Developing the Scale Adoption Framework for Evaluation (SAFE)

William Green, Greg Dunn, and Jettie Hoonhout
Philips Research
High Tech Campus 34
5656 AE, Eindhoven, NL
{first name.last name}@philips.com

ABSTRACT
A growing number of psychometric scales have been developed to measure a plethora of constructs within the umbrella term, *User Experience* (UX). Unfortunately, selecting an appropriate scale for UX can be difficult for the usability practitioner. Based on psychometric scale development literature, the Scale Adoption Framework for Evaluation (SAFE) has been designed to support practitioners in selecting appropriate scales. The SAFE is presented in this paper, with an example of how it could be used. However, utilizing the SAFE has emphasised the difficulty in selecting suitable measures, revealing that the SAFE will not solve this issue in isolation. Two main challenges are presented for the VUUM workshop: more explicit and transparent descriptions of psychometric scale development, and a need for more sensitive psychometric scales. Future work should build on these challenges by consulting usability practitioners directly.

Author Keywords
Psychometric scale development, evaluation, UCD.

ACM Classification Keywords
H.1.2.,User/Machine System: Design, Experimentation, Human Factors, Measurement.

INTRODUCTION
This paper is concerned with psychometric scale development and its adoption by practitioners within the User-Centred Design (UCD) process. There are many illustrative texts and practical examples of psychometric scale development in psychology literature, e.g., [4, 21, and 26]. Interpreting this literature can be difficult however; the majority of psychometric literature is written for an assumed audience that will have either psychology or statistics expertise. In this paper, the Scale Adoption Framework for Evaluation (SAFE) is presented and has been developed in an effort to support practitioners before selecting psychometric scales. A psychometric scale is a form of measurement used to assess abstract qualia subjectively experienced by an individual (e.g., individuals' personality, their emotional states, and interpersonal trust). From this point forward, 'psychometric scale,' will be referred to simply as, 'scale.' The SAFE is intended to help inform practitioners, who may not have a psychology or psychometric background, about the essential properties of valid and reliable measures. To achieve this, the SAFE illustrates the most important elements that a practitioner should look for when adopting a measure, to ensure that it is sufficiently robust and accurately measuring the construct that the practitioner needs, and it claims to measure.

Before presenting the SAFE, the UCD process is briefly reviewed emphasising how psychometric measurement of users' subjective experience supports the evaluation of products. The reasoning behind why it was believed that the SAFE would be helpful for practitioners is then presented. In the next section, we discuss development and evaluation of psychometric measures, specifically focusing on what is particularly *pertinent to practitioners* when selecting a measure. This led to the development of the SAFE. An example of how to utilise the SAFE is illustrated. Initial use of the SAFE to select a measure of emotion indicated several challenges, which has led to conclusions and recommendations for future research.

Measuring usability and user experience for UCD
User-Centred Design (UCD) is a design philosophy that focuses on the needs and interests of users and emphasises making products usable and understandable [20]. There have been many descriptions of UCD in the literature that vary in the nomenclature used. Nonetheless, [5] identified three principles for user-centred system design, which are still prevalent in the majority of the UCD descriptions, these are: 1) a user focus from the beginning and throughout the process, 2) measurement of system usage, and 3) iterative design. As [3] noted, the two editions of the Handbook for HCI, e.g., [7], also emphasise the three UCD principles proposed by [5]. Despite recent suggestion that

VUUM2008, June 18, 2008, Reykjavik, Iceland.

the emphasis on evaluation in [5] and [7] is dated [3] and not always appropriate [6], evaluation is undeniably an important aspect in UCD, often achieved via measurement.

Measurement of system usage by practitioners involved in UCD is often concerned with usability evaluation, i.e., are the products easy and comfortable to use, safe, effective, efficient, and easy to learn how to use? These usability evaluations have been done through objective (e.g., time or physiology) and subjective (e.g., perceptions, attitudes, and other scales of psychological constructs) measurement and are typically based on at least one of three dimensions outlined by ISO 9241-11 [13], which are: *Efficiency*, *Effectiveness*, and *Satisfaction*.

