Editorial for EJBRM Volume 11 Issue 1 2013

This issue has four papers of which two papers analyse the application of two types of quantitative methods (Finite Mixture Models and Likert scales), one paper gives an excellent review of the application grounded theory methods in IS research and the fourth paper introduces a less well known or used technique that of photo elicitation.

Both quantitative research methods papers evaluate their chosen analytical method under various practical scenarios, so that they can offer advice to other researchers on how best to taylor the method to differing practical circumstances. Tuma and Decker provide a comprehensive survey of the use of Finite Mixture Models methods in marketing, published since 2000. For marketers effective market segmentation is of great importance. But there are many problems in applying standard clustering techniques to this task. Tuma and Decker gives an excellent review of the potential value of using FMM techniques to improve the quality of segmentation. They provide advice on the appropriate selection of method under varying practical circumstances. Lantze investigates the phenomena of non equidistant Likert scales. He carries out an experiment to establish what type of question tends to generate these type of scale responses. He then analyses the effect of various types of non equidistant scales on the various statistical tests that can be applied to the date collected in this way. So that he can recommend the most effective type of test for the various types of non equidistant Likert data. The paper is clear in its objective, explains its methodology and results crisply and considers the implications of the findings in a simple, usable form.

The paper by Lawrence and Tar is an excellent introduction to the application of Grounded Theory methods (GT) for IS researchers. They establish how particularly effective this technique is for the interpretivist and case study researcher seeking to develop emergent theory. The paper gives a clear description of the GT approach of particularly Strauss & Corbin (1998). Each step is explained in detail. The case example researching internet adoption by a number of SME companies illustrates the key points in an illuminating way. This paper should be read by any researcher trying GT methods for the first time.

Finally Tonge et al give us a fascinating insight into the potential of a newer relatively underused technique – photo-elicitation. Potential interviewees are asked to record their impressions of the subject under investigation through photographs (rather than textual comments) .The paper eloquently describes an example inquiry into the phenomenon of *place attachment*, using this method. Visual learners will appreciate the photographs as means to better understand the research analysis and outcomes. The paper makes a convincing case this method for inquiries involving physical places and objects.

Ann Brown
EJBRM Editor
May 2013

Finite Mixture Models in Market Segmentation: A Review and Suggestions for Best Practices

Michael Tuma and Reinhold Decker
Bielefeld University, Bielefeld, Germany
mtuma@uni-bielefeld.de
rdecker@wiwi.uni-bielefeld.de

Abstract: Recently, Andrews, Brusco and Currim (2010) noted that some of the hesitancy on the part of practitioners to adopt model-based (MB) methods in market segmentation (MS) may stem from an insufficient awareness of their performance relative to their non-model-based (NMB) counterparts. Comparisons of MB and NMB methods should provide business researchers with information as to precise conditions in which the former should be preferred. Moreover, finite mixture models (FMMs) have grown in their use since 2000 and, as there is no recent survey-based empirical literature examining their application, a comprehensive review of their usage in segmentation research seems to be of use. This article discusses some of the critical issues involved when using FMMs to segment markets, takes a closer look at comparison simulation studies in order to highlight conditions under which a business analyst might consider the application of an FMM approach, discusses model selection as well as validation issues and provides suggestions for best practices and potential improvements. Furthermore, it presents an empirical survey that seeks to provide an up-to-date assessment of FMM application in MS.

Keywords: market segmentation, model-based clustering, finite mixture models, latent class models

1. Introduction

Marketers usually address consumer heterogeneity by grouping consumers into segments consisting of those consumers having relatively similar product or service needs. Cluster analysis (CA) is one of the most widely used methods in segmenting markets (Wedel and Kamakura, 2000). Most clustering done in MS practice is based largely on heuristic procedures like Ward's method and *k*-means (Tuma, Decker, and Scholz, 2011). However, the often insufficient statistical basis of such methods appears to be a major drawback for their use and crucial issues in segmentation, such as determining the optimal number of segments, can hardly be answered by heuristic procedures. Lubke and Muthén (2005:23), for instance, note in this respect that "clustering using *k*-means is achieved based on an arbitrarily chosen criterion which aims at minimizing the within-cluster variability while maximizing between-cluster variability".

FMMs or MB clustering are a principal alternative to heuristic-based algorithms. They are "viewed as elegant procedures that incorporate mixtures of parametric distributions to define the true cluster structure" (Steinley and Brusco, 2011:63). They are the main statistical approach to clustering and segmentation and some academic literature tout and advocate their usage "as a preferred approach because of the provision of a formal statistical model" (Andrews, Brusco and Currim 2010:609; McLachlan and Peel, 2000). In practice, a business analyst embarking on the use of FMMs for segmentation has to deal with several crucial issues like selecting the type of FMM, variable selection, determination of the number of variables and sample size, data pre-processing requirements, determination of the number of segments, validity and stability tests of the obtained results as well as the interpretation and substantial description of segments. All these issues highly influence the outcome and quality of the derived market segments regarding further market-centric activities. The seminal work of Wedel and Kamakura (2000) provides a comprehensive review of FMM applications in MS until the turn of the last century. Since then, however, there have been many new developments in FMMs applications. In the domain of marketing, new models have been designed, implemented and published in top-tier journals (e.g. Hahn et al., 2002). Andrews, Brusco and Currim (2010) suggest that some of the hesitancy on the part of practitioners to apply MB approaches to MS may stem from an insufficient knowledge of their performance relative to their NMB counterparts. Whereas the performance of NMB techniques has been evaluated using Monte Carlo simulations for more than three decades (Blashfield, 1976; Steinley and Brusco, 2008a/2008b), extensive simulation studies comparing MB and NMB methods are still slowly emerging in the literature and should ultimately provide business researchers with information as to precise conditions under which MB approaches are apt to be preferred. Moreover, as user friendly software packages like Latent GOLD have made their debut and FMMs have grown in their use since 2000 and, to the best of our knowledge, as

Reference this paper as: Tuma, M and Decker, R."Finite Mixture Models in Market Segmentation: A Review and Suggestions for Best Practices" *The Electronic Journal of Business Research Methods* Volume 11 Issue 1 2013 (pp 02-15), available online at www.ejbrm.com

there is no recent survey-based empirical literature examining their managerial application, an up-to-date review of their usage in segmentation research seems to be overdue.

Against this backdrop, the aim of this article is two-fold: First, it reviews the crucial issues in MS using FMMs. In order to inform business managers and marketing researchers about the performance of competing methods, it builds on the suggestion of Andrews, Brusco and Currim (2010) and, drawing on prior findings in marketing and other disciplines, provides a closer look at comparison simulation studies, with set-ups and key results regarding MB and NMB segmentation methods. The paper contributes to the current literature by integrating previous research results in order to present a holistic overview and comprehensive insights (e.g. concerning the performance of various types of FMMs) into new developments by addressing the critical issues. Furthermore, it presents an empirical study that investigates how some of the important problems in MS using FMMs have been addressed by researchers and, last but not least, it examines how the application of easy-to-use software packages have led to more rigorous applications of FMMs in MS.

Accordingly, the remainder of the paper is structured as follows. Section 2, among other things, takes a closer look at some of the critical issues when applying FMMs. Section 3 presents selected simulation studies. Section 4 specifies the research questions and the methodology underlying our study, whereas Section 5 is devoted to important results. This is followed in Section 6 by a brief conclusion, managerial implications and an outlook on future research.

2. Finite mixture models

Several extensions of the basic FMM approach have been suggested in the recent past. Finite mixture regression models (FMRMs) (Wedel and Kamakura, 2000) for example, are able to simultaneously derive segments and segment-specific weights that relate an outcome or dependent variable (e.g. product recommendation or rating) to a set of independent or explanatory variables (e.g. price of a product and product quality) and derive a unique regression model for each segment. Finite mixture approaches have also been developed that combine the strengths of the partial least squares (PLS) method or the covariance structure analysis that are used for better understanding heterogeneity within structural equation models with the advantages of classifying market segments according to FMMs. Jedidi, Jagpal and DeSarbo (1997) pioneered the development of the finite mixture structural equation model (FIMIX-SEM), an approach that combines FMMs, the Expectation-Maximization (EM) algorithm and covariance-based SEM. The original technique is inappropriate for PLS analysis because of divergent methodological assumptions. For this reason, Hahn et al. (2002) introduced the finite mixture partial least squares (FIMIX-PLS) method that combines a FMM procedure with an EM algorithm specifically coping with the ordinary least squares (OLS)-based predictions of PLS. Conceptually, FIMIX-PLS is equivalent to a mixture regression approach. However, the main difference is that the structural model can comprise a multitude of (interrelated) endogenous latent variables (Hahn et al., 2002). One of the assumptions of the aforementioned FMMs is that the available sample has only a single level, i.e., it consists of a sample of independent units, an assumption that is inadequate when the sample to be analyzed has multiple levels, i.e., when units are nested within clusters sharing common environments, experiences and interactions (Lukočiene, Varriale and Vermunt, 2010). The multilevel latent class (MLC) model with finite mixture distributions at multiple levels of a hierarchical structure has been developed for the analysis of data sets having such a multilevel structure.

2.1 Estimation of FMMs

Usually, the FMM parameters are unknown and have to be estimated from the data. There is a remarkable variety of estimation methods such as the method of moments, maximum likelihood (ML), minimum chi-square, and Bayesian approaches (McLachlan and Peel, 2000). Numerical methods for obtaining ML estimates primarily involve the use of gradient methods like Newton-Raphson, quasi-Newton, and Fisher's scoring. Other approaches rely on the EM algorithm (Dempster, Laid and Rubin, 1977), stochastic EM (SEM) or Markov Chain Monte Carlo (MCMC) methods. Parameter estimation in Bayesian methods is with MCMC using Gibbs sampler and Metropolis-Hastings algorithm.

The primary advantages of numerical optimization procedures are their speed, relative to the EM algorithm, and their ability to obtain standard errors for parameter estimates. The primary advantages of the EM algorithm are that (1) at each stage of the iterative process the likelihood is monotonically increasing and (2) under certain regularity conditions, the sequence of likelihoods will converge to at least a local maximum. The

latter is an iterative, hill-climbing procedure whose performance can depend severely on the particular starting point (McLachlan and Peel, 2000). Hence, numerous initialization procedures have been suggested in the literature (see, e.g., Melnykov and Maitra, 2010).

Distributional assumptions for the variables have to be made by the researcher when estimating model parameters. Checking the empirical distributions and consulting skewness and kurtosis measures or plotting the data for a visual representation can be very helpful in this regard (Wedel and DeSarbo, 2002).

Identifiability is an issue related to parameter estimation and determines whether a unique solution can be obtained. It can be investigated by inspecting the Hessian matrix of second derivatives of the likelihood (Bekker, Merckens and Wansbeek, 1994). Positive eigenvalues of the information matrix provide evidence of identifiability (Wedel and DeSarbo 2002). The non-identifiability of the components leads to so-called label switching. If this occurs, summary statistics of the marginal distributions will give inaccurate estimates (Dias and Wedel, 2004). Label switching can be detected through investigating iteration plots of the MCMC sampler (Ebbes, Grewal and DeSarbo, 2010). Frühwirth-Schnatter (2006) provides an overview of approaches to address label switching.

2.2 Model selection

2.2.1 Determining the Number of Segments

When applying FMMs to empirical data, the actual number of segments *S* is unknown and must be inferred from the data itself. The majority of methods devoted to estimating *S* can broadly be divided into two categories. The first group of methods relies on testing procedures while the second one is based on information criteria (IC). Particularly, the latter class of methods is frequently used for investigating the number of clusters (Sarstedt 2008). These methods determine the number of segments by minimizing the negative log-likelihood function augmented by some penalty function, which increases with the number of parameters and/or the number of observations to reflect its complexity. Table 1 presents some of the model selection criteria that have recently been used in MS.

Table 1: Some model selection criteria[1]

Criterion	Reference
Akaike IC (AIC)	Akaike (1974)
Bayesian IC (BIC)	Schwarz (1978)
Consistent AIC (CAIC)	Bozdogan (1987)
AIC3	Bozdogan (1994)
Normalized Entropy Criterion (NEC)	Celeux and Soromenho (1996)
Validation sample log-likelihood (LOVLG)	Andrews and Currim (2003a)
Sample size adjusted BIC (ssBIC)	Sclove (1987)
Classification error	Garver, Williams and Taylor (2008)
Markov switching criterion (MSC)	Ebbes, Grewal and DeSarbo (2010)
R^2	Garver, Williams and Taylor(2008)
Log marginal density (LMD)	Hofstede, Wedel and Steenkamp (2002)
Deviance IC (DIC)	Spiegelhalter et al. (2002)

2.2.2 Variable selection

Selecting the appropriate clustering variables actually used in segmentation is one of the most fundamental steps in the segmentation process. It has long been recognised that not all variables contribute equally to defining the underlying segment structure. In many multivariate datasets, for example, some of the variables are highly correlated with the others or just do not carry much additional information about the potential segments. Since the performance of segmentation algorithms can be severely affected by the presence of such variables, their elimination can potentially improve both estimation and clustering performance (Melnykov and Maitra, 2010).

[1] Most of the model selection simulation studies focus on IC. Sarstedt et al. (2011) consider classification criteria such as complete log-likelihood, Entropy criterion, etc.

The marketing researcher must also decide on the number of variables to be considered, the sample size and the relation between these and the resulting segments. The relationship between the number of objects to be grouped and the number of variables to be used is important, given that the number of variables used determines the dimensionality of the space within which the method or model is searching for groupings. Formann (1984) suggested a sample size of 2^k, where k represents the number of variables used in segmentation as a rough guide for this relation. Preferably, it should be $5 \cdot 2^k$ respondents.

Special care must be taken with LC regression (LCR) to ensure that the appropriate sample size exists for each segment in the model. Consistent with multiple regression, LCR generally requires at least five observations per independent variable per segment (Hair et al. 1995). A rule of thumb suggests that a minimum sample size of 30 observations/respondents per segment may be adequate (Garver, Williams and Taylor, 2008).

2.3 Other Important Issues

2.3.1 Stability and Validity

Various approaches, subsumed under the terms validity and stability, have been developed to analyse the quality of the final segment solution. The latter can be evaluated by running several clustering procedures or one clustering procedure several times (with different specifications) on the same dataset and testing whether the partitions remain constant and are thus stable. Validation includes attempts by the researcher to ensure that the segments are representative of the general population. Strategies for validation may be based on external, internal and relative criteria (Wedel and Kamakura, 2000).

2.3.2 Software packages

The R environment provides several packages for estimating mixture models, e.g. MCLUST for mixtures of multivariate Gaussian distributions, fpc for mixtures of linear regression models, mmlcr for mixed-mode latent class analysis, poLCA for polytomous outcome variables, and flexmix for FMRMs. In SAS, the package PROC NLP can be used to specify different mixture models. The Stata package fmm estimates FMRMs.

3. Simulation studies

3.1 FMMs

Andrews, Currim and Leeflang (2010) present a simulation experiment that provides insights into the conditions under which prediction bias may occur, and, when it occurs, to understand why by determining the effects of data aggregation (panel vs. store level), heterogeneity, endogeneity, and the number of households, stores, and weeks on bias in sales response predictions. Among other things, they manipulated the degree of heterogeneity between and within segments as well as the number of weeks, stores and households per store considered. Using choice data comprising more than 300 panel and store-level datasets and assuming two segments of consumers for all datasets, they estimated several nested logit models and nested logit FMMs. In this study models explaining within-store heterogeneity (i.e., heterogeneity across store visits) using random distributions for the coefficients produced predictions that were significantly more accurate than those of the other models. One implication of this finding is that if the sole objective of an analysis is to predict segment-specific market response to a new promotional environment, store-level data should suffice. Often, the latter is cheaper to obtain, more widely available, and more computationally efficient than panel data.

Andrews et al. (2010) compare the effectiveness of statistical MB clustering methods with that of more commonly used NMB procedures. They therefore manipulated the number of segments, consumers and characteristic variables, as well as the concordance between response-based and characteristic-based segments, the scale type for characteristics and the availability of predictor variables. From more than 800 generated datasets they found that if a manager's primary objective is to forecast responses for segments of holdout consumers for whom only characteristics are available, NMB procedures perform better than MB procedures. However, if it is important to understand the true segmentation structure in a market as well as the nature of the regression relationships within segments, the MB procedure is clearly preferred.
Andrews, Brusco, and Currim (2010) compare three approaches for forming a consensus segmentation scheme, namely clique partitioning, the SEGWAY algorithm and a method based on a latent class model.

Among others, they manipulated the number of customers and latent segments, the segment membership probabilities as well as the number of partitions and the number of classes within each partition. Using ANOVA to investigate the results of a total of 648 generated datasets, they found that the FMM approached yielded better average recovery of holdout validation segments than did the deterministic methods. For marketing researchers seeking to obtain consensus segmentation, FMM seems to be a promising option.

Andrews, Ainslie and Currim (2002) compare the performance of FM logit models and hierarchical Bayes-estimated mixed logit models with discrete versus continuous representations of heterogeneity in terms of the accuracy of household-level parameters, fit, and forecasting accuracy. The authors experimentally manipulated the number of mixture components, the separation between mixture components, the distribution and variance of coefficients within components, the number of households and purchases as well as the error variance. The set of predictors included one continuous variable (price) and two binary variables (store feature advertisement and aisle display). Based on 288 choice datasets the FMMs proved to have the best overall performance with regard to parameter recovery. In general, the models fit better when the separation between components is larger, the within-component distribution is normal, the within components variance is smaller, and when there is less error variance. The results indicate that FMM is the preferable method to use for marketing analysts seeking to obtain forecasting accuracy.

