Developing an instrument to assess the impact of attitude and social norms on user selection of an interface design: a repertory grid approach

Willem-Paul Brinkman and Steve Love

Brunel University

Uxbridge, Middlesex
UB8 3PH
United Kingdom
{willem.brinkman, steve.love}@brunel.ac.uk

ABSTRACT

This paper presents a questionnaire instrument to evaluate designs of a mobile phone and a multimedia player. The study adopted a bottom up approach by interviewing 20 participants using Kelly's Repertory Grid Technique. This resulted in two sets of 200 personal constructs participants considered relevant when evaluating a set of 15 designs for each device. Two initial questionnaire instruments were developed and their validity was examined in a survey among 156 university students. In the interview, data was also collected about participants' attitude, social norm and their intention to select a design. This data was used to develop two statistical models. These models suggest that beliefs about the preference of participants social reference group, such as peers, had a small, but significant impact on the users' selection on the more publicly noticeable mobile phone, but failed to have a significant impact on the selection of designs for the more private multimedia player.

Keywords

Repertory grid analysis, instrument, evaluation, attitude, social norm, theory of reasoned action.

INTRODUCTION

Where traditionally devices mainly had a single look and feel to them, new technologies allow designers to create multiple interface designs for a single device. For example, mobile phones with different covers, and software applications with different user interface skins to change their appearance, e.g. Microsoft Multimedia player, ICQ, and Winamp. This design flexibility leaves designers with the obvious question: which set of user interfaces should they design? One approach is to match each design to a specific user group. This requires information about the desires of each user group, which can be translated into relevant design factors. Some studies have looked directly at the interface properties and link these with specific user groups. For example, Brinkman and Fine (2005) have studied the user interface skins of a multimedia player, and suggested links between the colour, themes and user personality traits. Others (e.g. Kim, Lee, and Choi, 2003) have studied which factors designers use to create homepages that trigger specific emotions in users. They concluded that designers look at things such as colour, texture and shape. These studies however, focus directly on design properties, instead of focusing how users evaluate a design, and what they consider as relevant criteria. This may include traditional ergonomic criteria such as safety and workload. However, users may also consider the social impact or status that comes with selecting a specific design. This therefore led us to set up a study with two aims: first, to develop an instrument to evaluate how users perceive a device, secondly, to examine the impact of social reference groups on users' evaluation. The study takes the position that selecting a design is partly a social act. For example, consumers not only enjoy an iPod for its music but also for it social status. According to the Theory of Reasoned Action (TRA) (Ajzen and Fishbein, 1980), social behaviour can be understood by looking at individuals' attitude and the social norm towards that particular behaviour. In the context of this study this means that the intention to select a design would be influenced by individuals' general feeling of favourableness or unfavourableness towards a design (attitude), together with their belief about what they should do according to people who are important to them (social norm). A possible underlying factor here is the visibility of the application by other people, which could influence the evaluation criteria themselves. Therefore, the study considered two applications: a mobile phone, which can easily be seen by others when used, and a multimedia player on a PC, a more private application. The following sections will discuss how two sets of Likert-scales were developed to evaluate the two applications. This section is followed by a section that analyses the relationships between attitude, social norm and behavioural intention of selecting a design.

REPERTORY GRID ANALYSIS

The study used the repertory grid technique to establish two sets of Likert-scales. This interview technique, that complements the Personal Construct Theory (Kelly, 1963), helps individuals to talk about the constructs they use to interpret the world, in this case the design of two products. The way the repertory grid technique works is

for the individual to compare and contrast ten sets of, for example, three significant people in their lives. These sets are known as triads, and could include people such as the individual's father, mother and brother. In this case, the individual would be asked to compare two of the people in this group in some way and contrast them with the third member of the triad. For example, an individual may indicate that the mother and brother are very easy-going whereas the father is more tense and uptight. This resulting construct easy-going versus tense, is regarded as one construct that the individual uses to interpret their social world and their own role in it. By comparing a number of people in the life of individual it becomes possible to establish a grid of constructs used by that individual. Instead of comparing people, the technique can also be used to analyse objects, events, concepts, etc. Because of its flexibility, the repertory grid technique has been used in a diverse variety of research areas on topics such market research (Hamlin, 2000), the construction of information space (McKnight, 2000) and change management (Brooks, Davis, and Lycett, 2004). It has also been applied to study perceptions interactive systems, such as websites (Tan & Tung, 2003).

