Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

# Website Affective Evaluation: Analysis of Differences in Evaluations Result by Data Population

[1]Anitawati, M.L., [2] Afdallyna, F.H., [3] Nor Laila, M. N., [4]Nagamachi, M.

[1, 2]Faculty of Information Technology and Quantitative Sciences, University Teknologi MARA, 40450 Shah Alam, Malaysia
[1,3] {anita, norlaila} [at]tmsk.uitm.edu.my
[2] afdallyna.f.harun@gmail.com

[4]International Kansei Design Institute, Japan

**Abstract.** Studies involving consumer studies have suggested different mechanisms of subject selections. The paper elaborates results of subject's responses by the methodology adopted from Kansei Engineering. In the research, evaluation of subject's Kansei towards website interface design was performed, targeting to measure affective quality in website design. Principal Component Analysis was performed to identify semantic structure of Kansei Words. The analyses were based on the average of evaluation results obtained from subjects. Results of PC Loadings were analyzed to see differences of determinants by size of data population. It is evident from the study that population size does not affect determinants of affective web interface design. The study makes decent contribution in determining appropriate population size in designing research instruments for future studies involving website affective evaluations.

**Keywords.** Consumer science, website affective evaluation, Kansei, Population size, Principal Component Analysis

## 1 Introduction

Conducting a research that enables an insight into user behaviour from social standpoint can give us a lot of data to work with. However, given the nature of web development and testing, having to conduct experiments can be costly. A major part of this expense is the participant costs. Hence, it is desirable to reduce the number of participants without sacrificing the quality of the experiment. There would be a significant savings if there is a possibility of using smaller participant pool and yet get the same results as the entire pool. Therefore, for this research, we have conducted experiments with different amount of participant pool to see if the smallest pool would yield the same results as the entire population (biggest experiment population possible).

The context of web application chosen for this work is the design of online clothing websites where affective quality is assumed to be significant. Based on the

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

result, we discuss the differences in Kansei space, concluded to determinants, as output from different size of data population.

## 2 Subject Selection

Determining subjects for a study can be problematic as studies should be designed to fit the study's goals. Careful judgment need to be in place as errors in determining the number and type of participants or even the number of runs can be costly.

As this research concerns with web design, we have made reference to several usability studies theories due to its common nature of web testing. In usability studies, there is a common debate on the number of participants one must have when conducting experiments.

**i. Five-User Assumption**

"The best users come from testing no more than 5 users and running as many small tests as you can afford" [1]. Nielsen [1] elaborated that one user should be able to uncover a third of the findings and as more users are added, information redundancy occurs.

**ii. Five Users And Beyond**

According to Gilbert et al. [2], one study was done where five users were randomly chosen and only uncovered 35% of the findings, while the $13^{th}$ and $18^{th}$ user uncovered data that the original users missed. This result shows that if the study had been discontinued at five, those data would have been overlooked. In the same study, users $6^{th}$ to $18^{th}$ were able to find other new data that the original five were unable to find. This shows that, if the right users are not chosen, pertinent data can be left out [3]. Nonetheless, many has interpreted Nielsen's recommendation wrongly as Nielsen has highlighted that one should run as many tests as one can afford until the findings meet an "acceptable level".

Landesman & Perfetti [4] have conducted user testing on an e-commerce site using the recommendation "test four to five users with no more than eight". They have found this technique only yielded 35% of the problems in the system which would in return require 90 more tests runs to uncover the 600 problems in the system. From their study, it is learnt that in web testing, one need to apply a concept that fit the studies goals and needs. It must be noted that e-commerce websites have complex content which continuously and incrementally changes. Furthermore, there is a variety of e-commerce website users which implies that one sample group could not be used as a representation of the whole because each user who interacted with the system used the system differently.

Therefore, base on the reviews conducted, this study has used random sampling in selecting users and grouping them into a pool of 30, 60, 90 and 120 users. Random sampling is a chosen method for participant selection as it contains no bias and can be relatively representative of the targeted population [5]. It allows researchers to make

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

generalizations and justification about the majority of the population by a certain level of certainty [6].