3.2 Estimation of FMMs

Using a synthetic dataset Dias and Wedel (2004) provide simulation comparisons of ML estimation methods (EM, SEM, MCMC) on the basis of three convergence criteria and initialization methods such as starting with random centers (RC) based on McLachlan and Peel (2000:55), or with a random partition (RP) of the data. Furthermore, they investigated an approach to minimize the label-switching effect based on imposing identifiability constraints on the parameters (e.g. segment sizes, means and variances) and the methods proposed by Celeux (1998), Celeux, Hurn and Robert (2000) (CHR) and Stephens (1997).

For the EM algorithm, the relative criterion and the Aitken's absolute criterion underperformed the absolute criterion. They found that using RC for each convergence criterion decreases the proportion of false solutions (for a given number of iterations). For SEM, RP solutions outperformed the best EM solution in terms of the log-likelihood value. SEM proves to be faster and displays better convergence properties, but is less stable than the EM. For MCMC, the identifiability constraints negatively affect parameter recovery, in particular when the component sizes or variances are constrained. Furthermore, the absence of any label-switching procedure outperformed procedures in which identifiability constraints are imposed and Stephens (1997) and CHR relabeling procedures were the most effective.

Finally, Dias and Wedel (2004) conclude that MCMC is preferable over EM and SEM in recovering the parameters of mixture models. The absolute convergence criterion should be used in conjunction with the EM algorithm and RP with the SEM. Identifiability constraints should be avoided, but, if they have to be used, then the better performing ones, for instance that of Stephens (1997) and CHR, are recommended.

3.3 Model selection

3.3.1 Determining the number of segments

Simulation studies in the context of FMMs can be broadly classified according to the type of FMMs used in generating the datasets. Tables 2, 3 and 4 present set-ups and key results of model selection simulation studies in the contexts of FIMIX-SEM and -PLS, LC models and FMRMs, respectively.[2]

Table 2: Set-ups and key results for model selection simulation studies in the context of FIMIX-SEM and -PLS

	Sarstedt et al. (2011)	Henson, Reise and Kim (2007)	Jedidi, Jagpal and DeSarbo (1997)
Model	PLS	SEM	SEM
No. of manipulated data characteristics	6	5	2
No. of segments	2 or 4	1, 2 or 3	2 or 4

[2] Because of space limitations, we present only the best criteria. The interested reader should consult the respective studies for a complete list of the criteria considered.

	Sarstedt et al. (2011)	Henson, Reise and Kim (2007)	Jedidi, Jagpal and DeSarbo (1997)
Model	PLS	SEM	SEM
Disturbance term of the endogenous latent variables	10% or 25%	-	-
Distance between segment-specific path coefficients (separation between segments)	0.25 or 0.75	-	-
Sample size	100 or 400	500, 1500 or 2500	-
Model complexity	low or high	-	-
Relative segment sizes	Balanced or unbalanced	unbalanced	-
Factor level combinations	64	-	-
Mixture proportion	-	50%-50%, 70%-30%, 90%-10%	-
No. of indicators	-	-	3 or 6
No. of datasets	-	121500	-
Type of dataset	-	-	structured data
Distribution for datasets	-	binomial distribution	multivariate normal
Best criterion	AIC3, CAIC	ssBIC	BIC

Table 3: Set-ups and key results for model selection simulation studies in the context of LC models

	Dias (2006)	Lukočiene and Vermunt (2010)	Lukočiene, Varriale and Vermunt (2010)	Lukočiene, Varriale and Vermunt (2010)
Model	Binary LC	MLC	MLC with categorical indicators	MLC with continuous indicators
No. of manipulated data characteristics	5	3	7	6
No. of variables	5 or 8	6	6 or 10	6
No. of segments	2 or 3	3 at individual level, 2 or 3 at higher level	2 or 3 at both levels	2 or 3 at both levels
Segment sizes	equal proportion, unbalanced	-	0.7 or 0.8	0.7 or 0.8
Separation between segments	well, moderately or weakly separated	Lower level (moderately separated), higher level (from very low to very high)	from very low to very high separation	from very low to very high separation
Sample sizes	600, 1200 or 2400	Lower level (5, 10, 15, 20), higher level (50 or 500)	lower level (5, 10, 20, 50), higher level (30, 100, 1000)	lower level (5, 10, 20, 50), higher level (30, 100, 1000)
Measurement level of variables	-	discrete	discrete	-
No. of datasets	10800	2000	2880	-
Type of dataset	binary	binary with hierarchical or multilevel structure	hierarchical or multilevel structure	hierarchical or multilevel structure
Best criterion	AIC3, AIC	AIC3	AIC3	AIC3, BIC(K)

Table 4: Set-ups and key results for model selection simulation studies in the context of logit and FMRMs

	Andrews and Currim (2003a)	Andrews and Currim (2003b)	Sarstedt and Schwaiger (2007)	Sarstedt (2008)
No. of manipulated data characteristics	7	8	-	5
No. of segments	2 or 3	2 or 3	3	2

	Andrews and Currim (2003a)	Andrews and Currim (2003b)	Sarstedt and Schwaiger (2007)	Sarstedt (2008)
No. of manipulated data characteristics	7	8	-	5
No. of segments	2 or 3	2 or 3	3	2
Regression coefficients in each segment	-	-	S1 (1, 1, 1.5, 2.5); S2 (1, 2.5, 1.5, 4); S3 (2, 4.5, 2.5, 4)	S1 (1, 1,1.5, 2.5); S2 (1, 2.5, 1.5, 4)
Mean separation between segments	small (1.0), medium (1.5) or large (2.0)	small (0.5), medium (1.0) or large (1.5)	-	-
Sample size	100 or 300	100 or 300	Varied in 100-step intervals of [100:1000]	Varied in 10-step intervals of [50:500]
Mean no. of purchases per household	5 or 10	-	-	-
No. of observations per individual	-	5 or 10	-	-
No. of choice alternatives	3 or 6	-	-	-
No. of predictors	-	3 or 6	3	-
Error variance	1.645 or 50% higher	20% (R^2 = 0.80) or 60% (R^2 = 0.40)	-	-
Segment size	5%-10%, 10%-20% or 20% -30%	5%-10%, 10% - 20% or 20%-30%	(0.1, 0.1, 0.8), (0.2, 0.2, 0.6) or (0.3, 0.3, 0.4)	(0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6) or (0.5, 0.5)
Measurement level of predictors	-	continuous or discrete	continuous	-
No of datasets	864	1728	-	230000
Type of dataset	scanner panel data	normal data	normal data	normal data
Distribution for datasets	gamma distribution	normal distribution	-	standard normal
Best criterion	AIC3	AIC3	AIC3	AIC3

These tables show that AIC3 performs well. Sarstedt et al. (2011:52) conclude that "In summary, our key finding and decision rule are to use AIC3 and CAIC jointly when evaluating FIMX-PLS results."

To sum up, current evidence from the simulation studies considered in this research suggest that the accuracy of commonly used model selection criteria for determining the number of segments in a sample strongly depends on the usage context, including the types of distributions used to describe the data, the model specification, and the characteristics of the specific market. However, these results also indicate that AIC3 seems to be a good criterion to use across a wide variety of model specifications and data configurations.

3.3.2 Variable selection

Steinley and Brusco (2008b) recently evaluated eight variable selection techniques for MB (Law, Figueiredo and Jain, 2004; Raftery and Dean, 2006; and Dy and Brodly, 2004) and NMB clustering. They used 20412 datasets, each one generated with 250 observations and systematically manipulated factors, such as the number of segments, true structure variables, and masking variables, as well as the density of the segments, the average probability of overlap between segments on each true structure variable and the degree of within-segment correlation. The most effective method was the procedure proposed by Steinley and Brusco (2008a) for *k*-means. They found that variable selection methods used in conjunction with FMMs performed the worst suggesting that a business analyst should avoid using or use variable selection methods with FMMs only if it is necessary to do so.

3.4 Other important issues

3.4.1 Stability and validity

Brun et al. (2007) investigated the performance of internal (trace criterion, determinant criterion, invariant criterion, correlation with Euclidean distance matrix, silhouette index), relative (figure of merit, stability) and external (Hubert's correlation, Rand statistic, Jaccard coefficient, Folkes and Mallows index) validation indices[3] applied to the outcomes of several clustering algorithms (*k*-means, fuzzy *c*-means, self organizing maps (SOM), single, complete and average linkages, and MB clustering methods) under realistic conditions in order to evaluate their performances. They used several models with different mixtures (regarding dimensionality and shape of the mixture distributions considered). They conclude that the Rand statistic and the silhouette index are the best performing external and internal validation indices, respectively. A business researcher should therefore consider these measures for external and internal validation, respectively.

3.4.2 Software packages

Haughton, Legrand and Wolford (2009) recently compared three software packages, Latent GOLD, MCLUST and poLCA that can be used to perform LC-based market segmentation. Using a dataset having continuous, discrete or mixed variables and by applying each software package to develop a LC CA for this data, they were able to compare software features and the resulting clusters. The results obtained using MCLUST outperformed those by Latent GOLD according to the measure of heterogeneity they used. From the perspective of usability they concluded that Latent GOLD is the easiest to use with a well-written and usable documentation and a GUI interface that eliminates the need for user programming.

4. Methodology of this study

4.1 Instrument

The study instrument used to gather data for this research was developed based on a comprehensive review of the FMM literature. In addition to general information such as authorship and publishing data, the coded criteria reflect the basic structure of a typical FMM application in MS. These criteria can be divided into four broad stages representing the number of steps that are important to the quality of a segment solution. Table 5 specifies the data collected for the literature analysis.

Table 5: Data collected in the literature analysis

Criteria		Data collected
Model type		Type of mixture model used as stated by the authors
Parameter estimation		Methods used for parameter estimation (e.g. ML or Bayesian methods, initialization methods, convergence, identifiability, number of iterations, type of distribution, label switching)
Model selection	Determination of the number of clusters, variable selection and related issues	Number of segments; model selection criteria and reasons for their usage; segment sizes; segmentation variables; number of variables used in the segmentation; sample size
	Data pre-processing	Data pre-processed before being clustered; data pre-processing method; number of variables before and after data pre-processing; reasons for data pre-processing
Other important issues	Validity and stability	Evaluation of stability and validity; methods used
	Interpretation and description of segments	No description; partial description; full description
	Software used (FMM-related packages)	Type of software (e.g. Latent Gold or MCLUST)

[3] Please see Brun et al. (2007) for a detail description of the indices.

4.2 Data collection and analysis

As the nature of research on FMM applications is difficult to confine to specific disciplines, the relevant material is scattered across various journals. We therefore searched online journal databases to obtain a representative as possible bibliography of FMM applications in marketing and business literature. In doing so, most of the top-tier marketing, business, management and tourism journals were included in the literature analysis. The literature search was based on numerous descriptors, such as 'mixture model(s)', 'finite mixture model(s)', 'latent class model(s)' (LCMs), 'latent class analysis', 'multilevel latent class analysis', 'finite mixture structural equation model(s)', 'finite mixture partial least squares', 'Bayesian mixture model(s)' (BMM), 'latent class', 'mixture model', 'hidden Markov models' (HMM), etc. in conjunction with the words 'segmentation', 'marketing' and 'business'.

The full text of each article was reviewed to eliminate those articles that were not related to marketing/business and to the objectives of the study. The selection criteria were as follows: Only those FMM application articles that had been published in journals within the target timeframe (from 2000) in a marketing setting were selected. This search yielded 108 articles from 63 journals with about 40 percent of the articles coming from at least category A journals according to the 'Journal Quality List' of Harzing (2011).

5. Results

5.1 Finite mixture models used

An analysis of the data by the type of mixture models used as stated by the respective authors unveiled that FMMs and LCMs are dominant in MS. The other cited mixture model approaches – Bayesian methods, FIMIX-PLS, FIMIX-SEM, HMM, MLC with 2.8, 1.9, 1.9, 4.6 and 1.9 percent of applications, respectively – seemingly did not enjoy the same popularity as FMMs (29.6%) and LCMs (51.9%) in the considered time-period.[4] 4.6% of the studies used latent class mixture models. In one study the name of the software package was stated as the method used. In 18.5, 9.3 and 1.8 percent of the studies regression, logit and probit models were used respectively. Compared to Wedel and Kamakura (2000), these results show a considerable increase in the usage of new FMM types.

5.2 Parameter estimation in mixture models

ML (41.7%) is the most popular method used for parameter estimation. Bayesian methods were applied in 6.5% of the studies and, remarkably, in 51.9% the method used was not ascertainable. Despite some of its known shortcomings, the EM algorithm, used in 17.6% of the studies, was the most common method for parameter estimation. Numerical methods were used in 8.3% of the studies considered and in one study EM and MCMC were combined. Although MCMC clearly outperformed the EM algorithm in the study of Dias and Wedel (2004), it was implemented only in 6.5% of the studies considered.

The normal (Gaussian) distribution was used in 13.9% of the studies. Other types of distributions used include multivariate normal (9.3%), multinomial (5.6%), Poisson (3.7%), binomial (1.9%), Dirichlet-multinomial (1.9%), logistic (0.9%), Laplace (0.9%), Wishart (0.9%) and truncated bivariate normal (0.9%). Normal and multinomial distributions were combined in two studies (1.9%) and normal and logistic in one (0.9%). The distribution was not stated in 57.4% of studies.

Convergence was assessed in only 13.9% of the studies. The most popular method used to assess convergence was examining the iteration plots, followed by Geweke's (1992) convergence test (Rust and Verhoef, 2005). Other methods used to assess convergence include comparing the within to between variance for each parameter estimated across multiple chains (Netzer, Lattin and Srinivasan, 2008) and Heidelberger and Welch stationarity test (Rust and Verhoef, 2005).

Only 13% of the studies used different starting values to initialize the algorithm. Random starting points were used by almost all studies. Only 19.4% of the studies used several runs of the algorithm to overcome the problem of local maxima. Identifiability and label switching were addressed in 16.7% and 4.6% of the studies, respectively.

[4] In this context, it has to be mentioned that both terms, FMM and LCM, are often used in a synonymous way.

These results show that crucial factors that can be assumed to significantly impact the final solution are not adequately addressed in many of the studies – estimation methods used (51.9%), convergence of the algorithm used (86.1%), initializing the algorithm (87%), several runs of the algorithm (80.6%), identifiability (83.3%) and label switching (95.4%).

5.3 Model selection

5.3.1 Methods used to determine the number of segments

Figure 1 shows the methods used to determine an optimal/adequate number of segments in FMM applications. The category "Others" include those methods which were used only once. 12% of the studies did not indicate how they arrived at the number of segments. Interestingly, despite the repeated successes of AIC3 in determining the true number of segments in many simulation studies, its usage seemingly is not widespread in empirical applications. An investigation of the reasons for selecting a particular model selection criterion reveals that 87% of studies did not put forward any reasons for using a particular procedure to determine the number of segments.

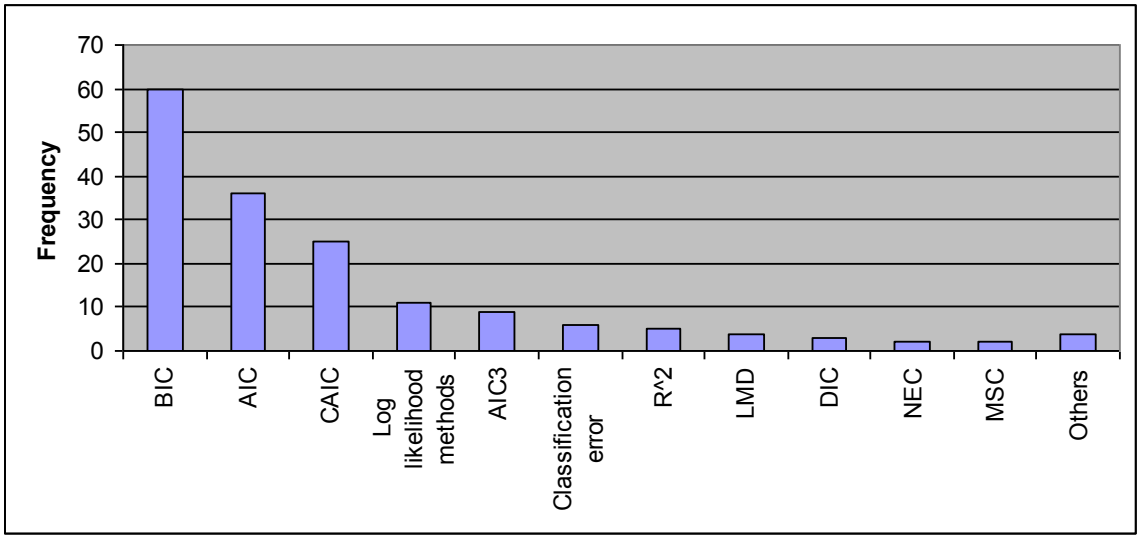

Figure 1: Methods used to determine the number of segments

Confidence in the number of clusters/segments is greater when multiple methods used to determine them converge. However, only 37% of the studies used this approach.

The minimum number of segments derived was two and the maximum was 14. On average 3.5 segments were derived. The number of segments was not ascertainable in 5.6% of the studies. Atella et al. (2004:660) state that in applications of FMMs most studies find that only a small number of latent segments are needed to describe the data adequately. Using additional data from Tuma, Decker and Scholz (2011) a one-way ANOVA test was carried out to check whether the clustering methods used (hierarchical, non-hierarchical methods and FMMs) have an influence on the number of segments derived. The test results ($p < 0.01$) are highly significant suggesting that FMMs maybe the appropriate method to use when the analyst suspects the existence of only a few segments in a dataset. The segment sizes were ascertainable in only 70.4% of the studies. This is a cause for concern considering that many managerial decisions are based on this information. The smallest segment used in the reviewed studies contains no more than seven respondents. The maximum segment size had 334083 respondents. 6.3% of the studies did not comply with the rule of thumb suggested in Garver, Williams and Taylor (2008).