Method

For the purposes of this study, participants were asked to complete a repertory grid in relation to mobile phone design and one related to a multimedia player. Twenty participants took part in this study. They were all either Brunel University students or members of the academic staff. There were 10 males and 10 females, with a mean age of 26.5 years (*SD* = 4.84).

Table 1: Phones and Multimedia-players skins.

Mobile phones	Skins
HSDPA, Samsung SGH-ZX20	Catwoman
BenQ-Siemens S88	Tiny Player
Samsung SGH-i300	Eye of Africa
OKWAP A236	Blue Martian
SPV C600	Ducky
Sony Ericsson J230	Winp Media Player
Sagem my301x	Xebish1-0
Sanyo A5520SA	Israeli
BlackBerry	Television
Nokia 7360 2	Back to the Future Trilogy
Netgear Skype Phone	Foo Fighters – One by One
Nokia 7600	PADD
Sanyo S750	wcnwms
O2 X2i	The Simpsons
Nokia's 9300	Halfskull

The design of the interview was counter-balanced with half of the participants completing the mobile phone repertory grid first followed by the multi-media player skin repertory grid and vice-versa. In the case of the mobile phone repertory grid, participants received 3 photos from the experimenter, each of a different mobile phone, and were asked to say which of the phones were similar and which was the odd one out. Next, they were asked to describe and put a label on the two groups of the triad, which resulted in a bipolar construct, such as immature – mature, complex – userfriendly, or ugly – appealing. The participants did this for 10 sets of 3 phone designs. A similar procedure was followed for the repertory grid for the multimedia players. These triads were randomly drawn from a set of 15 mobile phones, taken from the Mobile Digest news site (www.mobile-digest.com), and 15 skins selected from a previous experiment (Brinkman and Fine, 2005) based on their distinct user rating. After collecting the constructs, participants were asked to rate all the designs on the 10 constructs they had created. The phones and skins used in this study are shown in Table 1.

Results Grid Analysis

The main focus of the analysis was to see if the grids developed by each participant had some constructs in common. One possible way would be to study the semantics of the labels in the 20 grids. However, this approach is limited because individual participants might have verbalised the labels of the construct differently, while referring to a similar underlying construct. Therefore another more quantitative approach was applied. The assumption of this approach was that although participants might verbalise construct differently, if they refer to the same underlying construct, they would rate the designs in a similar way. For example, one participant mentioned the construct Femininity- Masculine and rated the phones on this scale as 6, 4, 2, 5, etc, while another participant mentioned the construct Female – Male and rated the phones on this scale as 7, 5, 2, 6, etc. Although the ratings are not exactly the same, a clear correlation in the rating patterns is visible. Factor analysis is a systematic method by which the correlations between all these ratings can be studied. It can establish a reduced number of components that accounts for the variance in the rating of the designs. If the grid constructs of participants representing the same underlying fundamental construct, these grid constructs should correlated highly with the same component of the factor analysis. Therefore, two factor analyses were conducted: one on the data from mobile phones grids and one data from the skin grids. Each factor analysis used the principal-component extraction method to establish the components from 200 constructs. The following sections will discuss the analysis factor loadings matrices after varimax rotation.

Mobile Phone Constructs

The aim of analysing the components was to identify common constructs themes that were used by multiple participants. Because the factor analysis was based on a relative small sample of 15 mobile phones designs compared with 200 constructs, a method of including all

components with an eigenvalue larger than 1 was inappropriate. Therefore instead, the following criteria were applied to select components:

- Factor loadings (correlations) below 0.69 were ignored.
- Components should have constructs loading from at least five different participants.
- A clear semantic relationship between the labels of the construct should exist.

