

MEASURING EMOTION: A NEW EVALUATION TOOL FOR VERY YOUNG CHILDREN

Yusrita Mohd Yusoff¹, Ian Ruthven², and Monica Landoni³

¹Universiti Utara Malaysia (UUM), Malaysia, yusrita@uum.edu.my.

²University of Strathclyde, United Kingdom, Ian.Ruthven@cis.ac.uk.

³University of Lugano, Switzerland, monica@usi.ch.

ABSTRACT. In this paper, we report our experience researching with very young children (age 3 to 5 years old) in a series of evaluation study in one local nursery in the UK. A new evaluation tool, named the Fun Semantic Differential Scale (FSDS) has been developed and tested with the nursery children to understand how they respond to it. We report on our experience of using these scales to evaluate the fun experienced by children after interacting with computer products.

Keywords: evaluation, very young children, measuring emotion, Semantic Differential (SD) scale, Self-Assessment Manikin (SAM)

INTRODUCTION

Nowadays young children use computer products everywhere, including in school and at home. One main issue for children's computer products is that the products are created by adults. Adults play major roles in the product development stages from early concepts until end products. This raises questions regarding whether adults understand what children need and prefer. From an evaluation perspective we could also ask whether a fun computer product from adult's perspective is the same from a child's perspective. Several researchers have incorporated children's perspectives in evaluating children's computer products. However these have been aimed at school age children. With more products being aimed at younger children, we could ask if it is worth involving very young children in evaluating computer products.

Some researchers might think they are too young and incapable to participate in evaluating computer products. However, issues like usability and fun are very important to understand from a child's point of view. Mohd-Yusoff, Landoni, & Ruthven (2010) discuss the importance of fun from researchers such as Dorman (2000), Read & MacFarlane (2000), Carroll (2004), Shneiderman (2004), and Blythe, Hassenzahl, & Wright (2004). As Read & MacFarlane (2000) note, children, play, and fun are very related and fun is something that children know about and are expert in. As children experience fun, they can talk about it, describing it as excitement, play, laughter, and feeling good. Thus, we want to develop a new evaluation tool that is suitable for use with and by very young children to measure emotion.

This paper reports on our experience researching with very young children in evaluations: what types of instruments can very young children understand and what differences are there from existing instruments to children and adult. Firstly we review the literature on measuring emotions, then we focus on Semantic Differentials and the Self-Assessment Manikin (SAM), then we introduce our new instrument, the Fun Semantic Differential Scale (FSDS), followed by reporting on findings of study conducted at a nursery in the UK with very young children.

MEASURING EMOTIONS

Research on emotion and computer is very interesting and crucial because until recently usability metrics have only focused on functionality and usability, whereas, what users feel is also very important. If a usable product makes users sad, stressed or angry while using it, the goals of the products will be defeated. Generally, there are two definitions of emotion given by Agarwal & Meyer (2009): “Emotion is a psychological reaction to events relevant to the needs, goals, or concerns of an individual” and “emotion is comprised of physiological, affective, behavioural, and cognitive components.” p.2920. They stress only a few case studies in Human Computer Interaction (HCI) context have investigated emotions, most publications discuss emotion metrics at the theoretical level rather than the application level, and the HCI community has overlooked the importance of emotional components in user experience. However, they admit it is very challenging to study human emotion as human have difficulties to describe their feeling and to differentiate their emotions.

For very young children in particular, fun is an important component in the success of a product and the ability to measure fun and other emotions has become crucial to the development of age-appropriate computer products (Mohd-Yusoff et al., 2010). However, finding an instrument that can assist young children at nursery age (3 to 5 years old in UK) to express their fun emotions after interacting with computer products is very difficult and challenging. The young children need an instrument which is simple, easy, clear, and suitable that can assist them to express their emotions towards computer products. Our approach is based on a visual version of semantic differentials.