Subjective evaluation is often done via questionnaires, described as both the most versatile, but also most often misused research tool for HCI [19]. Despite [19]'s caveat and support for questionnaire design, development, and appropriate use, it appears that the questionnaire remains misused almost twenty years later. For example, using [13]'s dimensions, [12] demonstrated that while the measurement of both *Efficiency* and *Effectiveness* is relatively straightforward and measured in similar fashion among UCD practitioners, the measurement of users' *Satisfaction* is diverse. This diversity is illustrated by findings indicating that only 11% of satisfaction scales used by practitioners are established (i.e., valid and reliable) [12]. Furthermore, [12] reported that the 89% of "homegrown" measures vary greatly in terms of reliability, and so, recommend that practitioners should use established scales. This is not a new recommendation. For example, [15] emphasised that, "questionnaires should be elegant in terms of their reliability, validity and human factors appeal" (pg. 210). It is in this vein that [11] proposed a model to support practitioners when selecting the best approach for usability evaluation. His model outlines six approaches for evaluating usability, both objectively and subjectively for each of the three usability dimensions noted within the ISO standard above. Extending from [11]'s model, this paper focuses on subjective approaches to usability evaluation.

Obviously, performance criteria related to *Efficiency* and *Effectiveness* are important for consumer products, especially in the case of safety, comfort, and learnability. Particularly for consumer products, however, it has been increasingly accepted that other requirements related to *Satisfaction* should also be considered. Using a product should be enjoyable, engaging, and appealing [e.g., 1, 8]. These requirements have often been discussed as part of the *User Experience* (UX). UX has been particularly important given the migration from the workplace, ubiquity and growth of technology in the home, and the emergence of 'intelligent' and perhaps more complex products. Nevertheless, like usability satisfaction, UX has numerous subjective meanings that are dependent not only on the individual, but also the context of interaction. Given the growing importance of UX evaluation, [12] findings regarding the low percentage of established satisfaction scales being used by practitioners becomes even more poignant. If practitioners are creating or implementing scales that do not provide valid or reliable results, this could seriously compromise evaluations of constructs critical to the success of a product (e.g., trust measures for users' perceived trustworthiness of an e-commerce website). This paper supports [12]'s position recommending that practitioners should use established scales, but greater emphasis is placed on supporting practitioners when selecting existing scales. To do this, steps are taken to further understand the benefits of using scales from a psychometric perspective, and their contribution to the UX evaluation process.

Using scales in user experience

Psychometric scales have often been employed for evaluation when measurements are required across a number of studies to determine some psychological construct (e.g., emotion, trust). This method enables users to subjectively quantify their experiences, which map onto a construct. These findings then indicate to practitioners how a product could be improved. This use of scales has provided a number of advantages: 1) it enables testing of large quantities of participants over short periods of time, at relatively low cost, which is practical since time and financial resources to conduct user-studies have usually been limited, 2) it is an easy to apply technique (although scale development is more time-consuming and complex), and 3) it has usually been non-intrusive for participants.

To support design decisions regarding the dimensions of UX through evaluation, scales must first be developed based on a UX construct definition. For example, emotion and affect are ambiguous terms but are often postulated as integral to designing UX, e.g., [14]. This ambiguity stems from the literature itself. [22] differentiate between the terminology 'affect', 'emotions', 'moods dispositions' and 'preferences' (pg. 124). They note that difficulty in answering the question, "what is an emotion?" is related to the interchangeable use of the terms 'emotion', 'emotional', and 'affect.' An example of this can be seen in [23] in an attempt to clarify this ambiguity in their use of terminology, "emotional and affective will be used interchangeably as adjectives describing either physical or cognitive components of emotion, although 'affective' will sometimes be used in a broader sense than 'emotional" (pg. 24). This ambiguity stresses the importance of evaluating UX dimensions with scales based on specific constructs.

One possible reason for the apparent lack of psychometric scale usage for UX could be due to practitioners simply not being familiar with the exhaustive process needed to create a new scale. Choosing whether a scale is appropriate for a particular product is difficult given that part of this decision involves understanding scale construction, which is perhaps limited to those with psychology or statistics expertise. It is anticipated that new psychometric scales will be developed to evaluate UX [e.g., 10]. Given the high percentage of

homegrown measures used to measure satisfaction and the similarities between UX and satisfaction (both are subjective and abstract constructs), it is possible that the nature of UX will inadvertently encourage its practitioners to develop their own scales without properly recognising issues such as scale validity and reliability. If this were to occur, then it would only serve to propagate questions concerning the validity and reliability of UX evaluation. To prevent this possibility, this paper presents the Scale Adoption Framework for Evaluation (SAFE), which is intended to encourage and support practitioners to select established scales for UX, rather than utilising homegrown ones. But if a homegrown one must be used, then the SAFE also provides a starting point for practitioners to consider what is vital when attempting to construct a valid and reliable scale.