## 3 Affective Website Interface Design

HCI issues related to website applications were formerly focused on cognitive aspects of websites. Since the early work of Nielsen in the 1990s, the emphasis was on the qualities of usefulness and usability in producing good website design. Li & Zhang [7] cited that most studies dedicated to e-Commerce website evaluation are based on two assumptions; (i) target customers spend at least a few minutes on a website and (ii) good website features usually elicit positive cognitive evaluations and shopping experience. These assumptions have ignored the primary affective reaction or primary emotional responses towards the website. They stressed that online shopping behavior is a complex phenomena and recognized that affective reaction is one factor that promotes online shopping. This is because e-commerce websites have gone beyond the function of conveying information to the extent of providing persuasive engagement with website visitors through the lively process of perception, judgment and action. Affect has also been discussed in literatures as a factor found to influence decision-making, perception, attention, performance and cognition [8] [9].

Align with these claims, we argue that e-Commerce websites should induce desirable consumer experience and emotion that influences users' perception of the websites, to enhance visitor's stickiness that promotes consumer conversions and retentions.

Despite the gained recognition, the emotional appeal of websites is often neglected as designers tend to pay more attention to issues of usefulness and usability [10] due to the availability of established design methodology addressing usefulness and usability. Design method that incorporates emotional design requirements is lacking. In addition, numerous studies conducted on emotional design tends to look at minimizing irrelevant emotions related to usability such as confusion, anger, anxiety and frustration [9]. Therefore, it is necessary to seek for a suitable design method to handle design requirements based on emotional signatures of websites.

## 4 Data Population

The principles and empirical findings of behavioral science are probabilistic in nature, whereby it is possible to describe the reactions of most individuals but there is also a need to recognize that not everyone will fit the general pattern [11]. Nonetheless, the influence of a population size on results validity is critical. The common belief has always been the larger the sample, the greater the statistical power.

The solution to the dilemma of number of participants lies in ensuring that the analysis can be placed in a structural context and the research objective. The key to managing it relies on the researcher's conscious self-understanding of the research process [12]. As Ward-Schofield [13] has suggested:

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

*"...Assumption of a qualitative research is very much influenced by the researcher's individual attributes and perspectives. The goal is not to produce a standardised set of result s[…]. Rather it is to produce a coherent and illuminating description of and perspective on a situation that is based on and consistent with detailed study of the situation."*

We attempted to determine and understand the effect of number of participants to the population generalization. A design method computing analysis of 30, 60, 90 and 120 participants were put into place to view its statistic generalisation effect.

## 5 Kansei Engineering

Kansei Engineering (KE) is a technology that combines Kansei and the engineering realms to assimilate human Kansei into product design with the target of producing products that consumer will enjoy and be satisfied with. The focus of KE is to identify the Kansei value of products that trigger and mediate emotional response. The KE process implements different techniques to link product emotions with product properties. In the process, the chosen product domain is mapped from both a semantic and physical perspective. In terms of a design methodology, the approach of KE is to organize design requirements around the emotions that embody users' expectations and interaction [14], [15], [16]. KE has been successfully used to incorporate the emotional appeal in the product design ranging from physical consumer products to IT artifacts. Due to its success in making the connection between designers and consumers of products, KE is a well accepted industrial design method in Japan and Korea. In Europe KE is gaining acceptance but is better known as emotional design.

## 6 Research Method

The following fig. 1 illustrates the research method.

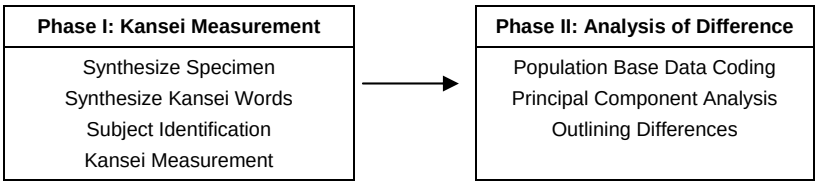


**Fig. 1.** Research method.

The research involves two phases; Phase I: Kansei Measurement, Phase II: Analysis of Difference, as shown in Figure 1. In Phase I, we adopted KE methodology to quantify website visitor's Kansei responses. Result form Phase I is then analysed statistically using Principal Component Analysis (PCA) to scrutinize the Kansei structure. The result enables the paper to conclude differences in Kansei

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

structure by data population based on 30, 60, 90 and 120 participants. Details to the phases are described in the following sections (7 and 8).