Semantic Differential (SD) Scale

Osgood, Suci, & Tannenbaum (1957) state that semantic differentials are a technique that was developed to measure the “meaning” of concepts or objects. This technique involves pairs of bipolar adjectives used as measurement scales. In Pearson & Bailey (1980), it is assumed that adjectives can be used in a rating process to measure the feelings associated with attitude. The measurement process involves rating the measurement object on a series of bipolar adjectives scales separated by a fixed number of intervals (usually seven). The participant is asked to place a cross in the interval between a pair of adjectives to express his or her feeling toward the measurement object. For example to measure feelings towards computer product, a series of bipolar scale is shown in Figure 1.

good: ____: ____: ____: ____: ____: ____: ____:bad
happy: ____: ____: ____: ____: ____: ____: ____:sad

Figure 1. Examples of Semantic Differential Scales

Pearson & Bailey (1980) list three steps to construct a semantic differential for attitude measurement: (p. 61)

1. Identify the attributes, concepts or objects that are to be assessed.
2. Identify pairs of adjectives which are to be used in the assessment and create bipolar scales.
3. Assign numbers to the scale to indicate the direction of emotion and the intensity.

They claim that semantic differentials have an important metric quality and it may be regarded as interval scales, thus the results can be summed and averaged. Advantages of

using the semantic differential scales are that are very quick and easy for participants to fill in and no special training required for usability engineers to use. It is also very easy to analyse the scale statistically. However, for children, disadvantages are the textual nature of the display, and (potentially) the use of sophisticated concepts that children cannot understand (Bradley & Lang, 1994).

Self-Assessment Manikin (SAM)

To overcome the textual difficulties in Semantic Differentials, Bradley & Lang (1994) propose a picture-oriented instrument, called the Self-Assessment Manikin (SAM). They claim that SAM is an easy method for quickly assessing pleasure, arousal, and dominance. The SAM is a nonverbal instrument that consists of 12 graphics characters representing three emotional dimensions; valence, arousal, and dominance. The arousal emotions in first row are represented by five smiley faces graphics from happy to sad faces. The second row contains five different graphics representing arousal emotions with a very big 'bang' shape in the middle of character body but decreasing to a small spot in the last graphic. The third row shows the dominance emotion graphics using five different characters' sizes, from a very small in-the-middle character to a very big out-of-the-box character.

SAM has been used effectively in measuring emotional responses in various situations including reactions to pictures, images, sounds, advertisements, and painful stimuli. For examples in study of emotion recognition system (Vera-Munoz et al., 2008), biometric sensor (Jones & Troen, 2007), affective computing (Axelrod & Hone, 2005), computer games (Lim & Reeves, 2010), and user experience in HCI (Mahlke, Minge, & Thuring, 2006). Generally there are two versions of SAM, a paper-and-pencil, and computer version. Figure 2 illustrates the original paper-and-pencil version of the SAM.

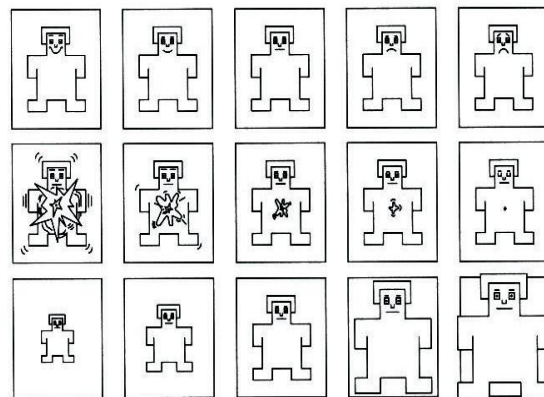


Figure 2. The original paper-and-pencil SAM version

SAM also has been used with children. For example, Greenbaum, Turner, Cook, & Melamed (1990) report their study using the SAM with 40 children to self-report the children feelings being at dentist. Participants were 3 1/2 to 7 years old children reported their feelings before and after getting treatment with the dentist by using the SAM questionnaires. However, the usage of SAM was not explained or validated in the paper. A possible weakness of using SAM with very young children is the large number of options displayed in the scale and that SAM requires very young children to interpret the images.

In our research, we have developed an evaluation tool based on semantic differential scales and picture-oriented instruments. A Fun Semantic Differential Scale (FSDS) has been developed to measure fun and has been evaluated within a nursery context to understand how

very young children respond to it. In the next section, we describe the FSDS and then our evaluation of the appropriateness of this tool.