THE SAFE DEVELOPMENT APPROACH

Based on several methodological and statistical texts that discuss (aspects of) scale development [e.g., 2, 21, 26], the SAFE was created to portray the key elements of scale development that are, debatably, essential for scale construction. As shown in Figure 1, the SAFE is composed of three elements. Within each element, on the left, a description of the aim that developers have when considering the scale in scale construction is provided. Additionally, for each of these elements three key aspects are emphasized for practitioner's to consider when deciding on a scale for their usability or UX evaluation. The pertinent aspects of these elements are as follows:

1. Construct definition – confirming that the scale construct (abstract qualia to be measured) is suitable.

1a) Theory – is concerned with whether scale developers have based their proposed construct on a strong theoretical foundation following a literature review, experiments, or consultation with domain experts. If this is not done, then there are fundamental questions regarding what the scale is measuring, despite possibly being valid and reliable (see below).

1b) Clarity – recommends that the theory behind the construct is clear, coherent, and straight to the point. If this is not the case, it becomes more difficult for the practitioner to decide whether the theory adequately describes the construct.

1c) Discriminating – acknowledges that it is important for scale developers to also consider what the construct does not include at a theoretical level. For the practitioner, this will simply mean that it will make it easier to see what the scale does and does not measure.

2) Scale validity – the extent to which a tool or method measures what it is intended to measure.

2a) Construct validity – it is important that practitioners see evidence from developers that the constructed scale shows some sort of relation (via prediction or correlation) with existing measures/behaviours connected to the construct. For further information see [2] and [26].

2b) Context – practitioners must be aware of the context of use that the developers first envisaged when they constructed their scale. If practitioners use a scale in a context that was not originally envisaged by developers, then the validity (and reliability) of the scale could be compromised.

2c) Sample – similar to Context, this aspect asks practitioners to be aware of the sample (i.e., users) that developers used when constructing the scale. If, in the opinion of the practitioner, this is distinctly different from the sample that s/he is intending to recruit, then this could compromise scale validity and reliability.

3) Scale reliability – the extent the tool provides stable results across repeated administrations.

3a) Inter-item reliability – Measured via Cronbach's alpha or split-half reliability, it determines what extent scale items measure the same construct that it is said to be measuring accurately. Generally, Cronbach's alpha of .7 is a recommended minimum, though higher Cronbach's alpha are said to be necessary depending on the construct that is said to be measured.

3b) Test-retest reliability – Much like Inter-item reliability, this aspect tests how reliable the scale is when individuals are retested with the same scale. The

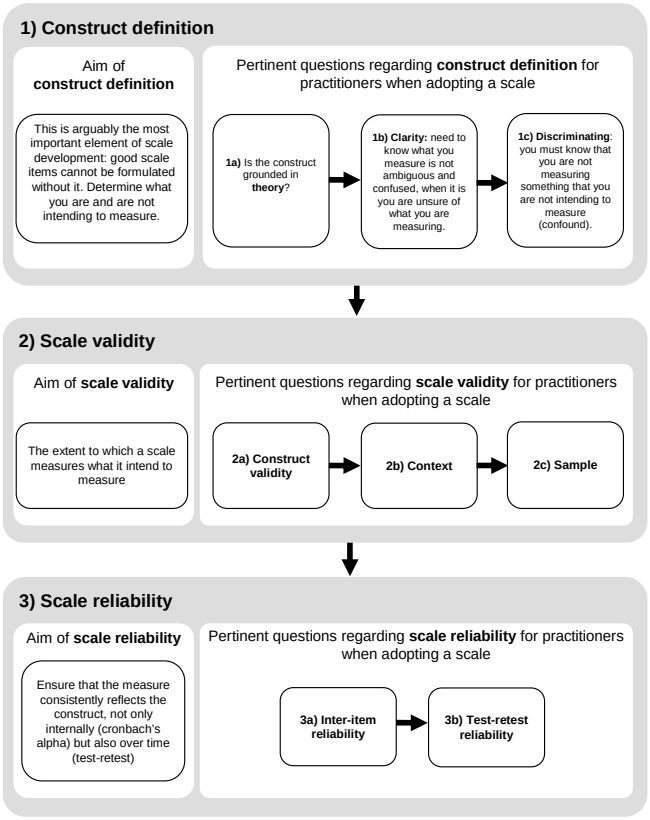

Figure 1: Scale Adoption Framework for Evaluation (SAFE)

level of test-retest reliability depends on what is said to be measured, but again, a good rule of thumb is .7.