5.3.2 Variable selection, number of variables used in clustering and sample size

The outcome of an analysis of the data regarding variable selection is illustrated in Figure 2. As the bar chart shows, there is some variety in the use of segmentation variables in FMM applications. In approximately 5% of the articles no information is provided about the segmentation variables used. Given that the segmentation

variables highly impact the resulting segmentation solution, this finding is a cause for concern. Despite the criticism of demographic variables, they continued to be used intensively in segmentation studies.

A promising new source for obtaining interesting segmentation variables is online consumer reviews as provided for example by Epinions.com. As Decker and Trusov (2010) point out, the analysis of freely expressed customer opinions is a promising alternative to conventional survey techniques since the reviewers or respondents have not been requested to communicate their opinion, but are doing this voluntarily. Accordingly, a high level of authenticity can be expected. None of the articles considered used this promising data source for segmentation variable elicitation. 70% of the segmentation data comes from opinion surveys (face-to-face interviews, mail interviews, online surveys, etc). 25.5% come from databases and the remaining 4.5% is unascertainable.

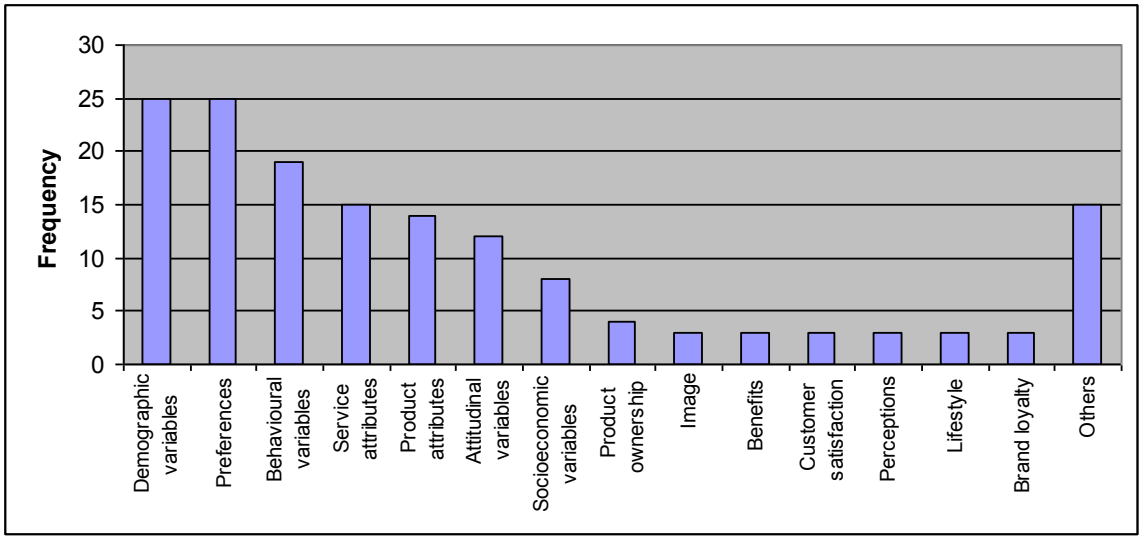

Figure 2: Segmentation variables used in FMM applications

The data were also analysed by looking at the number of variables used, the sample sizes and the relationship between these. The sample sizes were ascertainable in 97.2% of the studies and the numbers of segmentation variables was made available in 86% of the papers considered. A noteworthy 60% of the studies included in this literature analysis did not comply with the rule of thumb suggested by Formann (1984).

In 22.2% of the studies, the data were pre-processed, reducing the number of variables on average from 25 to 6 equalling a reduction of 76%. 9.3% of the studies included in this review factor analyzed the segmentation variables. The variables were standardized in an equal amount of studies. In 3.7% of the studies standardization and factor analysis were combined.

5.4 Other important issues

Validity and stability may be the most neglected issues in segmentation analysis. Surprisingly, they were not investigated in 83.3% and 65.7% of the studies respectively. The validation methods used include cross-validation (hold-out) sampling (26.9%), external validation (5.6%) and internal validation (1%). A combination of all three methods was reported in one study.

The segments were not described or interpreted in 16.7% of the studies. This is of some concern considering that the description and interpretation of segment solutions are the basis for formulating credible positioning and targeting strategies, for example.

In 64.8% of all FMM publications, the software used was not stated. Latent Gold, Gauss, LIMDEP/NLOGIT, Mplus, SmartPLS, LEM Software and SAS were used in 8.3, 6.5, 4.6, 3.7, 2, 2, and 2% of the studies, respectively. MCLUST, WinBUGS, Stata, poLCA and EMMIX were used in one study each. In one study Mplus and LIMPDEP\NLOGIT were combined.

6. Conclusion and outlook

MS remains the bedrock on which the positioning and targeting decisions of many businesses are built. Obtaining robust and valid solutions is therefore of crucial importance. FMMs have been promoted by some methodologist as a sophisticated method for obtaining good segmentation solutions.

In this study, we complement the toolbox of business practitioners by presenting FMMs as a comparatively new and interesting class of business research methods for deriving homogeneous groups of customers, for example. The presentation of set-ups and key results of simulation studies was intended to provide guidance for those seeking an appropriate FMM, model selection criteria, validation procedures, etc. for similar segmentation problems. Furthermore, identifying FMMs and methods for model selection, parameter estimation, variable selection and validation that are most robust to variations in data characteristics is of great importance for business researchers focusing on data-based market segmentation since they can use these methods with more confidence regarding the accurate recovering of consumer behaviour and characteristics. The presented results, among others, show that AIC3, MCMC, the absolute convergence criterion in conjunction with the EM algorithm and RP with the SEM, and MCLUST are useful methods or tools across a wide variety of model specifications and data configurations.

In order to review the common practices of marketing researchers, we carried out a comprehensive survey of 108 articles in which FMMs were applied in MS. The literature analysis was intended to serve as an orientation for marketing researchers interested in both best practices and pitfalls. The results show that crucial factors such as estimation methods, initializing and convergence of the algorithm used, identifiability and label switching, model selection criteria, stability and validity, etc. which may impact the final segmentation solution significantly are not always adequately addressed and reported in detail. Failure to provide specific information about the segmentation variables or method(s) used tends to inhibit replication and provides little guidance for researchers or business practitioners who might seek an appropriate method for a similar problem. Furthermore, the results show a remarkable discrepancy between simulation studies and applied FMMs methods. Methods or tools that performed well in simulation studies such as AIC3, MCMC and MCLUST are hardly used in practice.

Based on the available results, we believe that several issues have to be addressed by business researchers. Firstly, they should attempt to integrate best practice recommendations into their use of FMMs. Methods that performed well in simulation studies should be considered more intensively in empirical FMM applications. Furthermore, academic researchers should pay more attention to a detailed description of the FMM method used in their publications in order to provide better guidance to those who want to apply the respective (new) approach in practice. And last but not least, the idea of finite mixture modelling should increasingly find its way into education in business and marketing management in order to enable future generations of business and marketing researchers to successfully apply this powerful new class of market segmentation methods.

Future research should focus on large-scale simulation studies using a wide range of models and statistical distributions, and also investigate in-depth the effects of distributional misspecifications on the segmentation results. Furthermore, promising new data sources such as online consumer reviews should be considered in model-based segmentation. Model-based reviewer clustering, for instance, can help to better understand which reviewer characteristics have an influence on trust in this emerging source of word-of-mouth.

References

Akaike, H. (1974) 'A new look at statistical model identification'. *IEEE Transactions on Automatic Control*, AC-19: 716-723.

Andrews, R. L., Ainslie, A. and Currim, I. S. (2002) 'An empirical comparison of logit choice models with discrete vs. continuous representations of heterogeneity'. *Journal of Marketing Research*, 39: 479-487.

Andrews, R. L., Brusco, M. J. and Currim, I. S. (2010) 'Amalgamation of partitions from multiple segmentation bases: A comparison of non-model-based and model-based methods'. *European Journal of Operational Research*, 201: 608-618.

Andrews, R. L., Brusco, M. J., Currim, I. S., and Davis, B. (2010) 'An empirical comparison of methods for clustering problems: Are there benefits from having a statistical model?', *Review of Marketing Science*, 8: Article 3.

Andrews, R. L. and Currim, I. S. (2003a) 'A comparison of segment retention criteria for finite mixture logit models'. *Journal of Marketing Research*, 40 (2): 235-243.

Andrews, R. L. and Currim, I. S. (2003b) 'Retention of latent segments in regression-based marketing models'. *International Journal of Research in Marketing*, 20 (4): 315-321.

Andrews, R. L., Currim, I. S., and Leeflang, P. (2011) 'A comparison of sales response predictions from demand models applied to store-level versus panel-level data'. *Journal of Business and Economic Statistics*, 29 (2): 319-326.

Atella, V., Brindisi, F., Deb, P. and Rosati, F. C. (2004) 'Determinants of access to physician services in Italy: A latent class seemingly unrelated probit approach'. *Health Economics*, 13: 657-668.

Bekker, P., Merckens, A., and Wansbeek, T. J. (1994) *Identification, equivalent models and computer algebra*, New York: Academic Press.

Blashfield, R. K. (1976) 'Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods', *Psychological Bulletin,* 83: 377-388.

Bozdogan, H. (1987) 'Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions'. *Psychometrika*, 52 (3): 345-370.

Bozdogan, H. (1994) 'Mixture model cluster analysis using selection criteria and a new informational measure of complexity, in Bozdogan, H. (Ed.) *Multivariate Statistical Modelling*, Dordrecht: Kluwer Academic Publishers, 2: 69-113.

Brun, M., Sima, C., Hua, J., Lowey, J., Carroll, B., Suh, E. and Dougherty, E. R. (2007) 'Model-based evaluation of clustering validation measures'. *Pattern Recognition*, 40: 807-824.

Celeux G. (1998) 'Bayesian inference for mixture: The label switching problem' in Payne R. and Green P. (Eds.), *COMPSTAT 98*, Heidelberg: Physica, 227-232.

Celeux, G. and Soromenho, G. (1996) 'An entropy criterion for assessing the number of clusters in a mixture model'. *Journal of Classification,* 13: 195-212.

Celeux, G., Hurn, M., and Robert, C. P. (2000) 'Computational and inferential difficulties with mixture posterior distributions'. *Journal of the American Statistical Association*, 95: 957–970.

Decker, R. and Trusov, M. (2010) 'Estimating aggregate consumer preferences from online product reviews'. *International Journal of Research in Marketing*, 27 (4): 293-307.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the Royal Statistical Society Series B (Methodological)*, 39 (1): 1-38.

Dias, J. (2006) 'Model selection for the binary latent class model: A Monte Carlo simulation' in Batagelj, V., Bock, H.-H., Ferligoj, A., Žiberna, A. (Eds.), *Data Science and Classification*, Berlin: Springer, 91-99.

Dias, J. and Wedel, M. (2004) 'An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods'. *Statistics and Computing*, 14: 323-332.

Dy, J. G. and Brodley, C. E. (2004) 'Feature selection for unsupervised learning'. *Journal of Machine Learning Research*, 5: 845-889.

Ebbes, P., Grewal, R. and DeSarbo, W. S. (2010) 'Modeling strategic group dynamics: A hidden Markov approach'. *Quantive Marketing and Economics,* 8: 241-274.

Formann, A. K. (1984) *Die Latent-Class-Analysis: Eine Einführung in die Theorie und Anwendung*, Weinheim: Beltz.

Frühwirth-Schnatter, S. (2006) *Finite Mixture and Markov Switching Models*, New York: Springer.

Garver, M. S., Williams, Z. and Taylor, G. S. (2008) 'Employing latent class regression analysis to examine logistics theory: An application to truck driver retention'. *Journal of Business Logistics*, 29 (2): 233- 257.

Hahn, C., Johnson, M. D., Hermann, A. and Huber, F. (2002) 'Capturing consumer heterogeneity using a finite mixture PLS approach'. *Schmalenbach Business Review*, 54 (3): 243-269.

Hair, J., Ralph. F., Anderson, E., Tatham, R. L. and William C. Black (1998) *Multivariate Data Analysis,* Englewood Cliffs: Prentice Hall.

Haughton, D., Legrand, P., and Wolford, S. (2009) 'Review of three latent class cluster analysis packages: Latent GOLD, poLCA, MCLUST'. *The American Statistician*, 63 (1): 81-91.

Harzing, A.-W. (2011) *Journal Quality List*, University of Melbourne, available at: www.harzing.com

Henson, J. M., Reise, S. P., Kim, K. H. (2007) 'Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics'. *Structural Equation Modeling*, 14 (2): 202-226.

Hofstede, F. T., Wedel, M., and Steenkamp, J.-B. E.M. (2002) 'Identifying spatial segments in international markets'. *Marketing Science*, 21 (2): 160-177.

Jedidi, K., Jagpal, H. S. and DeSarbo, W. S. (1997) 'Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity'. *Marketing Science*, 16 (1): 39-59.

Law, M. H. C., Figueiredo, M. A. T. and Jain, A. K (2004) 'Simultaneous feature selection and clustering using mixture models'. *IEEE Transactions on Pattern Analysis and Machine Learning*, 26: 1154-1166.

Lubke, G. H. and Muthén, B. (2005) 'Investigating population heterogeneity with factor mixture models'. *Psychological Methods*, 10 (1): 21-39.

Lukočiene, O., Varriale, R., and Vermunt, J. K. (2010) 'The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis'. *Sociological Methodology*, 40 (1): 247-283.

Lukočiene, O. and Vermunt, J. K. (2010) 'Determining the number of components in mixture models for hierarchical data' in Fink, A., Lausen, B., Seidel, W., Ultsch, A. (Eds.), *Advances in Data Analysis, Data Handling and Business Intelligence*, Berlin: Springer, 241-249.

McLachlan, G. and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

Melnykov, V. and Maitra, R. (2010) 'Finite mixture models and model-based clustering'. *Statistic Surveys*, 4: 80-116.

Netzer, O., Lattin, J. M. and Srinivasan, V. (2008) 'A hidden Markov model of customer relationship dynamics'. *Marketing Science*, 27 (2): 185–204.

Raftery, A. E. and Dean, N. (2006) 'Variable selection for model-based clustering'. *Journal of the American Statistical Association*, 101: 294-304.

Rust, R. T. and Verhoef, P. C. (2005) 'Optimizing the marketing interventions mix in intermediate-term CRM'. *Marketing Science*, 24 (3): 477-489.

Sarstedt, M. (2008) 'Market segmentation with mixture regression models: Understanding measures that guide model selection'. *Journal of Targeting, Measurement and Analysis for Marketing*, 16: 228-246.

Sarstedt, M., Becker, J.-M, Ringle, C. M. and Schwaiger, M. (2011) 'Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments?'. *Schmalenbach Business Review*, 63 (1): 34-62.

Sarstedt, M. and Schwaiger, M. (2007) 'Model selection in mixture regeression analysis – A Monte Carlo simulation study', in Preisach, C., Burkhardt, H., Schmidt-Thieme, L. and Decker, R. (Eds), *Data Analysis, Machine Learning, and Applications*, Berlin: Springer, 61-68.

Schwarz, G. (1978) 'Estimating the dimension of a model'. *The Annals of Statistics*, 6 (2): 461-464.

Sclove, S. L. (1987) 'Application of model-selection criteria to some problems in multivariate-analysis'. *Psychometrika*, 52: 333-43.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) 'Bayesian measures of model complexity and fit'. *Journal of the Royal Statistical Society, Series B*, 64, 583-640.

Steinley, D., and Brusco, M. J. (2008a) 'A new variable weighting and selection procedure for *k*-means cluster analysis'. *Multivariate Behavioral Research,* 43: 77-108.

Steinley, D., and Brusco, M. J. (2008b) 'Selection of variables in cluster analysis: An empirical comparison of eight procedures'. *Psychometrika,* 73: 125-144.

Stephens, M. (1997) *Bayesian Methods for Mixtures of Normal Distributions*, Dissertation, University of Oxford.

Tuma, M. N., Decker, R. and Scholz, S. W. (2011) 'A survey of the challenges and pitfalls of cluster analysis application in market segmentation'. *International Journal of Market Research*, 53 (3): 391-414.

Wedel, M. and DeSarbo, W. S. (2002) 'Market segment derivation and profiling via a finite mixture model framework'. *Marketing Letters*, 13 (1): 17-25.

Wedel, M. and Kamakura, W. A. (2000) *Market Segmentation: Conceptual and Methodological Foundations*, Dordrecht: Kluwer.

Equidistance of Likert-Type Scales and Validation of Inferential Methods Using Experiments and Simulations

Bjorn Lantz
School of Engineering, University of Boras, Sweden
Bjorn.Lantz@hb.se

Abstract: Likert-type data are often assumed to be equidistant by applied researchers so that they can use parametric methods to analyse the data. Since the equidistance assumption rarely is tested, the validity of parametric analyses of Likert-type data is often unclear. This paper consists of two parts where we deal with this validity problem in two different respects. In the first part, we use an experimental design to show that the perceived distance between scale points on a regular five-point Likert-type scale depends on how the verbal anchors are used. Anchors only at the end points create a relatively larger perceived distance between points near the ends of the scale than in the middle (*end-of-scale effect*), while anchors at all points create a larger perceived distance between points in the middle of the scale (*middle-of-scale effect*). Hence, Likert-type scales are generally not perceived as equidistant by subjects. In the second part of the paper, we use Monte Carlo simulations to explore how parametric methods commonly used to compare means between several groups perform in terms of actual significance and power when data are assumed to be equidistant even though they are not. The results show that the preferred statistical method to analyse Likert-type data depends on the nature of their non-equidistance as well as their skewness. Under middle-of-scale effect, the omnibus one-way ANOVA works best when data are relatively symmetric. However, the Kruskal-Wallis test works better when data are skewed except when sample sizes are unequal, in which case the Brown-Forsythe test is better. Under end-of-scale effect, on the other hand, the Kruskal-Wallis test should be preferred in most cases when data are at most moderately skewed. When data are heavily skewed, ANOVA works best unless when sample sizes are unequal, in which case the Brown-Forsythe test should be preferred.