FUN SEMANTIC DIFFERENTIAL SCALE (FSDS)

The FSDS was an attempt to develop a new evaluation tool that could elicit emotions, particularly regarding fun, from very young children after interacting with computer products. We worked with children and staff in a local nursery over a period of 15 months designing and testing few versions before finalizing it. We iterated over various versions of the FSDS with nursery children and staff before arriving at design principles in Mohd-Yusoff, Ruthven, & Landoni (2011). The FSDS final version contains photographs of a nursery-aged child with nine different facial expressions showing positive, neutral and negative feelings. The neutral photograph, representing Don't Know was the same on each sheet. The photographs are displayed as four A4 sheets of three photographs with each sheet displaying one scale containing one positive emotion (*happy, good, love, or excited*), one neutral emotion, and one negative emotion (*sad, bad, hate, or bored*). The photographs on each sheet show a child displaying the labelled emotion. Each sheet thus displays one semantic differential.

The use of photographs was an attempt to make the evaluation tool intuitive to children; by displaying photographs of children of the same age the child participating in the evaluation could recognise which emotion was being described and indicate which as relevant to them. Two versions of the FSDS were evaluated in this study: one uses a boy's photographs (*Wafiy*) in Figure 3 and one uses a girl's photographs (*Alisya*) in Figure 4.

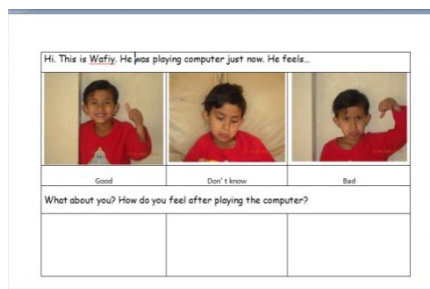


Figure 3. Wafiy FSDS



Figure 4. Alisya FSDS

STUDY IN NURSERY

The aims of the study were to observe how children respond/use the FSDS, and to detect any gender implications towards the versions. The study was conducted over a number of days, starting around 9.30am and ending no later than one hour because the nursery has its planned activities for the children. All the studies were conducted in room 4 and used only one computer which connected to internet. The participants were asked to play/explore the CBeebies website, which is based on a popular children television channel in the UK. There were 16 participations from 10 volunteer nursery children (eight boys and two girls); two children repeated the study three times, other two children repeated the study twice, and five boys and one girl participated once. However, only 12 responses gathered from the children and four responses missed due to several reasons e.g. lacking time to answer questions. All children who participated in the study described here had participated in earlier studies on different designs of the FSDS. Each participant was greeted, asked to sit on a 'playing' chair, and played any game from the CBeebies website for 5-10 minutes. Then the participant was asked to sit on an 'evaluation' chair and chose one FSDS versions. An introductory statement in the FSDS was read out and the scale was explained to the participant by pointing to each

photo and verbalizing the label. Sometimes, facilitator imitated certain emotions by using her facial expression and gesture. The participant was asked to indicate which photograph represented how he/she felt after playing the website, in effect saying which photograph was most similar to the participant's feeling. This was repeated for other three semantic differentials. Finally, the participant was thanked and given a high-five.

Findings

Children responded to the FSDS using many methods/ways; pointing finger to the photographs, verbalizing word(s), imitating photographs, nodding head, and combining few methods. We expected children will point to related photographs when answering questions but some children displayed more than pointing finger e.g., one child verbalized a reason, "hate..sometimes I don't know how to play". Some children verbalized a word or phrase to describe their emotions. Some children liked to imitate the emotions shown in the photographs, particularly when they felt positive emotions like *happy* or *excited*. Most children clearly and correctly responded to the FSDS but some children need further explanation from the facilitator before they used the tool to answer the questions. Most children chose the FSDS versions which matched their gender. Six boys and one girl chose the *Wafiy* FSDS, one boy who participated three times chose the *Wafiy* once, one girl who repeated the study twice consistently chose the *Alisya*, and one boy did not choose any tool. However, one boy who participated twice had chosen the opposite tools. Although the numbers are small this may indicate that the gender of the child shown in the FSDS is not a barrier to a child's use of the FSDS.

Limitations

Only ten children participated in the study. The biggest challenge conducting study in the nursery with very young children was to get children's complete participation in a very limited time. Each study was conducted less than an hour due to nursery has planned activities for the children every day. Besides, the studies were designed to be conducted in a nursery's natural environment; no children were forced to take part in the studies. We experienced, two boys who did not want to answer questions but only wanted to play the computer.