## 7 Kansei Measurement

Phase I begins with selection of specific domain. It is important to control the domain and subjects as different domain will induce different Kansei. Specific target market group must be used as experiments subject, so that the intended Kansei could be measured accurately. Failing which will lead to confusion during Kansei measurement and yield invalid result. The context of web application chosen for this work is the design of e-Clothing websites where emotional appeal is assumed to be significant. Correspondingly, the selected subjects are consumer with online shopping experience. Then, the study proceeds with synthesizing specimen, synthesizing Kansei Words, Subjects Identification, and Kansei Measurement.

### 7.1 Synthesize Specimen

Initially, 163 online clothing websites were selected based on their visible differences in design (i.e : colors, layouts, typography). An investigation was conducted to identify detail design elements in all websites in the context of what consumer's see in the interface feature of a website. As a result, the study has identified 77 categories in design element, and 249 items as specified values in each design category identifiable from all websites.

All websites were then analyzed following a set of predefined rules in the study. From the analysis, 35 website specimens were finally used.

### 7.2 Synthesizing Kansei Words

Since Kansei is the state of consumer's internal sensation, the measurement process can be very challenging. In the measurement of visitor's Kansei in e-Commerce website, measurements are psychological which deals with human emotional state. Hence, the most suitable measurement method is by self-reporting system. This is done by using words that describe the emotional expression associated to e-Commerce website. In KE, this expression is called Kansei Word (KW) [15].

In the study, KWs are used to represent emotional responses and were synthesized based on web design guidebook, experts and pertinent literatures. 40 Kansei Words were then selected according to their suitability to describe website. Among the synthesized words are 'adorable', 'professional' and 'impressive'. These KWs were used to developed checklist to rate websites, organized in a 5-point Semantic Differential (SD) scale.

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

### 7.3 Subjects Identification

Deciding on the number of participants was influenced with "What information we intend to capture?" As we intend to understand the pattern of experience economy while shopping online, we have selected male and female participants ranging from the age of 20 to 25. This age group is selected as they are the second most common consumers of online shopping [17]. Furthermore, this age group is readily available in UiTM. An equal distribution of male and female participants was acquired base on Freeman et al. [18], a research concerning participant behaviours and emotions. Furthermore, Horrigan [17] has also found that there is almost an equal distribution of male and female users within the age group of 20 to 25. We have also considered the length of time required to gather and analyse information systematically due to minimal resources of labour, time and money [19].

120 undergraduate students from the Faculty of Information Technology and Quantitative Science, Faculty of Architecture, Building, Planning and Survey, Faculty of Business and Management and Faculty of Electrical Engineering from UiTM participated in the Kansei evaluation. Exactly 30 students consisting of 15 males and 15 females were recruited from each faculty. All of them have Web experience.

### 7.4 Kansei Measurement

The participants were grouped according to their faculties. Four Kansei evaluation sessions were held separately for each group. During each session a briefing was given before the participants began their evaluation exercise. The 35 website specimens were shown one by one in a large white screen to all participants in a systematic and controlled manner. Participants were asked to rate their feelings into the checklist according to the given scale within 3 minutes for each specimen. They were given a break after the $15^{th}$ website specimen, to refresh their minds. The order of checklist was also change to avoid bias. Each Kansei evaluation session took approximately 2 hours to complete.

## 8 Analysis of Difference

We analyzed the website semantic space by PCA using the averaged evaluation value for each session. In Phase II, we coded the averaged data according to population size of 30, 60, 90 and 120. This is to organize information into set of orders where Kansei semantic space can be observed. PC loadings results show the degree of Kansei affecting variables which are used to obtain KWs structure. Figure 2 shows the distribution of all data population sizes while Table 1 helps to summarize the analysis.