Keywords: Likert-type scale; equidistance; Monte Carlo simulation; ANOVA; Kruskal-Wallis test; Brown-Forsythe test; Welch test

1. Introduction

Since the psychologist Rensis Likert (Likert, 1932) published his seminal work on measurement of attitudes, Likert items (often referred to as 'Likert-type scales') and true Likert scales have been important data collection methodologies in research on attitudes and opinions in the social sciences in general. In business and management, Likert-type scales are often used by researchers to collect data (Alexandrov, 2010). Even though the optimal number of steps has been debated over the years (Pearse, 2011), the classical Likert item based on a statement where the subjects are asked to choose one out of five possible degrees of agreement, ranging from 'strongly agree' to 'strongly disagree', that complies with their view on the matter, still seem to the most common choice among researchers. According to the traditional classification of measurement scales (Stevens, 1946), Likert-type scales must be either ordinal or interval, depending on whether or not the scale is equidistant, since rank-ordering is possible while a true zero point is missing. Whether Likert-type scales should generally be regarded as ordinal or interval has been extensively debated among researchers over the years (e.g. Carifio and Perla, 2007; Jamieson, 2004; Michell, 1986), since the choice of statistical methodology depends on the result of this debate. According to the standard textbook view, data having only ordinal properties should be analysed with non-parametric statistics based on ranks. However, parametric statistics that are more powerful are allowed for data with interval properties, utilising the actual values of the data instead of just their ranks.

To be able to use parametric statistics on ordinal data, several different methods for 'rescaling' ordinal scales to get interval properties have been proposed (e.g. Granberg-Rademacker, 2010; Wu, 2007; King et al., 2004; Harwell and Gatti, 2001; Bendixen and Sandler, 1995). The use of such methods in applied research seems rare in practical analysis of Likert-type data, perhaps because it appears improbable that an ordinal scale with unknown distance between the scale points could actually be rescaled to a true interval scale, or due to a concern that the informational content of the data will be affected by the procedure. Instead, the vast majority of researchers about to analyse Likert-type data seem to either obey the purist view on statistics and open the non-parametric toolbox or trust the methodology researchers who claim that t and F are statistics robust to minor violations in the underlying assumptions. To be allowed to use parametric methods, many researchers simply assume that their scale has interval properties (Albaum, 1997) while a few put a lot of

effort into showing that their scale demonstrates acceptable quality as an interval measurement device (e.g. Albaum et al, 1977). The main problem, however, is that presence of interval properties is a necessary but not a sufficient condition for statistical analyses with parametric methodologies. Assumptions regarding normality and homoscedasticity must also be addressed.

There has been a considerable amount of research during the years testing the robustness of parametric methods like the t-test and the analysis of variance (ANOVA), often based on Monte Carlo approaches, against violations of the normality and the homoscedasticity assumptions. For example, Glass et al. (1972) found that skewness only had a small effect on the efficiency of ANOVA. Feir and Toothaker (1974) compared ANOVA with the Kruskal-Wallis test in a Monte Carlo study and concluded, based on 'the instability of power for the Kruskal-Wallis test', that ANOVA was the recommended procedure even when normality and/or homoscedasticity is doubtful. Zimmerman (1998) evaluated the Wilcoxon test against the t-test in a simulation study under concurrent violation of two assumptions, namely normality and homoscedasticity. He showed that nonparametric methods do not generally provide protection against concurrent violation of normality and homoscedasticity. Under some conditions, the Wilcoxon test would even make the situation worse. Lantz (2012) showed that parametric methods are generally more sensitive to different degrees of sample non-normality when populations are distinctly non-normal. He concluded that the Kruskal-Wallis test should be preferred as soon as the underlying populations are not known to be normal or approximately normal in order to avoid a preliminary test for normality that makes the overall level of significance unclear. However, despite extensive search, we have been unable to find any study that assesses the violation of the equidistance assumption while using parametric methods from a robustness perspective.

When researchers are in the process of choosing a statistical methodology to analyse Likert-type data, they should consider the way subjects perceive the response scale. If the scale is perceived as equidistant, parametric methods can obviously be used to analyse the data. If not, the purist view on statistics requires a non-parametric methodology. Research based on rescaling of ordinal data indicate that subjects actually do perceive Likert-type scales as non-equidistant, at least for specific constructs (e.g. Lee and Soutar, 2010; Mundy and Dickinson, 2004; Kennedy, Riquier, and Sharp, 1996; Bendixen and Sandler, 1995). Hence, the purpose of this study is to explore whether subjects generally perceive the five-point Likert-type scale as non-equidistant, and to examine how non-equidistance can affect the choice of statistical method for analysing Likert-type data.

The remainder of the paper is organised as follows. In the next section, two experiments are conducted in order to explore subject perceptions of the five-point Likert-type scale. Thereafter, a simulation study compares the performance of one-way ANOVA with alternative methods under different types of non-equidistance, followed by the conclusion.

2. Experiments

There are several obvious problems related to the process of assigning values to the points (or the distances between points) on a scale that is qualitative rather than quantitative by nature. Both the subjective distances between the scale points and the subjective zero points may differ between occasions and/or between respondents. They may even shift within the occasion due to respondent stimulus. In addition, the subjective distances between scale points may differ between subsets of the scale. Because of these problems, it is meaningless to assume that the distance between different points on the scale is measurable in absolute terms in order to be able to make a claim about the absolute distance between the points on the scale. In general, if the premise (e.g. it is actually possible to measure the absolute distance between points on a scale) cannot be proven valid, the result (e.g. measured absolute distances between points on the scale) may be invalid. For this reason, we did not assume that the absolute distances between the scale point is possible to measure. Instead, we only assumed that respondents can compare pair-wise changes in opinion represented by different movements between points on a five-point response scale within a Likert item, and express whether they perceive one change in opinion as greater than the other, or if they view them as equal.

The response scale formats under study here were technically five-point discrete visual analogue scales. Verbal anchors were used at the ends as illustrated in figure 1a in the first experiment, and at all points as illustrated in figure 1b in the second experiment, in order to examine whether respondents perceive the scale differently when they have to attach their personal interpretation of the different scale points based on only the numbers

identifying them. Providing consecutive integers for the scale points was also assumed to maximize the likelihood that respondents actually would perceive the scale as interval, which meant that we would have a stronger case if the interval hypothesis were rejected.

	Strongly disagree				Strongly agree
Statement	1	2	3	4	5

Figure 1a: Verbal anchors only at the end points

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Statement	1	2	3	4	5

Figure 1b: Verbal anchors at all points

Two convenience samples of 178 and 101 respondents were recruited for participation in the experiments. No individual demographic or other background data were collected; however, all respondents were regular first- or second-year business students with an approximately even distribution between males and females. A majority of respondents were in their twenties. Hence, as always in experiments where students are used as subjects, the validity of the results rests on the assumption that students are representative of 'real people' (e.g., Cunningham et al., 1974). All respondents answered the questionnaire anonymously. The instructions specified that there was no 'correct' answer to any of the problems and that the aim of the study was to explore how people perceive this kind of response scale. All respondents were presented with a questionnaire where the scale was displayed and followed by nine different problems of the type illustrated in figure 2.

Compare a movement from 1 to 2 on the scale with a movement from 3 to 4 on the scale. Which one of them do you think represents a greater change in the level of agreement? Tick the appropriate box.

- ☐ 1 to 2 is greater
- ☐ 3 to 4 is greater
- ☐ No difference

Figure 2: Example of a problem

To ensure reliability, several versions of the questionnaire were used so that different respondents were exposed to the problems and the options in different orders using a randomized process. The direction of the response scale was also varied randomly. No significant differences were found between the groups of respondents using different versions of the questionnaire.

The data from the problems in both experiments were analysed with respect to two different null hypotheses:

H1: A majority of the subjects do not perceive a difference.

H2: Uniform distribution characterises the two types of perceived differences.

The results from experiment 1, where only the end points on the scale had verbal anchors, are presented in table 1. Note that, for clarity, the nine comparisons are presented in four groups with different characteristics.

Group I consists of the three symmetric problems. In group II, there are two asymmetric problems with one-point movements on each side of the scale. Group III consists of the two asymmetric problems with two-point movements on each side of the scale. Finally, in group IV, we have the two possible one-point movements on the same side of the scale.

Table 1: Results from experiment 1 with verbal anchors only at the end points

Group	Movement comparison		Frequency (percentage)			H1	H2	
	A	B	A is greater	B is greater	No difference	p-value*	p-value*	Effect size h
I	1 to 2	4 to 5	43 (24.2%)	40 (22.5%)	95 (53.4%)	0.815	0.742	
	1 to 3	3 to 5	36 (20.2%)	40 (22.5%)	102 (57.3%)	0.973	0.647	
	2 to 3	3 to 4	29 (16.3%)	26 (14.6%)	113 (63.5%)	0.999	0.686	
II	1 to 2	3 to 4	79 (44.4%)	49 (27.5%)	50 (28.1%)	<0.001	0.010	0.47
	2 to 3	4 to 5	52 (29.2%)	78 (43.8%)	48 (27.0%)	<0.001	0.025	0.40
III	1 to 3	2 to 4	85 (47.8%)	52 (29.2%)	41 (23.0%)	<0.001	0.006	0.49
	2 to 4	3 to 5	50 (28.1%)	79 (44.4%)	49 (27.5%)	<0.001	0.013	0.45
IV	1 to 2	2 to 3	84 (47.2%)	43 (24.2%)	51 (28.7%)	<0.001	0.001	0.66
	3 to 4	4 to 5	51 (28.7%)	78 (43.8%)	49 (27.5%)	<0.001	0.020	0.42
* Based on regular two-sample z-tests								

For all three problems in group I, the symmetric comparisons, the results indicate that a majority of subjects do not perceive a difference. Hence, they experience the scale as symmetrical, which we refer to as the *symmetry effect*. For group II with the asymmetric one-step comparisons, a majority of the subjects do perceive differences. Specifically, significantly more subjects perceive 1–2 as larger than 3–4. We see the same effect on the other side of the scale where significantly more subjects perceive 4–5 as larger than 2–3. Besides the symmetry effect, these results also indicate that respondents tend to think that the distance between scale points is greater near an end of the scale than near the middle. We will refer to this as the *end-of-scale effect*. For the asymmetric two-step comparisons in group III, the same pattern emerges as in group II. A majority of the subjects perceive differences, and significantly more subjects perceive 1–3 as larger than 2–4, while significantly more subjects perceive 3–5 as larger than 2–4. Thus, both the end-of-scale effect and the symmetry effect can be seen in these data. Finally, in group IV, a majority of the subjects perceive differences while both the end-of-scale effect and the symmetry effect are visible. On the lower part of the scale, significantly more subjects perceive 1–2 as larger than 2–3. Mirroring this, significantly more subjects perceive 4–5 as larger than 3–4. Again, the distance between points near the end of the scale is experienced as greater changes in opinion than between points in the middle of the scale. In addition, this effect exists in a symmetrical manner on both sides of the scale.

Table 2 presents the results from experiment 2, where all points on the scale had verbal anchors. Again, the nine comparisons are divided into four groups with different characteristics.

Table 2: Results from experiment 2 with verbal anchors at all points

Group	Movement comparison		Frequency (percentage)			H1	H2	
	A	B	A is greater	B is greater	No difference	p-value*	p-value*	Effect size h
I	1 to 2	4 to 5	19 (18,8 %)	27 (26,7 %)	55 (54,5 %)	0.814		
	1 to 3	3 to 5	16 (15,8 %)	30 (29,7 %)	55 (54,5 %)	0.814		
	2 to 3	3 to 4	19 (18,8 %)	18 (17,8 %)	64 (63,4 %)	0.995		
II	1 to 2	3 to 4	24 (23,8 %)	62 (61,4 %)	15 (14,9 %)	<0.001	<0.001	0.92
	2 to 3	4 to 5	68 (67,3 %)	24 (23,8 %)	9 (8,9 %)	<0.001	<0.001	1.00
III	1 to 3	2 to 4	23 (22,8 %)	65 (64,4 %)	13 (12,9 %)	<0.001	<0.001	1.00
	2 to 4	3 to 5	51 (50,5 %)	32 (31,7 %)	18 (17,8 %)	<0.001	0.042	0.46
IV	1 to 2	2 to 3	25 (24,8 %)	61 (60,4 %)	15 (14,9 %)	<0.001	<0.001	0.86
	3 to 4	4 to 5	55 (54,5 %)	28 (27,7 %)	18 (17,8 %)	<0.001	0.005	0.66
* Based on regular two-sample z-tests								

As in experiment 1, there were significant differences between the two types of movements in all cases where a significant minority perceived no difference. For all three problems in group I, the symmetric comparisons,

the results in experiment 2 were similar to the results in experiment 1. Hence, the perceived symmetry of the scale does not seem to be affected by the way the verbal anchors were used. In group II, a majority of subjects perceived differences. However, the differences are not of the same nature as in experiment 1. Significantly more subjects now perceive 3–4 as larger than 1–2, and 2–3 as larger than 4–5, in contrast to the previous situation when only the end points had verbal anchors. Similar changes can be seen in groups III and IV, where the movements between points near the middle of the scale are perceived as larger than movements near the end of the scale by significantly more subjects. Hence, the results in experiment 2 differ significantly from the results in experiment 1 since movements between points in the middle of the scale are experienced as greater changes in opinion than movements between points near an end of the scale in experiment 2. We will refer to this as the *middle-of-scale effect*. It should also be noted that the effect sizes were generally larger when all points on the scale had verbal anchors.

The main observation from these experiments is that there are systematic differences in the way respondents perceive the five-point Likert-type scale, depending on how verbal anchors are used. A possible explanation for why subjects perceive a scale with verbal anchors at the end points differently from a scale with verbal anchors at all points is that with the latter, it becomes more obvious that while a movement in the middle of the scale signals a change of opinion, a movement at the end of the scale only signals a change of intensity within the same opinion. With verbal anchors only at the ends, the subjects have to imagine the meaning of the other scale points themselves. Hence, without verbal anchors, it is not obvious that a movement between, for example, 2 and 4 is equivalent to a change in opinion.

There has been some previous research dealing with the effects of verbal anchors that these results can be related to (see Weijters et al., 2010, for a review). For example, verbal anchors are often assumed to make scale points more salient, which might attract respondents (Krosnick and Fabrigar, 1997). Hence, verbal anchors on intermediate points would create a shift towards those points at the expense of the extreme points (Simonson, 1989). The end-of-scale effect may also be seen as an explanation of the well-known central tendency bias (e.g. James, Demaree, and Wolf, 1984) in surveys. The relatively large perceived distance between points at the end of the scale makes it relatively hard to reach an end point when verbal anchors are used only at the end points. The perceived distance between points in the middle of the scale is smaller, which makes it easier to move between them.

The variance for a set of data of this type also depends on the way the verbal anchors are used, which should be taken into account in the statistical analysis. A rank-based test procedure (e.g. the Kruskal-Wallis test) is obviously immune to potential scale effects of the type we have discussed here; however, a parametric test procedure (e.g. the one-way ANOVA) is not. It is easy to see that assigning, for example, the values {0, 1.5, 2, 2.5, 4}, representing end-of-scale effect, to the five scale points will create a lower variance than, for example, the values {0, 0.5, 2, 3.5, 4}, representing middle-of-scale effect. Parametric test procedures are obviously affected by this problem.

Since neither of the two types of verbal anchor use creates a response scale that is perceived as equidistant, there are potential validity issues related to the choice of statistical methodology used in the analyses of data collected with Likert-type scales. In the next section, we will use a Monte Carlo approach to evaluate these issues.

3. Simulations

Design

An experimental design with three populations (k = 3) and four different combinations of small (defined as n = 5) and large (defined as n = 25) sample sizes were used. The simulations were based on random numbers from binomial distributions with a sample space {0, 1, 2, 3, 4} to represent five-level Likert-items, where $\mu_1 \leq \mu_2 \leq \mu_3$ and $|\mu_1 - \mu_2| = |\mu_3 - \mu_2|$. Symmetry was defined as a situation where the second population had a true mean value of 2. Moderate skewness was assumed to correspond with a true mean value of 1 for the second population, and severe skewness with a true mean value of 0.5. The three cases are illustrated in figure 3, with a symmetric distribution to the left, a moderately skewed distribution in the middle, and a severely skewed distribution to the right.

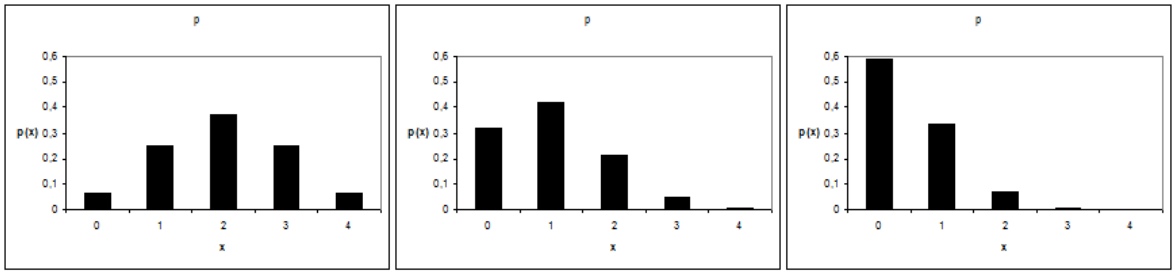

Figure 3: Symmetry, moderate skewness, and severe skewness

Table 3 shows the manner in which the true mean values of the distributions were shifted to achieve a suitable range of effect sizes (Cohen, 1992), ranging from no effect (f = 0) to a very large effect (f = 0.65). All mean values were calculated using G*Power version 3.1.2 (Faul et al., 2007). Note that the first population in the heavily skewed case with a very large effect size is technically not based on the binomial distribution as it consists exclusively of zeros.