Taking these criteria into account, the analysis produced three components (Table 2, Table 3 and Table 4). Examining Table 2 with the constructs loading on the first component, there seems to be a number of recurring themes, such as *gender*, *maturity* and *professionalism*. For example, for gender, participants mentioned constructs such as: girly – huge, girly – masculine, femininity – masculine, female – male, childish – manly, girly – standard, girly – manly, and feminine – masculine. Although these constructs are not all gender bi-polar, gender related constructs were explicitly mentioned by eight participants. Construct that related to maturity were: childish – classy, immature – mature, teen – professional, childish – sophisticated, childish – professional, playful – mature, babyish – sophisticated, childish – manly, childish – mature, and simple – mature. In total, ten participants had constructs with a reference to maturity that correlated on the first component. Finally, professionalism seems to be the related theme in the following constructs: fun – technical, teen – professional, unprofessional – professional, childish – professional, amateurish – professional, novelty – business, and toy – professional. As all these constructs loaded highly on one component, these themes seem to relate to a more general construct that can be described as the *appearance of the mobile phone*, i.e. the image or the look of the mobile phone.

Looking at Table 3, it seems that the constructs that loaded on the second component relate to the mobile phone's *ease of use*. Constructs that seems to suggest this are: confusing – appealing, complicated – simplistic stylish, chunky – simplistic, large – easy to use, comfortable to use – appealing, complex – attractive, not clear – clear, complex – simple, not understandable – understandable, uncomfortable in pocket – comfortable in pocket, unique – simple, and complicated looking – simple to use.

Finally from examining Table 4, it seems that *reliability* of the mobile phone could be a common construct. This time however the number of construct and participants that mentioned a related construct was smaller compared to the previous construct themes. Still, participants mentioned constructs such as: not easy to break – easy to break, protective of credit – non protective of credit, reliable – unreliable, concealing – open, high failure rate – low failure rate, and unsound – robust.

Table 2: Loading on component 1 of phones.

Ss	Label	Load
1	Girly - Huge	0.75
	Childish - Classy	0.85
	Delicate - Strong	0.78
	Confusing - Prestige	0.89
	Horrible - Nice	0.80
3	Immature - Mature	0.83
	Fun - Technical	0.97
4	Colourful - Plain	0.90
	Teen - Professional	0.86
5	Girly - Masculine	0.88
6	Embarrassed - Proud	0.92
	Unprofessional - Professional	0.93
	Lightness - Darkness	0.70
	Femininity - Masculine	0.91
	Ugly - Attractive	0.88
7	Childish - Sophisticated	0.95
	Fashionable - Plain	0.72
8	Childish - Professional	0.72
9	Playful - Mature	0.75
	Bright - Dull	0.75
10	Babyish - Sophisticated	0.84
	Fruity - Sleek	0.86
	Ugly - Gorgeous	0.80
	Novice technology - Advanced technology	0.78
	Dull - Impressive	0.78
	Female - Male	0.94
	Low capacity - High capacity	0.82
	Compact screen size - Spacious screen size	0.70
11	Colourful - Simple	0.84
	Childish - Manly	0.85
	Unique - Smart	0.74
	Girly - Standard	0.85
	Bright - Chunky	0.81
12	Childish - Mature	0.93
	Colourful - Plain	0.85
	Amateurish - Professional	0.92
	Simple - Mature	0.74
13	Less abilities - More abilities	0.79
	Colourful design - Mono colour	0.73
	Everyday use - Serious occasions	0.72
15	Colourful - Dull	0.78
	Female - Male	0.92
16	Childish - Sophisticated	0.82
	Garish colour - Neutral colour	0.87
	Limited functionality - Functional	0.77
	Girly - Manly	0.95
17	Feminine - Masculine	0.76
	Bright - Dark	0.73
	Colourful - Dull	0.70
18	Novelty - Business	0.89
	Feminine - Masculine	0.95
	Colourful - Plain	0.93
20	Funky colour - Serious colours	0.84
	Toy - Professional	0.77

Table 3: Loading on component 2 of phones.