Example of the SAFE usage

To initially ascertain whether or not the SAFE would support usability practitioners, the authors of this paper utilised the SAFE to review four established psychometric scales (Table 1). For this, the original publications of these psychometric scales were obtained. The scales have all been referenced in relation to their use in usability evaluations and are purposively concerned with disparate constructs that are all related to usability, satisfaction, and UX. In Table 1 the four psychometric scales and their adherence to the attributes of SAFE are tabulated.

As seen in Table 1, the four scales do not adhere to all of the SAFE attributes. Nonetheless, all of these scales are established and have been used to measure their respective constructs in publications. It is noted then, that it is not always possible or necessary to create scales that fully meet the SAFE requirements discussed above. This decision is down to the practitioner and whether or not the scale can be accepted. To determine how the use of SAFE can support practitioners in selecting appropriate scales, several case studies conducted in an industrial setting are now discussed.

Determining the affective response of participants: feedback from case studies

In many consumer application studies, a key question often addressed is, what is the affective response of participants when using the product being tested? In several projects, ranging from studies investigating users' affective response to environments that present coloured lighting effects [9, 16], to game applications [28], different measures to obtain participants' affective state were used: Self Assessment Manikin [17], Activation-Deactivation (AD-ACL, [27]), pleasure and arousal as evoked by environmental factors [24], and intrinsic motivation inventory (IMI [29]). These instruments are all claimed to be established measures. It is important to realise though that these instruments, as well as other affect scales that are often used in the context of UX studies, were never developed with possible application of usability testing for consumer applications in mind. These scales originate, for example, from research in the field of organizational psychology (e.g., IMI), or clinical psychology (e.g., AD-ACL).

Unfortunately, alternative measures do not seem to be readily available. Other means such as psycho-physiological measures have many drawbacks, especially in the context of testing consumer applications. So, it seems inevitable to adopt these scales, such as the SAM, which is being used extensively in usability testing. Without going into a detailed description of the above mentioned studies, this section provides a number of discussion points with respect to the use of these scales, and some lessons learned.

Attribute of SAFE		Scale and adherence to SAFE			
		Pleasure, arousal & dominance [17]	**Interpersonal trust** [29]	**Aesthetics** [18]	**Usability satisfaction** [19]
1) Construct	a) Theory	Yes	Yes	Yes	Yes
	b) Clarity	Yes: multi-dimensions	Yes: uni-dimentional	Yes: bi-dimensional	Yes: multi-dimensions
	c) Discriminat'	Yes	Yes	Yes	Yes and No
2) Validity	a) Construct validity	Yes	Yes	Yes	Yes
	b) Context	Yes	Yes	Yes	Designed for scenario based evaluation
	c) Sample	N=78: 45 Male	N=222	N=384: 211 Male	N=377
3) Reliability	a) Inter-item reliability	No	Split-half reliability: 0.92	Between 0.60 & 0.78	Exceeding 0.89
	b) Test-retest	No	No	Yes	No

Table 1: Example use of the SAFE[1]

- A general issue with several of the scales used is that quite often the scores obtained for different conditions do not show any significant differences. Obviously, a range of causes might have contributed; the conditions not being distinct enough, too many variables in play, selection and number of participants, or the scales not being sensitive enough. The first two possibilities are related to the fact that the studies in question are conducted in an ecologically valid setting, resulting in conditions that are not too extreme, and relatively rich in features. After all, in an industry context, testing applications in a realistic setting is very important.
- One could also wonder if these measures are effective in typical usability test conditions that often involve a relatively limited number of participants.
- The SAM consists of three subscales, pleasure, arousal, and dominance. The last subscale, however, is quite often not well understood by participants, even when provided with the official instructions as recommended [17]. A growing number of researchers have decided to simply no longer include the dominance subscale in their tests. This raises questions regarding the underlying construct of the scale, or at least questions with regard to applicability of part of the construct in a usability test.
- Valence measures (i.e., pleasure scales) quite often show ceiling effects; obtained scores tend towards the positive end of the scale. In many cases, such strong positive scores are already obtained in connection with a baseline condition. That is, even before the participants experienced the device or condition under investigation, they are in a positive state, and this continues for the test, unless something dramatic happens to impact mood.
- Finally, self-report measures are associated with a number of well-known issues:
  - Participants' answers might be biased or guided by what the participants think is the "right" answer, or by

[1] Table will be illustrated further in the VUUM workshop.

socially desirable answers. Furthermore, the discussed self-reports are retrospective and, thus, potentially subject to distortions.