Table 3: True mean values and implied effect sizes

Symmetrical			Moderately skewed			Heavily skewed			Effect size
μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	
2.000	2.000	2.000	1.000	1.000	1.000	0.500	0.500	0.500	f = 0.00
1.877	2.000	2.123	0.894	1.000	1.106	0.419	0.500	0.581	f = 0.10
1.696	2.000	2.304	0.737	1.000	1.263	0.299	0.500	0.701	f = 0.25
1.519	2.000	2.481	0.583	1.000	1.417	0.182	0.500	0.818	f = 0.40
1.243	2.000	2.757	0.344	1.000	1.656	0.000	0.500	1.000	f = 0.65

Four different test procedures were examined; the omnibus one-way ANOVA, the Kruskal-Wallis test (Kruskal and Wallis, 1952), the Brown-Forsythe test (Brown and Forsythe, 1974), and the Welch test (Welch, 1951). Every combination of test procedure and parent distribution was evaluated for every combination of sample sizes and effect size. For each combination, 3 × 50,000 sets of random numbers were generated, and the null hypothesis that corresponds to no difference between the locations of the populations was challenged at an alpha level of 0.05. The end-of-scale effect was simulated by adjusting the original scale to {0, 1.5, 2, 2.5, 4} and the middle-of-scale effect by adjusting it to {0, 0.5, 2, 3.5, 4}. All simulation procedures were conducted using Microsoft Excel 2010.

The null hypothesis of no difference in the true proportion of rejected tests between the four procedures for a certain combination was evaluated with chi-square tests in all cases. In cases where this null hypothesis was rejected, post hoc analysis was performed using pair-wise chi-square tests with Bonferroni correction.

Results

Table 4 shows the simulation results in the symmetric case with equidistant data. At (25, 25, 25), there is a significant difference between the methods only when the effect size is medium, where ANOVA (A) and the Brown-Forsythe test (BF) both are significantly more powerful than the Kruskal-Wallis test (KW). At (5, 5, 5), the methods differ significantly at all effect sizes. The overall picture from the detailed analyses is that A works best and that the Welch test (W) should be avoided in these circumstances. We also note that all methods except A are quite conservative and generate very few type I errors when all sample sizes are small. At (5, 5, 25), A is superior as well; however, it should be noted that KW dominates both BF and W at medium and larger effect sizes. At (5, 25, 25), KW has the lowest power when the effect size is small; however, when the effect size becomes larger, W performs worse while A and KW are more powerful. Hence, in the symmetric case with equidistant data, A should be recommended as it dominates the other three methods for many combinations of effect sizes and sample sizes, especially where at least one sample size is small, and is never dominated by any of them. This is obviously in line with what one could have expected according to theory, as data are metric and approximate homoscedasticity prevails in the symmetric case.

Table 5 shows the results from the symmetric case where the end-of-scale (EOS) effect is present. At (25, 25, 25), KW dominates W when the effect size is small. Furthermore, KW dominates all the other methods at

medium and large effect sizes. The tendency is similar, but reinforced, when at least one sample size is small: KW dominates the other methods, and BF and W are weak. Moreover, when at least one sample size is small, BF and W are quite conservative, which is the probable reason for their low power. Hence, in the symmetric case when the EOS effect is present, KW should be the recommended method as it dominates the other three methods for most combinations of effect size and sample sizes. Note that the percentages for KW are the same as in the equidistant case above (and as in the middle-of-scale case below) as KW is not affected by non-equidistance since the ranks stay the same regardless of the severity of the equidistance assumption violation. Thus, the differences in percentages for the other methods illustrate the impact of the EOS effect when data are incorrectly assumed to be equidistant.

Table 4: Proportion of significant tests, symmetric data and equidistance

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0503	0.0484	0.0499	0.0504	0.492
$n_2 = 25$	0.10	0.1105	0.1055	0.1099	0.1088	0.071
$n_3 = 25$	0.25	0.4605	0.4462	0.4598	0.4541	0.002
	0.40	0.8655	0.8564	0.8653	0.8602	0.343
	0.65	0.9992	0.9991	0.9992	0.9991	0.999
$n_1 = 5$	0.00	0.0488	0.0398	0.0428	0.0369	< 0.001
$n_2 = 5$	0.10	0.0584	0.0486	0.0514	0.0453	< 0.001
$n_3 = 5$	0.25	0.1088	0.0926	0.0964	0.0857	< 0.001
	0.40	0.2135	0.1838	0.1954	0.1725	< 0.001
	0.65	0.4878	0.4378	0.4584	0.4112	< 0.001
$n_1 = 5$	0.00	0.0516	0.0447	0.0492	0.0478	< 0.001
$n_2 = 5$	0.10	0.0701	0.0619	0.0650	0.0636	< 0.001
$n_3 = 25$	0.25	0.1927	0.1710	0.1518	0.1474	< 0.001
	0.40	0.4335	0.3890	0.3290	0.3162	< 0.001
	0.65	0.8455	0.8000	0.7009	0.6610	< 0.001
$n_1 = 5$	0.00	0.0513	0.0474	0.0557	0.0542	< 0.001
$n_2 = 25$	0.10	0.0751	0.0699	0.0789	0.0753	< 0.001
$n_3 = 25$	0.25	0.2297	0.2132	0.2154	0.1967	< 0.001
	0.40	0.5160	0.4915	0.4695	0.4398	< 0.001
	0.65	0.9100	0.8958	0.8574	0.8401	< 0.001

Table 5: Proportion of significant tests, symmetric data and end-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0499	0.0484	0.0493	0.0458	0.018
$n_2 = 25$	0.10	0.1015	0.1055	0.1003	0.0970	< 0.001
$n_3 = 25$	0.25	0.4163	0.4462	0.4140	0.4065	< 0.001
	0.40	0.8242	0.8564	0.8225	0.8167	< 0.001
	0.65	0.9973	0.9991	0.9973	0.9970	0.987
$n_1 = 5$	0.00	0.0336	0.0398	0.0206	0.0144	< 0.001
$n_2 = 5$	0.10	0.0418	0.0486	0.0257	0.0179	< 0.001
$n_3 = 5$	0.25	0.0789	0.0926	0.0516	0.0315	< 0.001
	0.40	0.1627	0.1838	0.1175	0.0656	< 0.001
	0.65	0.4085	0.4378	0.3343	0.1981	< 0.001
$n_1 = 5$	0.00	0.0546	0.0447	0.0262	0.0236	< 0.001
$n_2 = 5$	0.10	0.0704	0.0619	0.0395	0.0358	< 0.001
$n_3 = 25$	0.25	0.1627	0.1710	0.1212	0.0982	< 0.001
	0.40	0.3590	0.3890	0.2989	0.2496	< 0.001
	0.65	0.7620	0.8000	0.6637	0.6275	< 0.001
$n_1 = 5$	0.00	0.0543	0.0474	0.0420	0.0358	< 0.001
$n_2 = 25$	0.10	0.0726	0.0699	0.0580	0.0518	< 0.001
$n_3 = 25$	0.25	0.2065	0.2132	0.1709	0.1490	< 0.001
	0.40	0.4631	0.4915	0.4049	0.3640	< 0.001
	0.65	0.8656	0.8958	0.8207	0.7925	< 0.001

Table 6 shows the results from the symmetric case where the middle-of-scale (MOS) effect is present. At (25, 25, 25), there are no significant differences in performance between the methods; however, at (5, 5, 5) A dominates all the other methods at all effect sizes while KW generally performs worse than both BF and W. At (5, 5, 25), A continues to dominate the other methods at all effect sizes, while KW is now significantly more powerful than both BF and W at medium and larger effect sizes. At (5, 25, 25), BF and W are quite liberal and generate many type I errors, which also explains the superiority of BF and W at the small effect size. At larger effect sizes, BF and W, again, have less power than both A and KW, which in turn do not differ significantly. Hence, the MOS effect seems to affect BF and W more than it affects A. Therefore, in the symmetric case when the MOS effect is present, A should be recommended as it dominates the other three methods for most combinations of effect sizes and sample sizes.

Table 7 shows the results from the case where data are moderately skewed but equidistant. At (25, 25, 25), the only significant difference is that KW has less power at the medium effect size. At (5, 5, 5), all methods except A are excessively conservative, and A is more powerful while W is less powerful than KW and BF at all effect sizes. At (5, 5, 25) and (5, 25, 25), BF and W are more powerful than A and KW in most cases except for the very large effect size where W loses power, while KW is in turn generally more powerful than A. Hence, where data are moderately skewed but equidistant, A should be the recommended method under equal sample sizes as it is at least as powerful as any of the other methods in all situations. On the other hand, under unequal sample sizes, A should be avoided as it is the weakest of the methods in most cases. Here, BF must be recommended as it is the method with the highest power in most cases.

Table 6: Proportion of significant tests, symmetric data and middle-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0512	0.0484	0.0510	0.0511	0.156
$n_2 = 25$	0.10	0.1086	0.1055	0.1083	0.1082	0.391
$n_3 = 25$	0.25	0.4467	0.4462	0.4459	0.4435	0.878
	0.40	0.8529	0.8564	0.8525	0.8499	0.738
	0.65	0.9989	0.9991	0.9989	0.9987	0.999
$n_1 = 5$	0.00	0.0528	0.0398	0.0436	0.0458	< 0.001
$n_2 = 5$	0.10	0.0641	0.0486	0.0534	0.0554	< 0.001
$n_3 = 5$	0.25	0.1161	0.0926	0.0985	0.1034	< 0.001
	0.40	0.2225	0.1838	0.1918	0.2019	< 0.001
	0.65	0.4907	0.4378	0.4474	0.4549	< 0.001
$n_1 = 5$	0.00	0.0498	0.0447	0.0533	0.0610	< 0.001
$n_2 = 5$	0.10	0.0704	0.0619	0.0671	0.0785	< 0.001
$n_3 = 25$	0.25	0.1925	0.1710	0.1508	0.1625	< 0.001
	0.40	0.4327	0.3890	0.3167	0.3267	< 0.001
	0.65	0.8440	0.8000	0.6782	0.6457	< 0.001
$n_1 = 5$	0.00	0.0489	0.0474	0.0621	0.0623	< 0.001
$n_2 = 25$	0.10	0.0750	0.0699	0.0846	0.0849	< 0.001
$n_3 = 25$	0.25	0.2219	0.2132	0.2161	0.2059	< 0.001
	0.40	0.5023	0.4915	0.4590	0.4431	< 0.001
	0.65	0.8999	0.8958	0.8417	0.8289	< 0.001

Table 7: Proportion of significant tests, moderately skewed data and equidistance

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0510	0.0486	0.0503	0.0496	0.364
$n_2 = 25$	0.10	0.1089	0.1046	0.1083	0.1071	0.168
$n_3 = 25$	0.25	0.4609	0.4453	0.4587	0.4589	< 0.001
	0.40	0.8714	0.8585	0.8702	0.8706	0.089
	0.65	0.9993	0.9990	0.9993	0.9992	0.999
$n_1 = 5$	0.00	0.0500	0.0423	0.0436	0.0317	< 0.001
$n_2 = 5$	0.10	0.0592	0.0513	0.0511	0.0393	< 0.001
$n_3 = 5$	0.25	0.1064	0.0929	0.0929	0.0714	< 0.001
	0.40	0.2100	0.1876	0.1838	0.1417	< 0.001

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
	0.65	0.4862	0.4629	0.4294	0.2950	< 0.001
$n_1 = 5$	0.00	0.0487	0.0431	0.0514	0.0458	< 0.001
$n_2 = 5$	0.10	0.0542	0.0591	0.0784	0.0767	< 0.001
$n_3 = 25$	0.25	0.1391	0.1516	0.1890	0.1847	< 0.001
	0.40	0.3397	0.3639	0.3925	0.3831	< 0.001
	0.65	0.7885	0.8055	0.7824	0.6599	< 0.001
$n_1 = 5$	0.00	0.0488	0.0456	0.0551	0.0520	< 0.001
$n_2 = 25$	0.10	0.0652	0.0672	0.0882	0.0844	< 0.001
$n_3 = 25$	0.25	0.1916	0.1952	0.2444	0.2230	< 0.001
	0.40	0.4523	0.4573	0.5231	0.4625	< 0.001
	0.65	0.8792	0.8818	0.9113	0.7315	< 0.001

Table 8 shows the results from the case where data are moderately skewed and the EOS effect is present. At (25, 25, 25), KW is significantly more powerful than any other method at medium and large effect sizes. At (5, 5, 5), all methods except A are too conservative, an effect that stays as the effect size becomes larger. W is less powerful than KW and BF at all effect sizes, and KW in turn dominates BF. At (5, 5, 25) and (5, 25, 25), W is exceedingly liberal while A is more powerful than the other methods in most cases except KW at large effect sizes. KW is also generally more powerful than BF. Hence, when all sample sizes are large, KW should be recommended if data are moderately skewed and the EOS effect is present. However, A should be the preferred method when at least one sample size is small as it then dominates the other three methods for most combinations of effect sizes and sample sizes.

Table 8: Proportion of significant tests, moderately skewed data and end-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0504	0.0486	0.0501	0.0514	0.282
$n_2 = 25$	0.10	0.1051	0.1046	0.1043	0.1038	0.936
$n_3 = 25$	0.25	0.4225	0.4453	0.4213	0.4174	< 0.001
	0.40	0.8365	0.8585	0.8357	0.8316	< 0.001
	0.65	0.9985	0.9990	0.9985	0.9984	0.999
$n_1 = 5$	0.00	0.0506	0.0423	0.0347	0.0300	< 0.001
$n_2 = 5$	0.10	0.0599	0.0513	0.0428	0.0370	< 0.001
$n_3 = 5$	0.25	0.1054	0.0929	0.0790	0.0681	< 0.001
	0.40	0.2089	0.1876	0.1678	0.1370	< 0.001
	0.65	0.4958	0.4629	0.4334	0.2947	< 0.001
$n_1 = 5$	0.00	0.0465	0.0431	0.0474	0.0971	< 0.001
$n_2 = 5$	0.10	0.0736	0.0591	0.0538	0.0684	< 0.001
$n_3 = 25$	0.25	0.1853	0.1516	0.1136	0.0886	< 0.001
	0.40	0.4018	0.3639	0.2721	0.1845	< 0.001
	0.65	0.8167	0.8055	0.6965	0.4426	< 0.001
$n_1 = 5$	0.00	0.0477	0.0456	0.0608	0.1121	< 0.001
$n_2 = 25$	0.10	0.0757	0.0672	0.0715	0.0914	< 0.001
$n_3 = 25$	0.25	0.2079	0.1952	0.1705	0.1439	< 0.001
	0.40	0.4596	0.4573	0.3917	0.3098	< 0.001
	0.65	0.8731	0.8818	0.8314	0.6245	< 0.001

Table 9 shows the results from the case where data are moderately skewed and the MOS effect is present. At (25, 25, 25), KW is significantly more powerful than any other method at medium and large effect sizes, exactly as under the EOS effect. At (5, 5, 5), BF and W are very conservative, and A and KW have more power than BF and W at all effect sizes. When the effect size is large or very large, KW also has more power than A. At (5, 5, 25) and (5, 25, 25), W becomes excessively liberal, which suggests that it should be avoided. It should also be noted that the MOS effect actually reduces the percentage of significant As at (5, 5, 25) when the effect size increases from none to a small size. However, BF generally has more power than both A and KW without resulting in a large number of type I errors. KW is in turn also more powerful than A. Hence, when data are moderately skewed and the MOS effect is present, KW should be the preferred choice when sample sizes are equal and BF when sample sizes are unequal.

Table 9: Proportion of significant tests, moderately skewed data and middle-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0498	0.0486	0.0486	0.0535	0.001
$n_2 = 25$	0.10	0.1019	0.1046	0.0998	0.1064	0.006
$n_3 = 25$	0.25	0.4197	0.4453	0.4164	0.4260	< 0.001
	0.40	0.8322	0.8585	0.8298	0.8357	< 0.001
	0.65	0.9982	0.9990	0.9981	0.9984	0.999
$n_1 = 5$	0.00	0.0435	0.0423	0.0202	0.0150	< 0.001
$n_2 = 5$	0.10	0.0491	0.0513	0.0240	0.0201	< 0.001
$n_3 = 5$	0.25	0.0893	0.0929	0.0469	0.0396	< 0.001
	0.40	0.1689	0.1876	0.0967	0.0840	< 0.001
	0.65	0.3809	0.4629	0.2437	0.1922	< 0.001
$n_1 = 5$	0.00	0.0463	0.0431	0.0585	0.1266	< 0.001
$n_2 = 5$	0.10	0.0313	0.0591	0.0961	0.2174	< 0.001
$n_3 = 25$	0.25	0.0684	0.1516	0.2160	0.4026	< 0.001
	0.40	0.1956	0.3639	0.4069	0.5755	< 0.001
	0.65	0.6178	0.8055	0.7332	0.7021	< 0.001
$n_1 = 5$	0.00	0.0466	0.0456	0.0546	0.1541	< 0.001
$n_2 = 25$	0.10	0.0503	0.0672	0.0915	0.2374	< 0.001
$n_3 = 25$	0.25	0.1433	0.1952	0.2575	0.4028	< 0.001
	0.40	0.3581	0.4573	0.5365	0.5794	< 0.001
	0.65	0.7878	0.8818	0.9089	0.7390	< 0.001

Table 10 shows the results from the case where data are heavily skewed but equidistant. At (25, 25, 25), W is somewhat more liberal than BF and more powerful at small or medium effect sizes. KW is less powerful than all other methods at medium effect size. At (5, 5, 5), A and KW are more powerful than BF and W at all effect sizes, and A is also more powerful than KW at a very large effect size. At (5, 5, 25) and (5, 25, 25), BF is predominant at all effect sizes and KW in turn dominates A. It should be noted that W is completely ineffective in case of a very large effect size, since all observed values in one of the three groups become equal to the end point of the scale, corresponding with a zero variance. W is also generally inefficient when at least one sample size is small, as the probability of a zero variance in at least one group increases. Hence, when data are heavily skewed but equidistant, A should be the recommended method at equal sample sizes, while BF should be preferred when sample sizes are unequal. In other words, the recommendation for equidistant data under heavy skewness is identical to the one under moderate skewness.