Ss	Label	Load
1	Confusing - Appealing	0.77
2	Complicated - Simplistic stylish	0.91
	Chunky - Simplistic	0.94
	Large - Easy to use	0.92
	Uncomfortable to use - Appealing	0.94
	Complex - Attractive	0.92
	Not clear - Clear	0.94
	Gender biased - Gender neutral	0.70
3	Sophisticated - Plain	0.76
4	Complex - Simple	0.70
	Not understandable - Understandable	0.76
8	Uncomfortable in pocket - Comfortable in pocket	0.77
9	Unconventional - Conventional	0.81
15	Modern - Traditional	0.75
	High interactivity - Low interactivity	0.74
17	Modern - Traditional	0.81
	Unique - Simple	0.77
	Futuristic - Current	0.87
18	Fully functional - Minimalistic	0.79
	Futuristic - Modern	0.78
20	Complicated looking - Simple to use	0.86
	Hidden features - Fast access to options	0.90

Table 4: Loading on component 3 of phones.

Ss	Label	Load
8	Not easy to break - Easy to break	0.72
	Protective of credit - Non protective of credit	0.76
9	Cute - Not cute	0.80
12	Small - Big	0.86
14	Reliable - Unreliable	0.75
17	Compact - Expansive	0.79
19	Concealing - Open	0.83
	Rigid platform - High adaptability	0.70
	Simplistic - High functionality	0.73
	High failure rate - Low failure rate	0.91
	Unsound - Robust	0.75
	Standard functionality - Power functionality	0.71

On the basis of these findings, a set of scales was developed to measure these dimensions (or general construct themes). For the *ease of use* and the *reliability* construct at least five Likert scales were created, and for the *appearance* construct, four scales for *professionalism* and *maturity* theme were created and only three scales for the *gender* theme. Where possible, constructs mention by different participants were directly included. But in some cases the synonyms or labels from other constructs were used to make new bi-polar scale that corresponded with the dimension. For example, the construct playful – mature, was changed to playful – serious (M3), because the label mature was already used (M1). The following list presents the scales for the mobile phone:

Maturity

M1 Immature – Mature
M2 Childish – Sophisticated
M3 Playful – Serious
M4 Silly - Classy

Professionalism

P1 Unprofessional – Professional
P2 Fun – Technical
P3 Novelty – Business
P4 Amateur - Expert

Ease of use

E1 Complicated - Simple to use
E2 Difficult – Plain
E3 Hard to use - Easy to use
E4 Complex – Simple
E5 Difficult to carry - Easy to carry

Gender

G1 Feminine – Masculine
G2 Female – Male
G3 Girl - Boy

Reliability

R1 High failure rate - Low failure rate
R2 Unreliable – Reliable
R3 Unsound – Robust
R4 Easy to break - Not easy to break
R5 Unprotected - Protected

Skin Constructs

The aim of analysing the rotated component loading matrix of the skin constructs was to identify common constructs that different participants had used. The same three criteria used in mobile phone analysis were again used to select components. Four distinct components were identified which seems to relate to *progressive* (Table 5), *fun of use* (Table 7), *futuristic* (Table 6) and *ease* (Table 8). Starting with progressive, Table 5 shows constructs such as: unconventional – conventional, unexpected – expected, eyecatch – boring, strange – familiar, fashionable design – standard design, interesting – plain, funky – ordinary, non-conventional – conventional, and unfamiliar – familiar. Examination of Table 7 did not result in a clear common theme. Still it seems that fun or pleasurable arousal was the recurring construct as the following construct might suggest: dull – vibrant, dull – colourful, depressing – uplifting, menacing – humorous, and morbid – lively. The theme of the constructs that loaded on component 3 (Table 6) was also less obvious. However, a possible combining factor was science fiction or futuristic when considering constructs such as: abstract – futuristic, scary – space age, natural – techie and the construct outdated – futuristic had 0.67 loading on the third component. The common theme of constructs that loaded on fourth component seemed to be *ease* of using the media player or the clarity of the skin. Table 8 shows the constructs loaded on this component: tension – calm, scream – peaceful, uncomfortable – calm, complex – simplistic, and affirmed navigation – intuitive.

Table 5: Loading on component 1 of skins.