- In the case of SAM, such self-report issues might be strengthened because it is a measure that involves only three subscales, making it easy to remember what one filled in before the test condition, for example.

DISCUSSION

The aim of the SAFE is to support the selection of psychometric scales for usability and UX evaluations. The development of the SAFE has built on previous tools to support practitioners in selecting and adopting psychometric scales. For example, [11]'s proposed model that outlined six approaches to measuring usability. The SAFE has been developed to support the selection of psychometric scales for (subjective) usability satisfaction and UX, supporting [12]'s and [15]'s position recommending that practitioners should use established scales. The example use of the SAFE in this paper considered existing scales related to the broad notion of (subjective) usability satisfaction and UX. A number of discussion points emerged.

First, it became apparent that gathering the information to complete the SAFE was not as easy as initially anticipated. The main reason for this was that scales are developed in many different ways, (e.g., using different formats and statistical tests). This diversity is also reflected in the language used to describe scale development. Translating the original publications of the scales to complete the SAFE required an understanding of the development of psychometric scales. It is not anticipated that all practitioners wishing to conduct an evaluation will have this required expertise or time. It is proposed that developers are more transparent in their scale description when publishing their developed scales; [19] is a good example to follow.

Secondly, the majority of usability scales were developed in the 1990s, between 10 and 20 years ago. There are potential issues surrounding measures of this age, which practitioners should consider. For example, the majority of usability scales were developed for workplace systems where tasks are structured and efficiency is of primary importance. When choosing measures to evaluate systems for the home, these factors may not be as prominent in design.

In addition, the relative age of these scales may affect validity, considering the type of systems that would have been used when these scales were developed. Therefore, the practitioner should consider that the original context of scale development may have an effect on the reliability and validity versus today. Thus it is suggested that practitioners should at least acknowledge this and, if there is no time to develop another scale, be aware of the potential problems that could arise when reviewing findings.

Thirdly, to the knowledge of this paper's authors, there is no authority textbook to support the selection of psychometric measures specifically for UCD and UX. Given the abstract constructs that practitioners will be required to measure in UX, a textbook to support and explain to practitioners the advantages of using psychometric measures, and how they are developed, would perhaps increase the validity and reliability of UX scales utilised by practitioners, regardless of whether or not these scales are homegrown. It should, however, be noted that there are some good texts related to the development of scales for the HCI domain. For example, [15] describe a 'basic lifecycle of a questionnaire' and call for a 'battery' of scales to be developed for HCI. They describe the need to achieve validity and reliability, but not how this is done. Another important focus is the importance of language and choice of scale, amongst others. The SAFE builds on this work by supporting the choice of appropriate scales.

It seems that the call for a battery of psychometric measures [15] has been met and the HCI community is now entering a second generation of evaluation methods. Whilst the community has a battery of scales to support evaluation and design, it is clear that the context for usage is now being stretched. It is hoped that the SAFE will iterate to practitioners that just utilising a psychometric scale does not determine if constructs being measured are valid in every situation, and that using a psychometric scale does not guarantee significant findings. In other words, a psychometric scale can be the wrong measure for evaluating a system. Using the SAFE should encourage practitioners to question appropriate use of scales. Nonetheless, the example use of the SAFE to select a psychometric scale in the UX example SAM [17], among others, illustrates the need for new scale development.

Please note that this paper does not discuss scale norms. Psychometric scale norms are not reported in many UX studies. Norms would provide the researcher with the ability to compare findings with other studies (assuming the studies were sufficiently controlled) in the same way that personality and intelligence of individuals can be compared.

CONCLUSIONS

The preliminary development and example use of the SAFE, a tool to support practitioners in selecting appropriate psychometric scales, was illustrated in this paper. Two conclusions have emerged, these are:

- It can be daunting and difficult to gauge whether or not a scale is based on a robust construct, and if the scale is valid and reliable. This should be clearer in future publications of scales.
- New scales are required to measure new constructs and overcome ceiling effects due to the new emphasis on augmenting existing and enjoyable experiences.

These challenges are pertinent to the VUUM workshop, and can be elaborated on for presentation and discussion.

The SAFE has been developed to support practitioners in adopting established psychometric measures. Nonetheless,

the SAFE model still has to be evaluated by practitioners. We intend to build on the SAFE and develop it into a useful and usable design tool, via evaluation with practitioners.