Table 11 shows the results from the case where data are heavily skewed and the EOS effect is present. At (25, 25, 25), there are no significant differences in performance between the methods; however, at (5, 5, 5), A dominates the other methods at all effect sizes while KW generally performs better than BF and W. At (5, 5, 25) and (5, 25, 25), BF is again generally predominant at all effect sizes; however, A dominates KW. Thus, when data are heavily skewed and the EOS effect is present, A should be the recommended method at equal sample sizes, while BF should be preferred when sample sizes are unequal. Note that these recommendations deviate from those under moderate skewness.

Table 10: Proportion of significant tests, heavily skewed data and equidistance

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0481	0.0480	0.0473	0.0514	0.016
$n_2 = 25$	0.10	0.1082	0.1041	0.1065	0.1120	0.001
$n_3 = 25$	0.25	0.4569	0.4391	0.4529	0.4646	< 0.001
	0.40	0.8879	0.8757	0.8853	0.8906	0.066
	0.65	1.0000	1.0000	1.0000	0.0000	< 0.001
$n_1 = 5$	0.00	0.0435	0.0421	0.0365	0.0077	< 0.001
$n_2 = 5$	0.10	0.0529	0.0510	0.0438	0.0095	< 0.001
$n_3 = 5$	0.25	0.0930	0.0900	0.0765	0.0165	< 0.001
	0.40	0.1839	0.1839	0.1511	0.0269	< 0.001
	0.65	0.4487	0.4813	0.3633	0.0000	< 0.001

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 5$	0.00	0.0468	0.0417	0.0479	0.0086	< 0.001
$n_2 = 5$	0.10	0.0365	0.0414	0.0852	0.0082	< 0.001
$n_3 = 25$	0.25	0.0814	0.1092	0.2126	0.0278	< 0.001
	0.40	0.2219	0.2895	0.4347	0.0838	< 0.001
	0.65	0.6906	0.8133	0.8268	0.0000	< 0.001
$n_1 = 5$	0.00	0.0471	0.0441	0.0582	0.0163	< 0.001
$n_2 = 25$	0.10	0.0526	0.0557	0.1017	0.0190	< 0.001
$n_3 = 25$	0.25	0.1518	0.1605	0.2682	0.0632	< 0.001
	0.40	0.3735	0.3941	0.5697	0.1501	< 0.001
	0.65	0.8301	0.8769	0.9824	0.0000	< 0.001

Table 11: Proportion of significant tests, heavily skewed data and end-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0500	0.0480	0.0496	0.0520	0.046
$n_2 = 25$	0.10	0.1074	0.1041	0.1066	0.1083	0.196
$n_3 = 25$	0.25	0.4451	0.4391	0.4437	0.4480	0.206
	0.40	0.8769	0.8757	0.8760	0.8758	0.997
	0.65	1.0000	1.0000	1.0000	0.0000	< 0.001
$n_1 = 5$	0.00	0.0532	0.0421	0.0385	0.0102	< 0.001
$n_2 = 5$	0.10	0.0640	0.0510	0.0469	0.0126	< 0.001
$n_3 = 5$	0.25	0.1114	0.0900	0.0844	0.0215	< 0.001
	0.40	0.2198	0.1839	0.1758	0.0330	< 0.001
	0.65	0.5563	0.4813	0.4602	0.0000	< 0.001
$n_1 = 5$	0.00	0.0468	0.0417	0.0445	0.0144	< 0.001
$n_2 = 5$	0.10	0.0510	0.0414	0.0654	0.0107	< 0.001
$n_3 = 25$	0.25	0.1321	0.1092	0.1575	0.0124	< 0.001
	0.40	0.3346	0.2895	0.3569	0.0275	< 0.001
	0.65	0.8603	0.8133	0.8761	0.0000	< 0.001
$n_1 = 5$	0.00	0.0470	0.0441	0.0678	0.0186	< 0.001
$n_2 = 25$	0.10	0.0607	0.0557	0.1079	0.0179	< 0.001
$n_3 = 25$	0.25	0.1700	0.1605	0.2597	0.0477	< 0.001
	0.40	0.4087	0.3941	0.5361	0.1060	< 0.001
	0.65	0.8946	0.8769	0.9946	0.0000	< 0.001

Finally, table 12 shows the results from the case where data are heavily skewed and the MOS effect is present. At (25, 25, 25), KW dominates all other methods at all effect sizes, except when the effect size is small, in which case it does not differ significantly from W. At (5, 5, 5), KW is also predominant at all effect sizes. At (5, 5, 25) and (5, 25, 25), however, BF has higher power than the other methods at all effect sizes except the very large one. Hence, under heavy skewness and MOS effect, A should be the recommended method at equal sample sizes, while BF should be preferred when sample sizes are unequal. Again, these recommendations are identical to the recommendations under moderate skewness.

Table 12: Proportion of significant tests, heavily skewed data and middle-of-scale effect

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 25$	0.00	0.0434	0.0480	0.0400	0.0553	< 0.001
$n_2 = 25$	0.10	0.0898	0.1041	0.0848	0.1080	< 0.001
$n_3 = 25$	0.25	0.3764	0.4391	0.3658	0.4126	< 0.001
	0.40	0.8059	0.8757	0.7976	0.8369	< 0.001
	0.65	0.9999	1.0000	0.9998	0.0000	< 0.001
$n_1 = 5$	0.00	0.0292	0.0421	0.0143	0.0007	< 0.001
$n_2 = 5$	0.10	0.0355	0.0510	0.0165	0.0008	< 0.001
$n_3 = 5$	0.25	0.0602	0.0900	0.0282	0.0026	< 0.001
	0.40	0.1151	0.1839	0.0500	0.0049	< 0.001
	0.65	0.2746	0.4813	0.1167	0.0000	< 0.001

Sample sizes	Effect size	Test procedure				p-value
		ANOVA	Kruskal-Wallis	Brown-Forsythe	Welch	
$n_1 = 5$	0.00	0.0576	0.0417	0.0668	0.0041	< 0.001
$n_2 = 5$	0.10	0.0279	0.0414	0.1453	0.0162	< 0.001
$n_3 = 25$	0.25	0.0173	0.1092	0.3330	0.0762	< 0.001
	0.40	0.0405	0.2895	0.5322	0.1754	< 0.001
	0.65	0.2082	0.8133	0.7263	0.0000	< 0.001
$n_1 = 5$	0.00	0.0479	0.0441	0.0462	0.0183	< 0.001
$n_2 = 25$	0.10	0.0415	0.0557	0.0822	0.0417	< 0.001
$n_3 = 25$	0.25	0.1037	0.1605	0.2208	0.1343	< 0.001
	0.40	0.2536	0.3941	0.4668	0.2561	< 0.001
	0.65	0.6133	0.8769	0.8555	0.0000	< 0.001

4. Conclusion

The methodological issue of whether a Likert-type scale can be reasonably assumed to have metric properties and the type of statistical method that should consequently be used to analyse Likert-type data goes back a long time, and has been discussed in relation to a number of paradigms. Still, parametric methods are often used in contemporary research to analyse data that are not equidistant by nature (Jakobsson, 2004). In this study, we saw that respondents generally did not perceive a Likert-type scale as equidistant, and that the nature of the perceived non-equidistance depended on how verbal anchors were connected to the scale points. We also tested the sensitivity of common statistical methods to the two main types of non-equidistance under different circumstances. The overall conclusion from these simulations was that the best statistical method to compare different groups of Likert-type data seems to depend both on the expected scale effect, that is, the nature of the non-equidistance created by the use of verbal anchors and on the degree of skewness (see figure 4). Hence, this study contributes to the methodological literature in two different ways.

<table>
<tr><td rowspan="2">Skewness</td><td colspan="3">Expected scale effect</td></tr>
<tr><td>Equidistance</td><td>End-of-scale</td><td>Middle-of-scale</td></tr>
<tr><td>Approximate symmetry</td><td>ANOVA</td><td>Kruskal-Wallis</td><td>ANOVA</td></tr>
<tr><td>Moderate skewness</td><td rowspan="2">ANOVA (if sample sizes are approximately equal) or Brown-Forsythe (otherwise)</td><td>Kruskal-Wallis (if all sample sizes are large) or ANOVA (otherwise)</td><td rowspan="2">Kruskal-Wallis (if sample sizes are approximately equal) or Brown-Forsythe (otherwise)</td></tr>
<tr><td>Heavy skewness</td><td>ANOVA (if sample sizes are approximately equal) or Brown-Forsythe (otherwise)</td></tr>
</table>

Figure 4: The preferred method with respect to expected scale effect and skewness

Further research regarding subject perceptions of the Likert-type scale is required. The standard verbal anchors for the second and fourth scale points are 'Disagree' and 'Agree', respectively. Changing them to, for example, 'Somewhat disagree' and 'Somewhat agree' might change the way subjects perceive the scale in terms of distance between the points. Scales with more, or less, than five scale points should also be evaluated in a similar manner.

Further research is also required to evaluate the robustness of parametric methods to violations of the equidistance assumption. Other types of non-equidistant data can be used to evaluate the robustness of the parametric methods in more detail, as well as other types of statistical tests commonly used to analyse Likert-type data.

References

Albaum, G., Best, R. & Hawkins, D. (1977). The measurement properties of semantic scale data. *Journal of the Market Research Society*, 19 (1), 21-26.

Albaum, G. (1997). The Likert scale revisited: An alternate version. *Journal of the Market Research Society*, 39 (2), 331-348.

Alexandrov, A. (2010). "Characteristics of Single-Item Measures in Likert Scale Format". *The Electronic Journal of Business Research Methods*, 8 (1), 1-12.

Bendixen, M. T. & Sandler, M. (1995). Converting verbal scales to interval scales using correspondence analysis. *Management Dynamics: Contemporary Research*, 4 (1), 31-49.

Brown, M. B. & Forsythe, A. B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics*, 30, 719-724.

Carifio, J. & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3, 106-116.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

Cunningham, W. H., Anderson Jr, W. T. & Murphy, J. H. (1974). Are students real people? *The Journal of Business*, 47 (3), 399-409.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.

Feir, B. J. & Toothaker, L. E. (1974). The ANOVA F-Test Versus The Kruskal-Wallis Test: A Robustness Study. *The Annual Meeting of the American Educational Research Association*, Chicago, Il.

Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42 (3), 237–288.

Granberg-Rademacker, J. S. (2010). An Algorithm for Converting Ordinal Scale Measurement Data to Interval/Ratio Scale. *Educational and Psychological Measurement*, 70 (1), 74-90.

Harwell, M. R. & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71, 105-131.

Jakobsson, U. (2004). Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences*, 18, 437-440.

James, L. R., Demaree, R. G. & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38, 1212-1218.

Kennedy, R., Riquier, C. & Sharp, B. (1996). Practical Applications of Correspondence Analysis to Categorical Data in Market Research. *Journal of Targeting, Measurement and Analysis for Marketing*, 5 (1), 56-70.

King, G., Murray, C. J. L., Salomon, J. A. & Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*, 98 (1), 191-207.

Krosnick, J. A. & Fabrigar, L. R. (1997). Designing Rating Scales for Effective Measurement in Surveys. In: Lyberg, L.E., Biemer, P., Collins, M., de Leeuw, E.D., Dippo, C., Schwarz, N., and Trewin, D. (Eds.). *Survey Measurement and Process Quality*. NY: Wiley-Interscience.

Kruskal, W. H. & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis, *Journal of the American Statistical Association*, 47 (260), 583–621.

Lantz, B. (2012). The impact of sample non-normality on ANOVA and alternative methods, forthcoming in *British Journal of Mathematical and Statistical Psychology*. DOI: 10.1111/j.2044-8317.2012.02047.x

Lee, J. A. & Soutar, G. N. (2010), Is Schwartz's value survey an interval scale, and does it really matter?" *Journal of Cross-Cultural Psychology*, 41, 76-86.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology* 140, 1-55.

Michell, J. (1986). Measurement Scales and Statistics: A Clash of Paradigms. *Psychological Bulletin*, 100 (3), 398-407.

Mundy J. & Dickinson D. (2004). Factors affecting the uptake of voluntary HIV/AIDS counselling and testing (VCT) services in the workplace. *In: HIV/AIDS in the Workplace Research Symposium*. University of Witwatersrand, June 2004, 175-193.

Pearse, N. (2011). Deciding on the Scale Granularity of Response Categories of Likert type

Scales: The Case of a 21-Point Scale. *The Electronic Journal of Business Research Methods*, 9 (2), 159-171.

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effect. *Journal of Consumer Research*, 16, 158-174.

Stevens, S. S. (1946). On the Theory of Scales of Measurement, *Science,* 7 (103), 677-680.

Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236-247.

Welch, B. L. (1951). On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika*, 38, 330–336.

Wu, C-H. (2007). An Empirical Study on the Transformation of Likert-scale Data to Numerical Scores. *Applied Mathematical Sciences*, 1, 58, 2851-2862.

Zimmerman, D. W. (1998). Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions, *Journal of Experimental Education*, 67:1, 55–68.

The use of Grounded Theory Technique as a Practical Tool for Qualitative Data Collection and Analysis

Japhet Lawrence[1] and Usman Tar[2]
[1]Department of Applied Computing and Information Technology
University of Kurdistan-Hawler
Erbil, Kurdistan Region of Iraq
[2]Department of Politics & International Relations
University of Kurdistan-Hawler
Erbil, Kurdistan Region of Iraq
lawrenceje@yahoo.com
usmanatar@gmail.com

Abstract: When encountering qualitative research for the first time, one is confronted with both the number of methods and the difficulty of collecting, analysing and presenting large amounts of data. In quantitative research, it is possible to make a clear distinction between gathering and analysing data. However, this distinction is not clear-cut in qualitative research. The objective of this paper is to provide insight for the novice researcher and the experienced researcher coming to grounded theory for the first time. For those who already have experience in the use of the method the paper provides further much needed discussion arising out of the method's adoption in the IS field. In this paper the authors present a practical application and illustrate how grounded theory method was applied to an interpretive case study research. The paper discusses grounded theory method and provides guidance for the use of the method in interpretive studies.

Keywords: grounded theory; interpretive; case study; data collection; data analysis; qualitative; quantitative

1. Introduction

One of the main problems of conducting interpretive qualitative research is to decide an appropriate starting point for the research, and the basic framework within which the data will be collected and analysed. Qualitative studies tend to produce large amounts of data that are not readily amenable to mechanical manipulation, analysis and data reduction (Yin, 1984). It not only generates large amount of data, but it generates data in a non standard format which makes analysis problematic (Turner, 1983). Qualitative analysis provides an opportunity for the researcher to gain information and gather insights that may be overlooked with traditional data analysis techniques. The analysis of the case study is done in pursuant to guidance provided by many scholars in this field, (such as Glaser, 1978; Glaser and Strauss, 1967; Lofland and Lofland, 1984; and Taylor and Bogdan, 1984).

The process of data analysis in qualitative research involves working with data, organising it, breaking it down, synthesising it, searching for patterns, discovering what is important and what is to be learned, and deciding what you will tell others (Bogdan and Biklen, 1982). Spradley (1979) refers to analysis as a systematic examination of something to determine its parts, the relationship among parts, and their relationship to the whole. While Miles and Huberman (1984) describe data analysis as consisting of three concurrent activities - data reduction refers to the process of selecting, simplifying, abstracting and transforming the new case data. They argue that data collection and data analysis should overlap to allow for flexibility in data collection procedures so that the researcher remains open to new ideas or patterns which may emerge.

There are many ways of analysing qualitative data and a number of approaches were considered such as hermeneutics, content analysis and semiotics (Myers, 1997). These approaches come from diverse fields and all offer the possibility of different insights on the data. These approaches were evaluated from the perspective of whether the approach draws on all features of case study and whether the philosophy of the approach imposes any pre-existing theories. Eisenhardt (1989) suggests that theory building research must begin as close as possible to the ideal of no theory under consideration and no hypotheses to test since preordained theoretical perspectives may bias and limit the findings.

Reference this paper as: Lawrence, J and Tar, U. "The use of Grounded Theory Technique as a Practical Tool for Qualitative Data Collection and Analysis" *The Electronic Journal of Business Research Methods* Volume 11 Issue 1 2013 (pp 29-40), available online at www.ejbrm.com

The objective of the case study research was the development of a conceptual model that explains the adoption of Internet in small to medium-sized enterprises (SMEs) and the objective of the study fits well with the philosophical nature of grounded theory. Accordingly, proponents of the grounded theory advocate that an approach which concerns itself with the meanings, definitions, and interpretations which are made by the subjects of the study has greater potential for depicting their world and priorities more accurately than methods which begin by preconceiving the world and its meaning (De Burca and McLoughlin, 1996). The researcher did not come to the field with a well-defined set of constructs and instruments with which to measure the social reality; rather the researcher derives the categories from the field by in-depth examination and exposure to the phenomenon.