Ss	Label	Load
7	Commercial - Simple and convenient	0.77
9	Futuristic - Simple	0.83
	Unconventional - Conventional	0.72
10	Unexpected - Expected	0.85
11	Eyecatch - Boring	0.84
	Fun - Dull	0.75
12	Strange - Familiar	0.90
	Colourful - Plain	0.77
	Colourful - Dull	0.72
13	Less serious - Serious	0.92
	Complicated - Easy to use	0.71
	Fashionable design - Standard design	0.87
	Less comfortable to use - Comfortable to use	0.76
15	Interesting - Plain	0.74
	Fun - Function	0.78
	Colourful - Dull	0.71
	Themed - Standard	0.80
16	Interesting - Unappealing	0.79
	Appealing - Boring	0.75
18	Funky - Ordinary	0.75
19	Cluttered - Cleaner	0.75
	Waste of space - Straight to the point	0.78
	Complex - Simple	0.79
	Animated interface - Static	0.69
	Flowery - Sharp aesthetics	0.89
20	Non-conventional - Conventional	0.86
	Ugly - Acceptable looking	0.89
	Complicated - Simple	0.87
	Time consuming - Minimalistic	0.78
	Hard on eyes - Easy to look at	0.86
	Unnecessary metaphors - Plain	0.83
	Childish - Serious	0.82
	Unfamiliar - Familiar	0.84
	Toy - Mature	0.88

Table 6: Loading on component 3 of skins.

Ss	Label	Load
2	Abstract - Futuristic	0.87
6	Ugly - Attractive	0.96
	Not compact - Compactness	0.70
	Uncoolness - Coolness	0.94
	Desktop conflict - Desktop harmony	0.84
	Embarrassed - Proud	0.91
	Messiness - Cleanliness	0.80
11	Scary - Space age	0.71
14	Simple - More creative	0.84
15	Natural - Techie	0.73

Table 7: Loading on component 2 of skins.

Ss	Label	Load
2	Dull - Vibrant	0.69
3	Mature - Babyish	0.70
9	Dull - Colourful	0.72
	Dull - Bright	0.84
	Depressing - Uplifting	0.76
11	Gothic - Cartoon	0.82
	Menacing - Humorous	0.89
14	Mature - Childish	0.72
	Complex – User friendly	0.78
	Advanced - Beginners	0.71
15	Less colourful - Colourful	0.83
	Morbid - Lively	0.88
	Dull - Relaxing	0.76

Table 8: Loading on component 4 of skins.

Ss	Label	Load
1	Musical - Decorative	0.77
3	Tension - Calm	0.74
4	Scream - Peaceful	0.78
	Uncomfortable - Calm	0.76
5	Informative - Abstract	0.78
8	Complex - Simplistic	0.82
	Many features - Few features	0.88
16	Affirmed navigation - Intuitive	0.71

Again a number of scales were developed to measure skin evaluation dimensions by using the constructs created by participants. Each dimension is represented by five scales. In a number of cases the labels of the scale were adjusted by taking synonyms or labels of other construct to fit the overall dimension. The following list presents the dimensions with the scales.

Progressive

PO1 Expected – Unexpected
PO2 Standard – Fashionable
PO3 Conventional – Unconventional
PO4 Boring – Appealing
PO5 Plain – Interesting

Ease

EA1 Complex – Simplistic
EA2 Scream – Peaceful
EA3 Complicated – Intuitive
EA4 Cluttered – Simple
EA5 Tension - Calm

Fun of use

FO1 Serious – Humorous
FO2 Sombre – Lively
FO3 Dull – Relaxing
FO4 Depressing – Uplifting
FO5 Bland – Vibrant

Futuristic

FU1 Outdated – Ultramodern
FU2 Traditional – Futuristic
FU3 Basic – Creative
FU4 Natural – Techie
FU5 Old - Space age

Survey

The next step was to validate the scales. This was done in a survey among 156 students of the School of Information System, Computing and Mathematics (Brunel, UK). All these students received a questionnaire that included these seven-point Likert scales and two sheets with pictures of nine skins and nine mobile phones to present a frame of reference. The students were asked to evaluate the same skin and a mobile phone on the sheet. To avoid potential order effects, the scales were not ordered by dimension. Instead, the scales were taken from alternating dimension groups. Next, the polarity of half of the scales was revered by swapping the labels of the scales. The data collected in the survey was again analysis.