In Figure 2, we can observe a good distribution of variables to axis-x and axis-y, which proves that the measurement was successful. It is evident from the plot results for all population sizes that the KW that produced large negative first PC loadings (x-axis) are mostly "Beautiful", "Gorgeous", "Stylish", "Impressive" and "Appealing". The dense area of the left hand side of the chart corresponds to such KW. On the

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

other hand, KW that produced large positive PC loadings are “Boring” and “Old-fashioned”. Thus, we label this PC as the axis of “Attractiveness”.

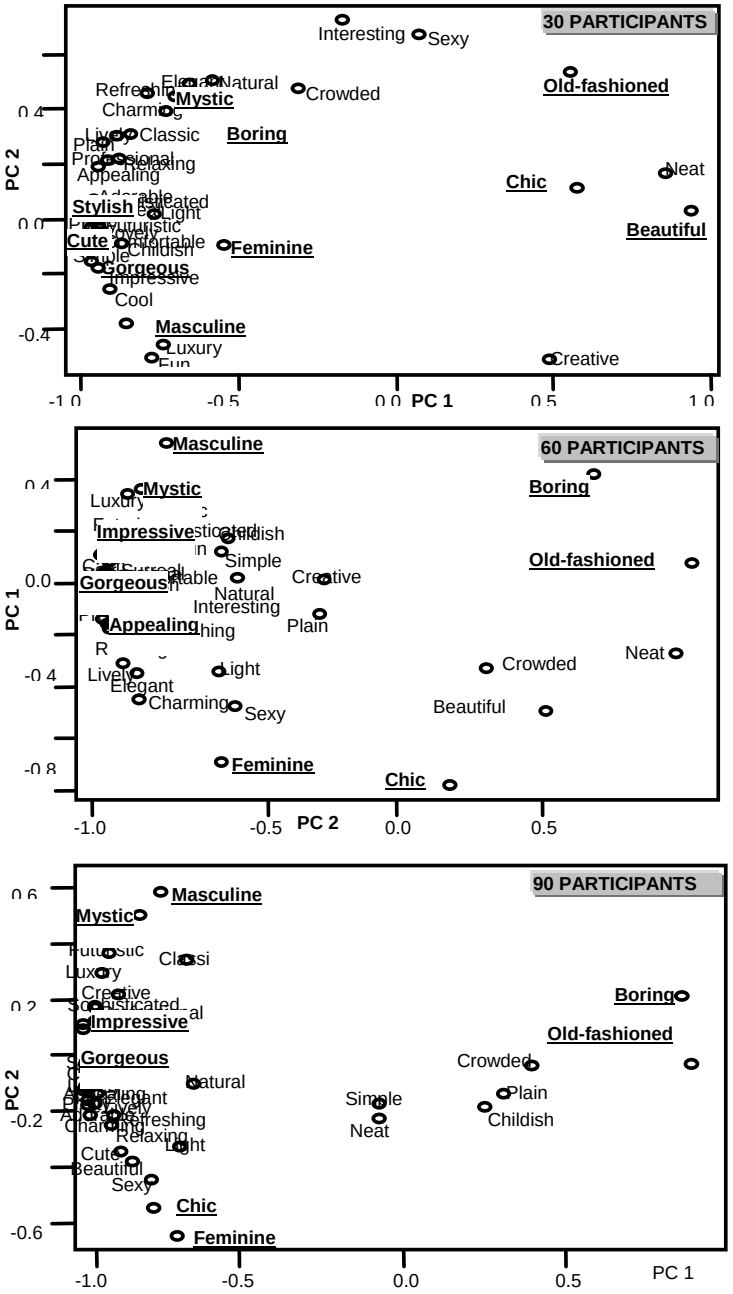


**Fig. 2.** PC Loadings by 30, 60, 90 and 120 data population.

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

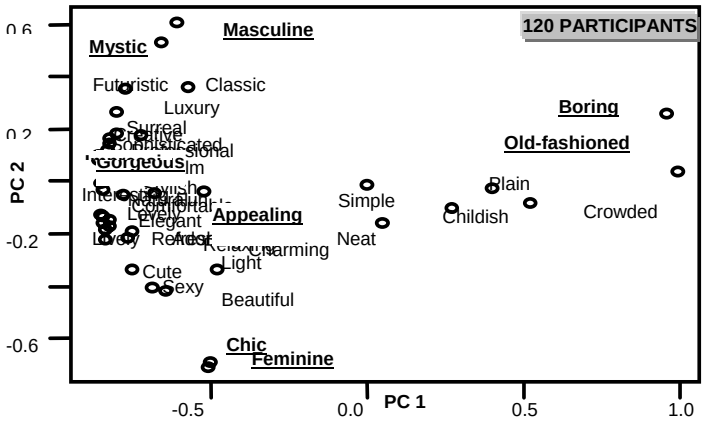


**Fig. 2.** cont.,

From the result, we can expect that websites with lower scores on this component are likely to have higher sense of attraction and conversely. The second PC loadings (y-axis) shows that KW with positive large loadings are "Masculine" and "Mystic"; the negative are "Cute", "Feminine", and "Chic". Thus, we label this PC as the axis of "Masculine-Feminine". We can expect that websites with high scores on this component will tend to have high characteristic of masculinity and conversely.

In Table 1, we see a contribution ratio of over 70% indicating that the first two principal components represent the total variability. Thus, most of the data structure can be captured in two underlying dimensions. This means, the KW structure are highly influenced by the first two principle components. The remaining principal components account for a very small proportion of the variability and can be ignored. Table 1 also shows that the Kansei structure on website design has two components, which are attractiveness and masculine-feminine. Blending and balancing these two components are determinants of affective website design. Furthermore, all groups of data population suggest same determinants in designing affective website. The difference in number of subjects seems to produce similar Kansei structures.

**Table 1.** PC Loadings Results.

| **Population Size** | **Contribution Ratio** | **1st PC Loadings (x-axis)** | | **Axis Label** |
|---|---|---|---|---|
| | | **Large Negative** | **Large Positive** | |
| **30** | 73.0% | Beautiful, Gorgeous, Stylish | Boring, Old-fashioned | Attractiveness |
