribution to the AIED community is to encourage the investigation of using multiple data sources when attempting to infer the emotional state of a learner.

This work also raises the question of whether there is a logical flow of emotions within a learning experience. If a student is interacting successfully with the teacher, displaying few visible signs of discomfort a logical model of emotional experience might conclude that this student, based on their context, is experiencing low anxiety and an increased feeling of pride and confidence, yet our study has found this logical flow of emotions is not what is experienced by our learner, or at least not what the learner reports they have experienced.

The findings of this study are furthermore capable of giving designers some insight into the differences in the temporal lives of emotions. For example, we have shown emotions such as boredom and hopelessness are less liable to rapid change over a learning session, and in contrast emotions such as relief and anxiety change rather more rapidly. This finding might be particularly interesting to designers of affectively intelligent systems if we correlate it with Craig's work [16] which concluded that boredom is negatively correlated with learning, since it provides the beginnings of a framework which starts to identify which learning emotions an educational system could *realistically* and with the most impact identify and react to.

When severe emotional changes are experienced, the tool, based on this case study, shows itself to be fairly reliable. However, it is entirely possible that when smaller emotional changes (as hypothesised to have occurred in Clare's emotional experience in Experience 1) are experienced, the tool becomes far less reliable. If affectively intelligent tools are to be effective they must be able to pick up on small changes in emotional experience which appear to be less prevalent in the type of body language, facial expression or vocalisations displayed by the learner, in order to avert some of the potentially larger changes in emotional experience.

A more detailed analysis of the self-report data alongside the video coding data together with triangulation will clarify the reliability of the data collection tool, in terms of the participant's self report of emotional experience versus an observer's report, as well as providing deeper insight into the extremities of emotional expression in the context of a formal learning setting.

References

- [1] Del Soldato, T., du Boulay, B, *Implementation of motivational tactics in tutoring systems*. International Journal of Artificial Intelligence and Education, 1995. **6**: p. 337-378.
- [2] Kort, B., Reilly, R. *Analytical Models of Emotions, Learning and Relationships: Towards an Affect-Sensitive Cognitive Machine*. in *Conference on Virtual Worlds and Simulation*. 2002. San Antonio, Texas.
- [3] Porayska-Pomsta, K., Pain, H. *Providing cognitive and affective scaffolding through teaching strategies: applying linguistic politeness to the educational context*. in *Intelligent Tutoring Systems*. 2004. Brazil: Springer-Verlag.
- [4] Zhang, L., Barnden, J., Hendley, R. *Affect detection and metaphor in E-Drama*. in *Workshop on motivation and affect in education software associated with the 12th international conference on artificial intelligence in education 2004*. Amsterdam: IOS press.
- [5] Conati, C., Zhou, X. *Modelling students' emotions from cognitive appraisal in educational games*. in *ITS 2002*. 2002. Biarritz, France.
- [6] Vygotsky, *Thought and language*. 1962, Cambridge, Mass.: The MIT Press.
- [7] Vygotsky, *Mind in society: the development of higher psychological processes*, ed. M. Cole, Scribner, S., Soubberman, E. 1978, Cambridge, Mass: Harvard University Press.
- [8] Luckin, R., *Learning contexts as ecologies of resources: issues for the design of educational technology (CSR Technical Report 578)*. School of Cognitive Science Research Paper, 2005.
- [9] Luckin, R., du Boulay, B., Smith, H., Underwood, J., Fitzpatrick, G., Holmberg, J., Kerawalla, L., Tunley, H., Brewster, D., Pearce, D, *Using mobile technology to create flexible learning contexts*. Journal of interactive media in education, 2005. **22**.
- [10] Litman, D., Forbes-Riley, K. *Predicting student emotions in computer-human tutoring dialogues*. in *42nd annual meeting of the association for computational linguistics*. 2004. East Stroudsburg.
- [11] Scherer, K., *What are emotions? And how can they be measured?* Social Science Information, 2005. **44**(4): p. 695-729.
- [12] Do, S., Schallert, D, *Emotions and classroom talk: toward a model of the role of affect in students' experiences of classroom discussions*. Journal of educational psychology, 2004. **96**(4): p. 619-634.
- [13] Pekrun, R., Goetz, T., Perry, R., Titz, W, *Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research*. Educational psychologist, 2002. **37**(2): p. 91 - 105.
- [14] O'Regan, K., *Emotion and e-learning*. JALN, 2003. **7**(3): p. 78-92.
- [15] Wosnitza, M., Volet, S, *Origin, direction and impact of emotions in social online learning*. Learning and instruction, 2005. **15**: p. 449-464.
- [16] Craig, S., Graesser, A., Sullins, J., Gholson, B, *Affect and learning: an exploratory look into the role of affect in learning with AutoTutor*. Journal of educational media, 2004. **29**(3): p. 241 - 250.
- [17] Goleman, D., *Emotional intelligence: why it matters more than IQ*. 1996, London: Bloomsbury.
- [18] Russell, J., *A circumplex model of affect*. Journal of personality and social psychology, 1980. **39**(6): p. 1161-1178.
- [19] Cowie, R., Cornelius, R, *Describing the emotional states that are expressed in speech*. Speech Communication, 2003. **40**(1-2): p. 5-32.