Mobile Phone Scales

The firs step in the analysis of the survey data was to study the reliability of the scales in each dimension. The Cronbach's alpha value for the *appearance* dimension (gender, maturity, professionalism) was 0.78. For the *ease of use* dimension the E5 scale was removed to obtain an alpha value of 0.72, which is was above the threshold value of 0.70. The *reliability* dimension failed to obtain an alpha value above the threshold, even after removing scales from the dimension. All remaining scales were entered into a factor analysis using the principle component extraction method. The analysis extracted 4 components with eigenvalues above 1. The last step was to repeat the factor analysis only including scales that had loadings of 0.70 or more on the components of the varimax rotated component loading matrix.

Table 9: Rotated component loading matrix of mobile phone scales used in the survey.

	Components			
Scales	**1**	**2**	**3**	**4**
G2 Female - Male	0.89	-0.01	0.07	0.01
G3 Girl - Boy	0.85	0.05	0.21	0.17
G1 Feminine-Masculine	0.85	0.00	-0.02	0.17
E3 Hard to use - Easy to use	0.05	0.80	-0.03	-0.26
E4 Complex-Simple	-0.02	0.79	-0.08	0.22
E1 Complicated - Simple to use	0.01	0.79	0.23	0.07
M2 Childish - Sophisticated	0.15	-0.12	0.81	0.03
M4 Silly - Classy	-0.04	0.14	0.75	-0.10
P3 Novelty - Business	0.14	0.09	0.70	0.30
P2 Fun - Technical	0.09	-0.08	0.05	0.85
M3 Playful-Serious	0.20	0.13	0.07	0.83

The resulting factor analysis again identified 4 components. Table 9 presents the rotated component loading matrix after varimax rotation. The *gender* scales seems to load clearly on the first component and the *ease of use* scales on the second component. The third component has loadings both from the *maturity* dimension (M2 and M4) and *professionalism* dimension (P3). Looking at the labels it seems that the common theme can be described as *sophistication*. The last component only has two scales with high loadings, P2 and M3, which makes it difficult to consider this as a separate dimension.

Taking into account these findings, the revised mobile phone evaluation instrument seems to include the dimensions gender, ease of use, and sophistication, with the following scales:

Gender	**Sophistication**
Feminine - Masculine	Childish - Sophisticated
Female - Male	Silly - Classy
Girl - Boy	Novelty - Business

Ease of Use

Hard to use - Easy to use

Complex - Simple

Complicated - Simple to use

Multimedia Player Scales

Again the first step of the analysis of the survey data was to conduct a reliability test on the scales of each dimension. The Cronbach's alpha values were all higher than threshold level of 0.7, with the exception of the progressive dimension. However, after removing PO1 and PO3, the alpha value became 0.72. The next step was to conduct a factor analysis on these scales, which extracted four components with eigenvalues above 1. Studying the component loading matrix after varimax rotation reveals that some scales loaded on multiple components, and therefore the factor analysis was repeated this time including only scales with a loading higher than 0.70 and excluding the EA2 scale as it was the only scale with a loading higher than 0.7 on a single component.

Table 10: Rotated component loading matrix of skin scales used in the survey.

	Component	
Scales	**1**	**2**
PO4 Boring – Appealing	0.83	0.15
FO4 Depressing –Uplifting	0.82	0.13
FO5 Bland –Vibrant	0.73	-0.08
PO5 Plain – Interesting	0.72	-0.28
FU5 Old – Space age	0.63	-0.25
EA3 Complicated – Intuitive	0.05	0.87
EA1 Complex– Simplistic	-0.11	0.83
EA4 Cluttered – Simple	-0.09	0.79

The new factor analysis identified only 2 components. Table 10 with the rotated loading matrix shows that the

loading on the first component is a combination of the *progressive* and *fun of use* dimensions. FU5, the only scale of futuristic dimension included in the factor analysis, has a relative weak loading on this component. Next, the loading on the second component seems to validate *ease* as a separate dimension in skin evaluation. Considering these results the revised skin evaluation instrument seems to include only two dimensions — *stimulation* and *ease*— with the following scales:

Stimulation	**Ease**
Boring – Appealing	Complicated – Intuitive
Depressing – Uplifting	Complex – Simplistic
Bland – Vibrant	Cluttered – Simple
Plain – Interesting	

APPLICATION OF THEORY OF REASONED ACTION

Calculation Attitude, Social Norm, and Intention

In the two-hour interviews to establish the repertory grid, additional data was also collected about the participants' attitude toward each skin and phone, the social norm, and the participants' intention to select a skin or phone. This data was collected to study the relationship between these factors, to ascertain if they were different for a more publicly noticeable mobile phone, or a more private multimedia player. In the interview, participants used Likert scales to rate: 1) the weight given to each construct, 2) their belief on how others (peers, family, or authority) think their phone or multimedia player should score on these constructs, 3) their willingness to comply to these people's opinion in relation to their design choices, and 4) their intention to select a design. For example, one participant mentioned a skin construct *explaining – basic*, this participant was therefore asked to rate the 15 skins on a 7-point Likert scale from (1) *explaining* to (7) *basic* as part of the repertory grid analysis. Additionally participants were also asked to rate on 7-point Likert scale from (-3) *bad* to (+3) *good*, each label of their construct. For example, the participant mentioned earlier was asked to respond to the following question: *For me, having a skin that is "explaining", or that is associated with this, is?* The weight and direction of the construct was calculated by taking the average of the bad-good rating of both labels of a construct after reversing the polarity of the bad-good scale of the first label. For example, if the participant had given a bad-good score of +3 for the *explaining* label and score of -2 for the *basic* label, the weight value would be (+3 × -1 + -2)/2 = -2.5. This absolute value of 2.5 represented the weight given to this construct by the participant, ranging from (0) nothing, to (3) maximum. Next, since -2.5 is a negative value the score of the skins on this construct and the labels of the construct were reversed to *basic – explaining*. This procedure ensured that all constructs had the same direction, from bad to good. To reduce the effect caused by individual variation in responding to a Likert scale, the rating on the constructs was also standardized using a z-score transformation that was based on the scores of construct rating of the skins or the mobile phones by an individual. Finally, the attitude (A) of each participant towards a design (j) was calculated by adding together all the scores of the design on a construct (e_{ij}), times the absolute weight (w_i) given to this construct by the participant, or more formally:

$$A_j = \sum_{i=1}^{10} \left(w_i \times e_{ij} \right) \quad \textbf{(1)}$$

A high attitude value meant that participants evaluated the design favourability and a low value meant that they evaluated the design not favourably.

To determine the social norm, participants were asked to rate their beliefs on how peers, family members, or authority figures would their (participants) phone or skin to score on a construct. For example a participant was asked respond to the following question on 7-point Likert scale ranging from (-3) *explaining* to (+3) *basic*: *Members of my peer group (for example, friends or colleagues) think that I should have a skin that is (or is associated with).....* Again in the calculation the direction of these normative belief scores were reversed if the weight rating of the construct was negative. Furthermore, the normative belief score was also standardized using a z-score transformation that was based on the participant's scores on these normative belief ratings over all three social groups (peers, friends, and authority). In the interview participants were also asked to rate their willingness to comply with the social groups. For peers, friends and authority participants used a four-point Likert scale ranging from (0) *not at all*, to (3) *strongly*, to respond to the following question: *In general, how much do you want to do what the following people think you should do?* Finally, the social norm (SN_j) towards a design (j) was calculated by the sum of the social norm set by a group of important others multiplied by the willingness to comply with this reference (g_k). The social norm of each group was calculated by the sum of normative beliefs (b_{ik}) on how a design should score on a construct, multiplied by the evaluation of the design on that construct (e_{ij}). Or again more formally:

$$SN_j = \sum_{k \in O} \left(g_k \times \sum_{i=1}^{10} \left(b_{ik} \times e_{ij} \right) \right) \quad \textbf{(2)}$$

$$\textit{whereby} \quad O = \{ \textit{peers}, \textit{family}, \textit{authority} \}$$

A high *SN* value meant that the social norm was in favour of a design, whereas a low *SN* value meant that the social norm was against a design. The last set of data collected in the interview was the behavioural intention of selecting a design. For example participants were asked to provide a response for each skin on 7-point Likert scale, ranging from (1) *unlikely* to (7) *likely*, to the following question: *I would try this skin on my media player?* As before, in the analysis scores of

the behavioural intention was standardized with a z-score transformation.