| | | **2nd PC Loadings (y-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| | | Cute, Feminine, Chic | Masculine, Mystic | Masculine-Feminine |
| **Population Size** | **Contribution Ratio** | **1st PC Loadings (x-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| **60** | 75.0% | Beautiful, Gorgeous, Stylish | Boring, Old-fashioned | Attractiveness |
| | | **2nd PC Loadings (y-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| | | Cute, Feminine, Chic | Masculine, Mystic | Masculine-Feminine |

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

**Table 1:** cont.,

| Population Size | Contribution Ratio | 1st PC Loadings (x-axis) | | Axis Label |
|---|---|---|---|---|
| | | **Large Negative** | **Large Positive** | |
| **60** | 75.0% | Beautiful, Gorgeous, Stylish | Boring, Old-fashioned | Attractiveness |
| | | **2nd PC Loadings (y-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| | | Cute, Feminine, Chic | Masculine, Mystic | Masculine-Feminine |
| **Population Size** | **Contribution Ratio** | **1st PC Loadings (x-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| **90** | 78.7% | Impressive, Gorgeous, Appealing | Boring, Old-fashioned | Attractiveness |
| | | **2nd PC Loadings (y-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| | | Feminine, Chic | Masculine, Mystic | Masculine-Feminine |
| **Population Size** | **Contribution Ratio** | **1st PC Loadings (x-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| **120** | 77.9% | Gorgeous, Impressive, Appealing | Boring, Old-fashioned | Attractiveness |
| | | **2nd PC Loadings (y-axis)** | | **Axis Label** |
| | | **Large Negative** | **Large Positive** | |
| | | Feminine, Chic | Masculine, Mystic | Masculine-Feminine |

## 9 Conclusion

The study was performed to identify differences in Kansei structure by size of data population. The result shows that Kansei structure from different data populations are similar. With the result, we could conclude that the size of data population used in the research instruments does not affect the result of determinants in affective web design. We have shown that to first begin testing, one need to understand the type of product one intends to test, who are the users and what work with the system. Once criterions of data population are set suitable to the research methods and objectives, one can make accurate estimates about the population based on what they have learned from the sample, henceforth concluding a more accurate generalization.