The selection of grounded theory (Glaser and Strauss, 1967; Glaser, 1978; Strauss, 1987; Strauss and Corbin, 1990, 1998) amongst a myriad of other qualitative methods is not arbitrary but rather because it has been a dominant paradigm for social research (Hughes and Jones, 2003) and its use is increasing in the IS field. This is evidenced by the growing literature that is either discursive on philosophy and application or detailed about method (Toraskar, 1991; Orlikowski, 1993; Baskerville and Pries-Heje, 1999; Trauth, 2000; Hughes and Howcroft, 2000; Urquhart, 2001). Hughes and Jones (2003) note that more researchers are taking up qualitative studies, it is therefore worth reflecting on lessons learned from the practical application of the method. The purpose of this paper is to provide insight for the novice researcher and the experienced researcher coming to grounded theory for the first time. For those who already have experience in the use of the method the paper provides further much needed discussion arising out of the method's adoption in the IS field.

The remainder of the paper is structured as follows. It begins with an overview of the grounded theory method and a discourse on the use of grounded theory in Information Systems. This will be followed by a description of the procedures involved in collecting and analyzing data in grounded theory method. Following on from that is the justification for using grounded theory to collect and analyse the case study data. An illustrative piece of research is then presented in which grounded theory was used in interpretive, qualitative case studies and finally, the paper presents some conclusions.

2. The grounded theory method

This section discusses grounded theory (Glaser and Strauss, 1967; Martin and Turner, 1986; Turner, 1983; Glaser, 1978; Strauss, 1987; Strauss and Corbin, 1990, 1998; Dey, 1999; Charmaz, 2003, 2006, 2008). Grounded theory originates in the work of Glaser and Strauss (19967) and is a method that has been used extensively across a variety of social science disciplines. A grounded theory is one that is discovered, developed, and provisionally verified through systematic data collection and analysis of data pertaining to a particular phenomenon (Strauss and Corbin, 1990). It is invaluable when conducting empirical research; it has some attraction for a researcher using qualitative techniques for the first time and it offers well sign-posted procedures. In the method conceptual properties and categories may be 'discovered' or generated from the qualitative data by following a number of guidelines and procedures. The grounded theory is iterative, requiring a steady movement between concept and data, as well as comparative, requiring a constant comparison across types of evidence to control the conceptual level and scope of the emerging theory.

The goal of grounded theory is seeking a theory that is intimately tied with the evidence, so that the resultant theory is likely to be consistent with empirical data (Orlikowski 1993; Eisenhardt 1989). Data collection, coding rationale, integration of categories, abstracting from the data and construction of theory are thus guided by theory as it emerges. Hughes and Wood-Harper (1999) report that the main application areas of grounded theory are most notably in Glaser and Strauss' own research into status passage, but also in a number of other, usually medical or nursing related areas such as - experiences with chronic illness (Charmaz, 1980); the management of a hazardous pregnancy (Corbin, 1992) and homecoming (Hall, 1992). Additionally much work has been done with respect to guidance on the use of grounded theory method. Most notable amongst them include Turner (1983); Martin and Turner (1986); Strauss (1987); Strauss and Corbin (1990); Dey (1999); Charmaz (2003, 2006, 2008); Jones and Alony (2011).

The use of grounded theory has also spread to other disciplines including research in information systems (Torasker, 1991; Pidgeon *et al*, 1991; Oliphant and Blockley, 1991; Pries-Heje, 1992; Orlikowski, 1993; Pettigrew, 1990; Calloway and Ariav, 199l; Baskerville and Pries-Heje, 1995, 1998). The most notable use of grounded theory in

IS research is that by Orlikowski (1993) in which she presents findings of a study into the adoption and use of CASE tools. In this study the use of grounded theory was justified on the basis that it provided 'a focus on contextual and processual elements as well as the action of key players associated with organizational change elements that are often omitted in IS studies. The grounded theory fitted well with the interpretivist rather than positivst nature of this research. Grounded theory studies in this interpretivist tradition have become increasingly common in the IS research literature precisely because the method is useful in developing context-based, process-oriented descriptions and explanations of phenomenon (Myers, 1997; Urquhart, 2001).

Hughes and Howcroft (2000) consider that the individual researcher plays a critical role in an interpretive study. They maintain that using the grounded theory procedures may be a way for a researcher to deal with some of the uncertainty that some researchers feel when faced with data collection and analysis in interpretive studies. Notably they point to the fact that for novice researchers (or experienced researchers new to interpretive studies) grounded theory 'provides a useful template…and as such serves as a comfort factor in the stressful and uncertain nature of conducting qualitative research (Hughes and Jones, 2003).

Grounded theory is a general style of doing analysis that does not depend on any particular disciplinary perspectives (Strauss 1987) and, therefore, would seem to lend itself to information systems research, which can be described as a hybrid discipline. The main aspect of grounded theory, which differentiates it from other qualitative research methods, is its emphasis upon theory development (Strauss and Corbin, 1998). Theory is grounded when it emerges from and generates explanations of relationships and events that reflect the life experiences of those people and processes that the researcher is attempting to understand. It also differs from other qualitative approaches, because traditional qualitative approaches collect data first before commencing the analysis and long after they have left the research site. In contrast, grounded theory uses the emerging theoretical categories to shape the data collection while doing the fieldwork (data collection and analysis proceed simultaneously). By analysing data from the lived experience of the research participants, the researcher can, from the beginning attends to how they construct their world.

The use of grounded theory is founded on the premise that the generation of theory at various levels is indispensable for a deep understanding of social phenomena (Glaser and Strauss 1967; Glaser 1978). It requires that the researcher demonstrates theoretical sensitivity (Glaser and Strauss 1967; Glaser 1978) by being well grounded in technical literature as well as from personal and professional experience in collection and analysis of data (Strauss and Corbin 1990). It encourages researchers to steer their thinking out of the confines of technical literature and avoid standard ways of thinking about the data. The interplay between emergent theory and technical literature comes to the fore when extending generalisations from the study, which is achieved by either integrating supplementary or conflicting analysis into the theory by including them as categories or conditions, or criticising them in terms of what has emerged (Strauss 1987). Grounded theory is particularly suitable for a case study aimed at exploring the factors that influence the adoption of Internet in Small to Medium-sized Enterprises. It is useful for understanding contextual elements (Orlikowski 1993) that constituted the main focus of this case study.

One very practical problem with grounded theory is that the method is extremely labour intensive, requiring the investment of considerable cognitive effort by the researcher. However, the author believes that grounded theory technique is a suitable approach to use, especially when a researcher needs to analyse large quantities of unstructured or semi-structured qualitative data. This section has presented and discussed grounded theory as a practical tool for collecting and analysing qualitative data. The description of the procedures involved in collecting and analysing data in grounded theory is the topic of the next section.

3. Grounded theory procedures

The previous section explicated grounded theory method. The defining characteristic of grounded theory is that of a general methodology for discovering theory that is grounded in data systematically gathered and analysed. The theory evolves during actual research, and it does this through a continuous interplay between analysis and data collection; data analysis guides future data collection. In this section consideration will be given to the details of the procedures associated with data collection and analysis in grounded theory method.

Strauss and Corbin (1990) identify three levels of analysis - (a) to present the data without interpretation and abstraction, the participants tell their own story; (b) to create a rich and believable descriptive narrative using field

notes, interview transcripts and researcher interpretations; and (c) building a theory using high levels of interpretation and abstraction. This research combines the second and third approaches, to present rich and detailed descriptions, which allows the reader to make sufficient contextual judgements to transfer the case study findings to alternative settings. The concern here is with the multiple constructions of reality as experienced by Small to Medium-sized Enterprises.

Data analysis in grounded theory involves specific procedures which, when applied appropriately and with vigilance will result in theory that is rigorous and well grounded in the data. Strauss (1987) points out that the procedures should be thought of as rules of thumb, rather than hard or fixed rules, and advises researchers to study these rules of thumb, use them, and modify them in accordance with the requirements of their research. In addition, Strauss and Corbin (1998) warn researchers that rigid adherence to any procedure can hinder the analytic process and stifle creativity.

An example application that demonstrates how grounded theory was applied to an interpretive in-depth case study research is presented in the section titled application of grounded theory. Recording of data may be thought of as a pre-analytic step of grounded theory method and it is said to be essential to the successful generation of grounded theory (Hutchinson, 1988). The grounded theory approach involves coding the assignment of themes and concepts to a selected unit such as sentences taken from an interview transcript. The concepts are combined into related categories; links between categories are identified and verified against the data, and selective coding attempts to integrate the categories into a theory, which accounts for the phenomenon being investigated. The subsection below discusses the process of analysis in grounded theory which is coding data (that comprises open, axial and selective); memo writing and theoretical sampling.

4. Coding

Codes can take the form of a straight forward category label or a more complex one (example a metaphor) (Miles and Huberman, 1994). The coding of data such as field notes and interview transcripts poses questions such as what does this incident indicate? Coding gets the researcher off the empirical level by fracturing the data, and then conceptually grouping it into codes that then become the theory which explains what is happening in the data (Glaser, 1978). Researchers use codes to pull together and categorise a series of otherwise discrete events, statements, and observations which they identify in the data (Charmaz, 1983).

Open coding is the analytic process through which concepts are identified and their properties and dimensions are discovered in data. It is the part of analysis that pertains specifically to the naming and categorising of phenomena through close examination of the data. … During open coding the data are broken down into discrete parts, closely examined, and compared for similarities and differences, and questions are asked about the phenomena as reflected in the data (Corbin and Strauss, 1990). The researcher compares incident to incident with the purpose of establishing the underlying uniformity and its varying conditions (Glaser, 1978). Events, happenings, objects and actions/ interactions that are found to be conceptually similar in nature or related in meaning are grouped under more abstract concepts termed "categories" (Strauss and Corbin, 1998).

Glaser (1978) describes a set of three questions that should guide the open coding. What is this data a study of? This question continually reminds the researcher that his original intents on what he thought were going to study just might not be. What category does this incident indicate? The continual asking of this question keeps the analyst from getting lost in the re-experiencing of his data by forcing him to try and generate codes that relate to other codes. It forces code that earns its way into the theory by its grounding in the data. What is actually happening in the data? What is the basic social psychological problem(s) faced by the participants in the action scene? These three types of questions keep the researcher theoretically sensitive and transcending when analysing, collecting and coding the data. They force the researcher to focus on patterns among incidents, which yield codes, and to rise conceptually above fascinating experiences. It is important to emphasise that researchers' make codes fit the data, rather than force the data into codes.

Axial coding involves re-building the data (fractured through open coding) in new ways by establishing relationships between categories and their subcategories. It is termed "axial" because coding occurs around the axis of a category, linking categories at the level of properties and dimensions (Strauss and Corbin, 1998). Axial codes typically represent categories that describe the open codes. The researcher continues to code and compares the concept to more incidents (Glaser, 1978). Comparison enables the identification of variations in

the patterns to be found in the data. Data coding at this level is intended to elevate the data to higher levels of abstraction (Hutchinson, 1988). During axial coding, the analyst begins to fit the pieces of the data 'puzzle' together, which were fractured during open coding. Each piece (e.g., category and subcategory) has its place in the overall explanatory scheme. When building a puzzle, the analyst might pick up a piece and ask, "Does this go here or there?" The analyst's first attempts are often trial and error. Later, as he becomes more theoretically sensitive, making the fit between conceptual indicator and category becomes easier.

Selective coding: the aim of selective coding is to integrate and refine the categories into a theory, which accounts for the phenomenon being investigated (Darke *et al*, 1998) and validates the statements of relationships among concepts, and fills in any categories in need of further refinement. In selective coding the researcher reduces data from many cases into concepts and sets of relational statements that can be used to explain, in a general sense, what is going on (Strauss and Corbin, 1998).

5. Memo writing

Memos are devices that depict the relationship among concepts. It is an important ways of keeping records of analysis. Martin and Turner (1986) and Strauss (1987) discuss the processes involved in detail. Memo writing takes place throughout the research process starting with the first interview. They serve a dual purpose of keeping the research grounded and maintaining awareness for the researcher. Memos provide an opportunity to generate and develop explanations of the emerging concepts, and to discern some of the interrelationships which exist between them. The memo informs what the code is about and provides the pivotal step of breaking the categories into components and elaborating the codes. Glaser (1978) considers writing of theoretical memos as the core stage in the process of generating theory. Glaser defines memo as "the theorising write-up of ideas about codes and their relationships as they strike the analyst while coding ... memo can be a sentence, a paragraphs or a few pages ... it exhausts the analyst momentary ideation based on data with perhaps a little conceptual elaboration". Memos don't just report data; they tie together different pieces of data into a recognisable cluster, often to show that those data are instances of a general concept.

Memos are one of the most useful and powerful sense-making tools at hand for researchers to use during analysis. The advice is to 'stop and memo' as coding sparks off ideas. You are writing memos to yourself, secondarily to colleagues. Memoing helps the analyst move easily from empirical data to conceptual level, refining and expanding codes further, developing key categories and showing their relationships, and building towards a more integrated understanding of events, processes, and interactions in the case. Memoing develops the core category around which the other categories integrate. The core category integrates the theory according to the emergent perspective of investigation and thereby defines its cut-off points.

However, the core category has earned its relevance through the grounding of the theory in the domain. 'It must be central, that is related to as many other categories and their properties as possible... and account for a large portion of the variation in a pattern of behaviour' (Glaser, 1978). It must also occur frequently, be completely variable, and 'have a clear and grabbing implication for formal theory' (Glaser, 1978). Memos are a rapid way of capturing thoughts that occur throughout data collection, data reductions, data display, conclusion drawing and final reporting. It saturates dimensions of the main categories that have emerged through coding, and constantly generates open questions for further coding and data collection. At the end of the process memos have to be sorted and integrated. Sorting memos simply means putting those that elucidate the same category together in order to clarify its dimensions and to distinguish it from the other categories.

6. Theoretical sampling and comparing

Two analytic processes contribute to raising categories to conceptual categories - *constant comparison*, which is central in generating grounded theory and *theoretical sampling* (Glaser and Strauss, 1967). Both of these processes are achieved through a process Glaser (1978) calls theoretical sampling and the selective sampling of the literature. Essentially, the researcher needs to confront the conceptual categories with more data in order to define them carefully, delineate their properties, explicate their causes, and demonstrate the conditions under which they operate, and spell out their consequences.

The constant comparative is central to the data analysis in generating grounded theory. The purpose is to build and clarify a category by examining all the data it covers and variations from it. The researcher takes a

limited set of codes that were developed in the initial phase and applies them to large amounts of data. The coded data are compared with other data and assigned to clusters or categories according to obvious fit. Glaser (1978) labels this process of comparison as the constant comparative method, where bits of data are compared with other data and where coded data is constantly confronted with new data for verification purposes. "Comparative analysis forces the researcher to 'tease out' the emerging category by searching for its structure, temporality, cause, context, dimensions, consequences and its relationship to other categories" (Hutchinson, 1988). Additionally, it is appropriate and desirable to compare the data categories and constructs that emerge between various groups of participants in the study. In this way the process of constant comparison is intended to generate a theory rich in detail. It moves the researcher more quickly away from describing the specifics of a case to thinking more abstractly about what the various cases share in common and what is different about them.

Theoretical sampling begins during the data collection phase of the study and involves searching the transcripts for emerging categories that characterise the narrative and seem significant. "*Theoretical sampling* is the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyses... data and decides what data to collect next and where to find them, in order to develop... theory as it emerges. This process of data collection is controlled by the emerging theory, whether substantive or formal (Glaser, 1978). Theoretical sampling primary function is to provide the researcher with the opportunity to discover properties of the core variable under study by collecting new data to check, fill out and extend conceptual categories. Theoretical sampling and constant comparison reflect cyclical processes which are fluid and flexible, but at the same time they ensure that the analysis is planned and well grounded in the data, rather than haphazard, which can lead the analyst down to unproductive paths and away from the focus of study.

Maximising opportunities for comparing concepts along their properties for similarities and differences enables researchers to densify categories, to differentiate among them, and to specify their range of variability. Once the analyst has some categories, sampling is aimed at developing, densifying and saturating those categories. It is also related to the sensitivity that a researcher has developed to the emerging concepts. The more sensitive a researcher is to the theoretical relevance of certain concepts, the more likely he is to recognise indicators of those concepts in the data. Sensitivity usually grows throughout the research project and enables the researcher to decide what concepts to look for and where he might find indicators of them.

As constructs are derived from the data, repeated theoretical sampling can be used to increase the depth of focus and to ensure consistency; that is, to ensure that data are gathered in a systematic way for each category (Strauss and Corbin, 1990). "Theoretical sampling is used as a way of checking on the emerging conceptual framework rather than being used for the verification of preconceived hypotheses" (Glaser, 1978). After developing a set of focused codes into categories the researcher has to weave them together in developing a grounded theory. The emerged-grounded categories, derived from the data, are the basic building blocks for the theoretical understanding of the area under study. The conceptual framework developed from the conceptual categories is tested by collecting data which provides support (or not) for the framework hypotheses and reveals the relationship between the categories, which forms the basis for the subsequent emergent theory.