TRA Models

Once the value of the attitude, social norm and behavioural intention to select a design were calculated, it was possible to analyse the relationship between these three factors. Figure 1 and 2 show the correlations and partial correlations between attitude, social norm, and intention to select a design averaged over the statistical models of the 15 designs. The first observation is that the average correlations between attitude and selection intention, and between social norm and selection intention were significant in these 15 models. However, more interesting are the partial correlations between social norm and intention controlled for by attitude. Whereas it remains significant for the phone models, for the skins model this correlation was not significant. In other words, participants seem to have taken social norms, which were not correlating with their attitude, into consideration when selecting a mobile phone design. However this finding was not replicated when the skin analysis was undertaken.

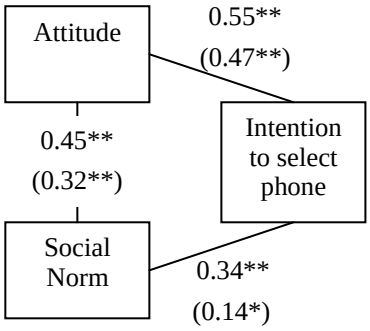

Figure 1: Mean correlations (partial correlation) between attitude, social norm and intention to select a phone. *p<.05, **p <.01.

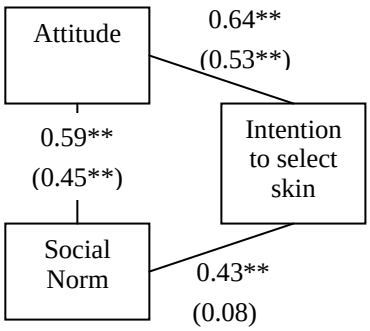

Figure 2: Mean correlations (partial correlation) between attitude, social norm and intention to select a skin for multimedia player. *p<.05, **p <.01.

CONCLUSION AND FINAL REMARKS

The study resulted into two validated design evaluation instruments: one for mobile phones and one for the multimedia player. The study also reveals that people use different type of criteria when evaluating these two devices, but also that they give different importance to social pressures when evaluating the design of a public or private application. These findings suggest that people are aware of what Goffman (1959) calls front-stage and back-stage behaviour when personalising interfaces for their mobile phones and multimedia player skins. In terms of front-stage behaviour, people are more aware that in a public place people (such as strangers sitting close to them on a train) will be able to see their mobile phone and make assumptions about that person based on their mobile phones. In contrast, if an individual is more inclined to use an application such as multimedia player in the comfort of their own home, they may be less inclined to worry about their choice of skin. This is based on the idea that no-one will see them using the skin and they may, in fact, prefer another skin when they would use their multimedia player in public locations such as the class room or in a café. Therefore, the next stage of this work is to investigate the effect of context of use on design choices.

REFERENCES

Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs : Prentice-Hall.

Brinkman, W.-P., and Fine, N. (2005). Towards customized emotional design: An explorative study of user personality and user interface skin preferences. *Proceedings EACE'05*, 107-114.

Brooks, L., Davis, C.J., and Lycett, M. (2004). Change management: the contribution of personal construct theory (PCT). *Proceedings ECIS2004.*

Goffman, E. (1959) The Presentation of Self in Everyday Life. New York: Doubleday.

Hamlin, R.P. (2000) A systematic procedure for targeting market research *European Journal of Marketing*; 34(9/10), 1038-1052.

Kelly, Y. (1963). *A theory of personality: The psychology of personal constructs*. New York: W.W. Norton & Company.

Kim, J., Lee, J. and Choi D. (2003). Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. *International journal of human-computer studies*, 59, 899-940.

McKnight, C. (2000) The Personal construction of Information, *Journal of the American Society of Information Science*, 51(8), 730-733.

Tan, F.B., and Tung, L.L. (2003). Exploring website evaluation criteria using the repertory grid technique: A web designers' perspective. *Proceedings of the second annual workshop on HCI research in MIS*, pp. 65-69.