**Acknowledgement.** The research is supported by grants from the Ministry of Science, Technology and Innovation, Malaysia, under the ScienceFund grant scheme [Project Code: 01-01-01-SF0029].

## References

[1] Nielsen, J. (2000) *Why You Only Need to Test With 5 Users*. Retrieved February 2, 2007 from Alertbox Website: http://www.useit.com/alertbox/20000319.html.

Lokman, A.M., Harun, A.F.H., Noor, N.M., Nagamachi, M. (2009). Website Affective Evaluation: Analysis Of Differences In Evaluations Result By Data Population. Human Centered Design. Series: Lecture Notes In Computer Science. Subseries: Information Systems And Applications, Incl. Internet/Web, And HCI. Vol. 5619. Kurosu, Masaaki (Ed.). ISBN: 978-3-642-02805-2.

[2] Gilbert, J. E., Williams, A. and Seals, C. D. (2007). *Clustering for Usability Participant Selection*. Journal of Usability Studies, 3(1), pp. 40-52

[3] Faulkner, L. (2003) *Beyond The Five-User Assumption: Benefits Of Increased Sample Sizes In Usability Testing*. Behavior Research Methods, Instruments, & Computers, 35(3), pp. 379-383.

[4] Landesman, L. & Perfetti, C. (2001) *Eight is not Enough*. Retrieved 30 September 2008 from User Interface Engineering Website: http://www.uie.com/articles/eight_is_not_enough/

[5] *Arteology: Sampling.* Retrieved 1 September 2008 from www.uiah.fi/projects/metodi/152.htm.

[6] Rosenstein, A. (2001) *Managing Risk with Usability Testing, Classic System Solutions*. Retrieved 1 September 2008 from CLASSIC Website: http://www.classicsys.com/css06/cfm/article.cfm?articleid=23

[7] Li, Na and Ping Zhang. Consumer Online Shopping Behavior, in Customer Relationship Management. Fjermestad, J. and Romano, N. (eds), Series of Advances in Management Information Systems, Zwass, V. (editor-in-chief), M.E. Sharpe Publisher, 2005.

[8] Tractinsky, N, Katz, A.S. and Ikar, D. What is Beautiful is Usable. Interacting with Computers, 13: . 127-145, 2000

[9] Norman, D. A. Emotional Design: Attractive Things Work Better in Interactions: New Visions of Human-Computer Interaction, IX: pp. 36-42, 2002

[10] Buchanan, R. Good Design in the Digital Age. AIGA Journal of Design for the Network Economy, 1(1), American Institute of Graphic Arts: pp: 1-5, 2000

[11] Leary, M. R. (2008). Introduction to Behavioural Research Methods (4th Ed). Pearson

[12] Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in practice*. London: Tavistock Publications.

[13] Ward-Schofield, J. (1993). Increasing the generalisability of qualitative research. In M. Hammersley (Ed.), *Social research: Philosophy, politics & practice* (pp. 200-225). London: Open University/Sage.

[14] Anitawati, M. L. and Nor Laila, M.N. Kansei Engineering: A Study on Perception of Online Clothing Websites. Proceedings of the 10th International Conference on Quality Management and Operation Development 2008 (QMOD '07), Sweden: Linköping University Electronic Press, ISSN 1650-3740, 2007

[15] Nagamachi, M. The Story of Kansei Engineering (in Japanese), Japanese Standards Association, 6. Tokyo, 2003.

[16] Nagasawa, S. Present State of Kansei Engineering in Japan. 2004 IEEE International Conference. Vol.1. pp:333-338, 2004.

[17] Horrigan, J. B. (2008). Online Shopping. PEW Internet and American Life Project. Retrieved 20th August from Pewinternet Website: http://www.pewinternet.org/pdfs/PIP_Online%20Shopping.pdf

[18] Freeman, J, Lessiter1, J., Pugh1, K. and Keogh, E. (2005). "*When presence and emotion are related, and when they are not.* Proceedings of PRESENCE 2005, pp: 213-219.

[19] Freedman, R., Taub, S., Silver, J. F., Pell, E., Rowe, J. and Chiri, G. (2006). Sampling: A Practical Guide for Quality Management in Home & Community-Based Waiver Programs. Thomson & Medstat.