ACKNOWLEDGMENTS

We thank Jan Engel, Nele Van den Ende, and Janneke Verhaegh for lending ideas and comments in support of this research. We thank Marie Curie for funding William Green's research (SIFT project: FP6-14360) and Greg Dunn's research (CONTACT project: FP6-008201).

REFERENCES

1. Blythe, M. A., Overbeeke, K., Monk, A. F., & Wright P. C. (Eds.). Funology: From usability to enjoyment. Dordrecht, The Netherlands: Kluwer Academic, 2003.
2. Clark, L.A., Watson, D. Constructing validity: basic issues in objective scale development. *Psychological assessment*, 7, 3 (1995), 309-319.
3. Cockton, G. Revisiting usability's three key principles. In *Proc. CHI 2008:alt.CHI.* Florence, Italy, ACM, (2008), 2473-2484.
4. Cohen R. J. Swerdlik, M. E., and Philips, S. M. *Psychological testing and assessment. An introduction to tests and measurement.* Mayfield, 1996.
5. Gould, J. D. and Lewis, C. Designing for usability: key principles and what designers think. *Communications of the ACM* 28, 3 (March 1985), 300-311.
6. Greenberg, S. and Buxton, B. Usability evaluation considered harmful (some of the time). *Proc. CHI 2008*. Florence, Italy, ACM, (2008), 111-120.
7. Helander, M.G. Landauer, T.K. and Prabhu, P. (eds.). Handbook of Human-Computer Interaction, 2nd Edition. Amsterdam, North-Holland, 1997.
8. Helander, M.G. Khalid, H.K., and Tham, (Eds.). *Proc of the Int. Conf. on Affective Human Factors Design*. London: Asean Academic Press, 2001.
9. Hoonhout, H.C.M., Human Factors in Matrix LED illumination. In: Aarts, E., Diederiks, E. (eds.). *Ambient Lifestyle*. Amsterdam: BIS Publishers, 2006.
10. Hoonhout, H.C.M. How was the experience for you just now? Inquiring about people's affective product judgments. In: Westerink, J., Ouwerkerk, M., Overbeek, T., Pasveer, F. & de Ruyter, B. (Eds.). *Probing Experiences*, Springer, Dordrecht, NL, 2008.
11. Hornbæk, K. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64, 2* (2006), 79-102.
12. Hornbæk, K. and Law, E.L. Meta-analysis of correlations among usability measures. In *Proc. CHI 2007*, ACM Press (2007), 617-626.
13. ISO 9241. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. *ISO 9241-11*, Geneva, CH, 1998.
14. Isomursu, M., Tähti, M., Väinämö, S., and Kuutti, K. 2007. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human-Computer Studies.* 65, 4 (2007).
15. Kirakowski, J, and Corbett, M, *Effective Methodology for the Study of HCI.* Amsterdam: North-Holland, 1990.
16. Kohne, R. Subjective, behavioral and psychophysiological responses to lighting induced emotional shopping. Unpub master thesis. Utrecht / Eindhoven: Universiteit Utrecht / Philips Research, 2008.
17. Lang, P. J. Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems*. Norwood, NJ: Ablex, (1980), 119-137.
18. Lavie, T., and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 3 (2004), 269-298.
19. Lewis, J. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7, 1 (1995), 57-78.
20. Norman, D.A. *The psychology of everyday things*. New York: Basic Books, 1988.
21. Nunnaly, J. C. *Psychometric theory.* New York: McGraw-Hill, 1978.
22. Oatley, K, & Jenkins, J. M. *Understanding Emotions.* Cambridge MA: Blackwell Publishers, 2006.
23. Pickard, R. *Affective Computing.* Cambridge, MA: MIT Press 1998.
24. Russell, J.A., and Pratt, G. A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology.* 38 (1981), 311-322.
25. Ryan, R.M. Intrinsic Motivation Inventory: http://www.psych.rochester.edu/SDT/measures/intrins.html. Accessed on April 13 2008.
26. Spector, P.E. *Summated rating scale construction. An Introduction.* Newbury Park: Sage, 1992.
27. Thayer, R.E. Measurement of activation through self-report. Psychological reports, 20 (1967), 663-678.
28. Verhaegh, J., Fontijn, W., and Hoonhout, J. TagTiles: optimal challenge in educational electronics. Proc. of the Tangible and Embedded Interaction conference, Baton Rouge, LA, ACM, 2007.
29. Wheeless, L. and Grotz, J. The Measurement of Trust and Its Relationship to Self-Disclosure. *Human Communication Research*, 3, 3 (1977), 250–257.