CONCLUSION

The aim of our research is to obtain a better understanding on what instrument is understandable and representable for use with and by very young children to assist them to express feelings after interacting with computer products. The very young children are a special computer user groups nowadays that need a specific evaluation tool to assist them to express their feelings about computer products. Our work on several iterations of the Fun Semantic Differential Scale (FSDS) has led to a stable instrument for evaluation by very young children. We planned to explore the use of the FSDS in Malaysia context and scenarios to further validate the FSDS with Malaysian kids.

REFERENCES

- Agarwal, A., & Meyer, A. (2009). Beyond usability: evaluating emotional response as an integral part of the user experience. *Proceedings of the 27th international conference* (pp. 2919-2930). Boston, MA: ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1520420>
- Axelrod, L., & Hone, K. (2005). E-emotional advantage: performance and satisfaction gains with affective computing. *CHI '05 extended abstracts on Human factors in computing systems*. (pp. 1192-1195). Portland, OR, USA. Retrieved from <http://dl.acm.org/citation.cfm?id=1056874>

- Blythe, M., Hassenzahl, M., & Wright, P. (2004). Introduction. *interactions*, 11(5), 36-37. ACM. doi:10.1145/1015530.1015545
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. Retrieved from <http://www.sciencedirect.com/science/article/B6V5V-45XTYKY-JM/2/98910d0512a8b5fe462cdec44c2a5ee9>
- Carroll, J. M. (2004). Beyond Fun. *interactions*, 11(5), 38-40. ACM. doi:10.1145/1015530.1015547
- Dorman, C. (2000). Engaging consumers, using humour in electronic commerce. Retrieved February 5, 2009, from http://www3.sympatico.ca/renn_scott/CandF3abs.pdf
- Greenbaum, P. E., Turner, C., Cook, E. W., & Melamed, B. G. (1990). Dentists' voice control: Effects on children's disruptive and affective behavior. *Health Psychology*, 9(5), 546-558. US: Lawrence Erlbaum Associates. doi:10.1037/0278-6133.9.5.546
- Jones, C. M., & Troen, T. (2007). Biometric Valence and Arousal Recognition. *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*. (pp. 191-194). Adelaide.
- Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies* (2010), 68(1-2), 57-68. Elsevier. doi:10.1016/j.ijhcs.2009.09.008
- Mahlke, S., Minge, M., & Thüring, M. (2006). Measuring Multiple Components of Emotions in Interactive Contexts. *CHI '06 extended abstracts on Human factors in computing systems*. (pp. 1061-1066). Montréal, Québec.
- Mohd-Yusoff, Y., Landoni, M., & Ruthven, I. (2010). Assessing fun: young children as evaluators of interactive systems. *ACM Workshop on Accessible Search Systems at ACM SIGIR 2010*. (pp. 11-18). ACM.
- Mohd-Yusoff, Y., Ruthven, I., & Landoni, M. (2011). The fun semantic differential scales. *Proceedings of the 10th International Conference on Interaction Design and Children* (pp. 221-224). ACM. Retrieved from <http://portal.acm.org/citation.cfm?id=1999066>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press. USA.
- Pearson, S. W., & Bailey, J. E. (1980). Measurement of computer user satisfaction. *SIGMETRICS Perform. Eval. Rev.*, 9(1), 59-68. doi:DOI= <http://doi.acm.org/10.1145/1041872.1041881>.
- Read, J. C., & MacFarlane, S. (2000). Measuring Fun – usability testing for children. University of York, York, UK. Retrieved February 2, 2009, from http://www3.sympatico.ca/renn_scott/CandF3abs.pdf
- Shneiderman, B. (2004). Designing for Fun : How Can We Design User Interfaces to Be More Fun ? *interactions*, 11(5), 48-50. Retrieved from http://delivery.acm.org/10.1145/1020000/1015552/p48-shneiderman.pdf?ip=130.159.184.116&acc=ACTIVE_SERVICE&CFID=78953224&CFTOKEN=78112260&__acm__=1327061648_e5b2169b13573c6f6504d6745383062a
- Vera-Munoz, C., Pastor-Sanz, L., Fico, G., Arredondo, M. T., Benuzzi, F., & Blanco, A. (2008). A Wearable EMG Monitoring System for Emotions Assessment. *Probing Experience* (pp. 139-148). Springer. Retrieved from <http://www.springerlink.com/index/r4n4738415tw2380.pdf>