An important issue in reaching closure is when to stop sampling. Ideally, researchers should stop sampling when theoretical saturation is reached (Glaser, 1992; Eisenhardt, 1989). Theoretical saturation is the point at which incremental learning is minimal because the researcher is observing phenomena seen before (Glaser and Strauss, 1967). The general rule when building theory is to gather data until each category is saturated (Glaser, 1978; Glaser and Strauss, 1967). This means until (a) no new or relevant data seem to emerge regarding a category, (b) the category is well developed in terms of its properties and dimensions demonstrating variations and (c) the relationships among categories are well established and validated. Unless a researcher gathers data until all categories are saturated, the theory will not be evenly developed and it will lack density and precision. The next section presents justification for using grounded theory to collect and analyse the case study data.

7. Justification for using grounded theory method for interpretive case study research

Grounded theory is chosen for collecting and analysing the case study data, with the aim of generating a descriptive and explanatory theory of the adoption of Internet rooted in the experiences of Small to Medium-sized Enterprises. Strauss (1987) emphasises the usefulness of the case study approach when used with grounded theory. It is an inductive, theory discovery method that allows the researcher to develop a theoretical account

of the general features of a topic while simultaneously grounding the account in empirical data (Martin and Turner, 1986; Glaser and Strauss, 1967). This generative approach seems particularly useful here given the objective of the study was the discovery of theory that explains the adoption and use of Internet in Small to medium-sized enterprises.

Grounded theory offers a way of attending in details to qualitative material in order to develop systematically theories about the phenomena being studied. Turner (1981) suggests that grounded theory is particularly well suited to dealing with qualitative data of the kind gathered from participant observation, from the observation of face-to-face interaction, from semi-structured or unstructured interviews, from case-study material or documentary sources. Typically, these particular kinds of inquiry generate large amounts of data, which accumulates in non-standard and unpredictable formats. The grounded theory approach offers the researcher a strategy for sifting and analysing material of this kind. A particular strength of utilising grounded theory is that a documented record of the progress of the analysis is generated. Hence, it is always possible to trace the derivation of any concept or model by checking back through the data and memos.

Grounded theory makes its greatest contribution in areas in which little research has been done. Little or no research has been conducted specifically into adoption of innovations in SMEs context. Most of the research in this field has tended to focus on descriptive Internet usage and on demographics, or some other form of Internet pattern (Abell and Lim, 1996; Abell and Black, 1997). The paucity of research about Internet usage and adoption in SMEs means that many of the variables relevant to the concepts of this phenomenon are yet to be identified.

Grounded theory is an appropriate method for this study as it generates theory that can be used as a precursor for further investigation of this phenomenon and related issues. Other qualitative research techniques, quantitative methods, or a combination of both can then be used in subsequent studies to test, verify or extend the qualitative propositions that emerge from this research.

A major premise of grounded theory is that to produce accurate and useful results, the complexities of the organisational context have to be incorporated into an understanding of the phenomenon, rather than be simplified or ignored (Orlikowski, 1993; Martin and Turner, 1986; Pettigrew, 1990). This mode of research requires that broader, contextual issues, which are shown to influence the phenomenon under study, be given appropriate recognition in the development of theory. Rather than focusing the investigation by disregarding these broader conditions, every effort was made to acknowledge and account for them.

A number of theoretical approaches emphasise the criticality of organisational context in shaping technology use in organisations, such a conviction also informs this research, and the use of a grounded theory method allows the inclusion and investigation of this key organisational element. This inductive approach relies on the researcher systematically collecting, coding, categorising and analysing data to derive the theory that explains the phenomena. The method facilitates "the generation of theories of process, sequence, and change pertaining to organisations, positions, and social interaction" (Glaser and Strauss, 1967).

Most Information Technology usage models tend to neglect the contextual aspect of technology use in literature. An approach that specifically includes elements of process and context is thus particularly appropriate here. The three characteristics of grounded theory inductive, contextual, and processual fit with the interpretive orientation of this research. The focus here is on developing a context-based description and explanation of the phenomenon, rather than an objective, static description expressed strictly in terms of causality (Boland, 1979, 1985; Chua, 1986; Orlikowski and Baroudi, 1991). The research develops theory which describes and explains the adoption and usage of Internet in terms of an interaction of contextual conditions, actions, and consequences, rather than explaining variance using independent and dependent variables (Orlikowski, 1993). This section has discussed the reasons for using grounded theory as analysis tool for the case study. The next section presents by way of illustration a piece of practical research project in the interpretive tradition in which grounded theory was used.

8. A practical application of grounded theory in the context of IS research

The discussion in this section is based on Lawrence (2002). The research concerned a case study that focused on in-depth understanding of the factors that influence SMEs decision to adopt and use the Internet in

business. The aim of the study was to describe, interpret, analyse, and understand Internet adoption and usage from SMEs' perspective.

The number of SMEs that were approached to participate in the study was 26. The criteria for inclusion were based on a need for each participating SME to conform to the definition of SMEs and a willingness on the part of the SME owners/mangers to disclose details of their business. Several potential SMEs were rejected on the grounds that they did not satisfy the criteria. Seven SMEs that satisfied the criteria were chosen to participate in this study, amounting to total of seven separate field studies which were carried out within the broad tradition of interpretive case study (Zuboff, 1988; Orlikowski, 1991; Walsham, 1993). The case study involved extensive interviewing of key participants (e.g. company owner or manager in each of the SMEs), coupled with the use of documentary evidence such as company reports.

The primary details of the participants are shown in table 1 in no significant order. These SMEs were chosen across business sectors so that the study could investigate the existence of sector-independent issues. This was important to avoid observations specific to a particular sector. The first SME was selected at random from the seven SMEs that participated in this study, to provide the first body of data. Then subsequent data collection was guided by the theoretical sampling principle of grounded theory as defined by Strauss and Corbin (1990); that is sampling on the basis of concepts that have proven theoretical relevance to the evolving theory. In a grounded theory theoretical sampling cannot be fully planned before the study commences.

Table 1: SMEs that participated in the case study

SMEs	Type of Business	Size (employees)
BIL	Peugeot cars franchise holder	20
BPC	Publishing	25
SAH	Health care	200
MGL	Manufacturer of contract carpets	9
AL	Manufacturer and seller of educational engineering equipment	40
FP	Specialist flooring manufacturer	110
CLR	Cigarette paper manufacturer	180

Source: Lawrence, 2002

The detailed write-up of the cases and all the data generated by interviews, and documentation were examined and coded by focusing on the factors that influence adoption. This technique uses open coding (Strauss and Corbin, 1990) where the data were read and categorised into concepts that are suggested by the evidence rather than imposed from outside (Orlikowski, 1993).

Some of the initial codes that emerged from the open coding process were *promotional and advertising, communication medium to improve organisational efficiency, easy entry into new markets, and the Internet generates new business opportunities,* all these codes contributed to the category 'benefits of using the Internet'. Once all the evidence were examined, the codes were organised by recurring theme, for example benefits of using the Internet, cost effectiveness, compatibility and perceived usefulness. These themes became prime candidates for a set of stable and common categories, which linked a number of associated codes. This is known as axial coding (Strauss and Corbin, 1990) and it relies on a synthetic technique of making connections between subcategories to core categories to construct a more comprehensive scheme. The categories of benefits of using the Internet, cost effectiveness, compatibility and perceived usefulness became subcategories of the technological (Internet characteristics). The first SME case data were then re-examined and re-coded using this proposed scheme, the goal being to determine that set of categories and concepts covered as much of the data as possible. This iterative examination yielded a set of broad categories and associated concepts that described the salient conditions, events, experiences, and consequences associated with the adoption and use of the Internet in the first SME case, see table 2.

These initial concepts guided the remaining field study, allowing the process of data collection, coding, and analysis to be more targeted. Following the constant comparative analysis method (Glaser and Strauss, 1967), the remaining SMEs case's experiences were systematically compared and contrasted with those of the initial SME case. This analysis also used Miles and Huberman's (1984) technique for across-case pattern comparison and clustering that involves matrix displays to compare key events, triggers, and outcomes.

The iteration between data and concepts ended when enough categories and associated concepts had been defined to explain the factors that influence SMEs decision to adopt and use the Internet in business, and when no additional data were being collected from the SMEs to develop or add to the set of concepts and categories, a situation Glaser and Strauss (1967) refer to as 'theoretical saturation'. The resultant theoretical model is empirically valid as it can account for the unique data of each case in the study, as well as generalise patterns across the cases (Eisenhardt, 1989). The participants in the study provided commentary, correction, and elaboration on drafts of the findings and theoretical model.

Table 2: Sample of the initial concepts that emerged from the analysis of the first SME case

Emergent Core category	Subcategory	Codes
Technological	Benefits of using the Internet	*Global markets reach* *Communication medium to improve organisational efficiency* *Promotional and advertising* *Better customer service* *Easy access to global information* *The Internet generates new business opportunities*

Source: Lawrence, 2002

The case study findings provide new insights into the factors that influence SMEs decision to adopt and use the Internet in business. The result shows that many of the factors that influenced adoption of the Internet in SMEs were similar to the factors identified in large organisations (Igbaria *et al*, 1997; Davis, 1989; Kwon and Zmud, 1987). However, technological, organisational and environmental factors emerged from the case analysis and were supported by the literature as significant in influencing SMEs decision to adopt and use the Internet, while barriers hindered adoption of the Internet. This result highlights the fact that SMEs have special needs due to their unique organisational characteristics. The categories constituting these factors were combined to formulate a framework to serve as a summary for SMEs Internet adoption, see table 3 for across case comparison. The presence of these categories is indicated by a 'Yes' in the table. The results indicated a positive relationship between technological, organisational factors and the decision to adopt the Internet and less for environmental and barriers to Internet adoption. In all the cases, technological and organisational factors (management support and organisational resources) were congruent with the adoption decision, except for one SME that did not indicate a yes for organisational resources.

Table 3: Across case pattern comparison

Core categories	**Subcategories**	**Small to Medium-sized Enterprises**						
		AL	*SAH*	*BIL*	*BPC*	*FP*	*MGL*	*CLR*
Technological	Compatibility	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Cost effectiveness	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Benefits of using the Internet	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Perceived usefulness of the Internet	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Perceived richness of the Internet	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Perceived ease of use of the Internet	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Organisational	Management support	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
	Organisational resources	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*		*Yes*
	Organisational size						*Yes*	
Environmental	Competitive pressure	*Yes*			*Yes*	*Yes*	*Yes*	*Yes*
	External pressure			*Yes*	*Yes*		*Yes*	
Inhibiting	Security	*Yes*	*Yes*				*Yes*	
	Lack of Internet knowledge		*Yes*	*Yes*			*Yes*	
	Cost of investment		*Yes*				*Yes*	
	Complexity			*Yes*				
	Limitation of infrastructure					*Yes*		*Yes*
	Uncertainty about the Internet			*Yes*	*Yes*			
	Limitation of personal contact			*Yes*				*Yes*
	Lack of universal electronic payment systems							*Yes*

Source: Lawrence, 2002

The results confirmed that organisational factors (management support and organisational resources) played a critical role in the adoption decision of the Internet. Adoption and use of the Internet tended to take place in the firms where the management was enthusiastic about Internet technology and had the necessary organisational resources. The result shows that SMEs were generally very enthusiastic in their involvement with the adoption and use of the Internet in business. They viewed the advent of the Internet as a useful business tool that provided global connectivity; access to low-cost information and its use enabled them to be involved in electronic commerce. The low financial commitment required implied a high level of perceived trialability among SMEs; an important consideration since smaller firms tended to have more limited financial resources.

Table 3 shows a less positive relationship between environmental factor, barriers to Internet adoption and the decision to adopt the Internet, in all the cases, only four cases indicated yes for competitive pressure and three indicated yes for external pressure, especially from trading partners. The barriers to adoption showed various categories that inhibited the adoption of the Internet in SMEs. The strongest barriers to Internet adoption were security and lack of Internet knowledge with three yes each, and two yes for cost of investment, limitation of infrastructure, uncertainty about the Internet and limitation of personal contact. While complexity and lack of universal electronic payment systems have only one yes each.

Finally, the research has highlighted the issues of Internet uptake among the examined SMEs. This study has built on existing research on innovation adoption by showing the factors that influence their decision to adopt the Internet. The results showed that technological, organisational and environmental factors were very important factors that influenced SMEs decision to adopt and use the Internet while barriers to Internet adoption hindered adoption. The findings suggest that the adoption of Internet technology was influenced to a greater extent by the technological and organisational factors than by environmental factors.

9. Conclusion

In organisational research, methods grounded theory can be useful in providing deep insights and understanding of social life that is consistent with interpretive case-based field studies dealing with social and organisational contexts. The interpretive approach and the research methods chosen for this study were described and the justification for choosing grounded theory method was discussed. The grounded theory as a practical tool for qualitative data collection and analysis was presented and a description of the procedures involved in doing data analysis in grounded theory was discussed.

The use of grounded theory helped in providing useful data in original and rich research findings and theory because of its close tie to the data and the rigour in the method. The method explicitly seek to discover the underlying assumptions, the contexts and the experiences of those involved in the IS phenomenon under study. This research study has illustrated that grounded theory, a method more commonly associated with the social science perspective, can indeed assist with rich, context-based interpretive IS research. The empirical work has illustrated that it is possible to successfully use grounded theory in qualitative IS adoption studies where the social aspect is paramount.

The paper reveals that grounded theory is no longer a preserve of particular field of study. It can be employed in both social and natural sciences. Further, methodologies should be seen as means to an end; any discipline could employ specific methodologies in so far as such methods could be transparently applied.

References

Abell, W., and Lim, L., (1996), *Business use of Internet in New Zealand: An Exploratory Study*, http://www.scu.edu.au/ausweb96/buiness/abell/paper.htm

Abell, W., and Black, S., (1997), *Business use of Internet in New Zealand: A Follow-up Study*, http://www.lincoln.ac.nz/ccb/staff/abell/webnet.htm

Baskerville, R. L. and Pries-Heje, J. (1999), *Grounded Action Research: A Method for Understanding IT in Practice,* Accounting, Management and Information Technology, vol. 9, no. 1, pp. 1-23.

Baskerville, R., and Pries-Heje, J. (1998), *Information Technology Diffusion: Building Positive Barriers*, European Journal of Information Systems, 7, 17-28.

Baskerville, R., and Pries-Heje, J. (1995), *Grounding the Theory in Action Research,* Proceedings of the 3rd European Conference on Information Systems, G. Goukidis, B. Galliers, T. Jelassi, H. Krcmar, and F. Land, eds., Athens University of Economics and Business, Athens Greece, 837-349.

Bogdan, R.C., & Biklen, S.K., (1982), *Qualitative research for education: An introduction to theory and methods*, Allyn and Bacon, Boston.

Boland, L. A., (1985), *Reflections on Blaug's Methodology of Economics: Suggestions for a Revised Edition*, Eastern Economic Journal, Eastern Economic Association, vol. 11(4), pages 450-454.

Boland, L. A., (1979), *A Critique of Friedman's Critics*, Journal of Economic Literature, American Economic Association, vol. 17(2), pages 503-22.

Calloway, L. J. and Ariav, G. (1991), *Developing and Using a Qualitative Methodology to Study Relationships among Designers and Tools*, in Nissen, H.E., Klein, H.K. and Hirschheim, R. (eds.), Information Systems Research: Contemporary Approaches and Emergent Traditions, North Holland Amsterdam, pp. 175-193.

Charmaz, K. (2008), *Grounded theory,* in J. A. Smith (Ed.), Qualitative psychology: A practical guide to research methods. Los Angeles sage Publications, pages 81-110.

Charmaz K (2006), *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*, London Sage Publications

Charmaz, K. (2003), *Grounded theory - Objectivist and constructivist methods*, in N. K. Denzin & Y. S.Lincoln (Eds.), Strategies of qualitative inquiry, London Sage Publications, pages 249-291.

Charmaz, K., (1983), *The grounded Theory Method: An Explication and Interpretation,* in Emerson, R.M., (ed.) Contemporary Field Research, Waveland Press Inc., Illinois

Charmaz, K., (1980), *The construction of self pity in the chronically ill,* Studies in Symbolic Interaction, volume 3, pp. 123-145

Chua, Wai Fong, (1986), *Radical Developments in Accounting Thought*, The Accounting Review, 61 (1986), 601-632

Corbin, C.B., (1992), *The effects of verbal feedback on intrinsic motivation and perceived competence of cricketers*, Australian Journal of Science and Medicine in Sport (AJSMS), 24(4) 94 - 99

Darke, P., Shanks, G., & Broadbent, M., (1998), *Successfully completing case study Research: Combining rigour, relevance and pragmatism,* Info Systems Journal, 8,273-289

Davis, F. D., (1989), *Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,* MIS Quarterly (13), pp. 319-340.

De Burca, S., and McLoughlin, D., (1996), *The Grounded Theory Alternative in Business Network Research,* http://www.dcu.ie/business/research_interests.html

Dey, I., (1999), *Grounding grounded theory: Guidelines for qualitative inquiry*, San Diego Academic Press.

Eisenhardt, K.M., (1989), *Building theories from Case Study Research,* Academy of Management Review, volume 14, number4, pp. 532-550

Glaser, E.G., (1978), *Advances in the Methodology of Grounded Theory: Theoretical Sensitivity,* Sociology press, Mill Valley, California.

Glaser, E.G., and Strauss, A.L., (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research,* Weidenfeld and Nicplson, London, England

Hall, S., (1992), *New Ethnicities' - In Race, Culture and Difference,* Edited by A. Rattansi and J. Donald, Sage Publications

Hughes, J., and Jones, S., (2003), *Reflections on the use of Grounded Theory in Interpretive Information Systems Research*, Electronic Journal of Information System Evaluation, issue 1

Hughes, J., and Howcroft, D., (2000), *Grounded Theory: I mentioned it once but I think I got away with it,* UKAIS Conference, York

Hughes, J., and Wood-Harper, T., (1999), *Systems development as a research act,* Journal of information technology, 14, 83-94

Igbaria, M., Zinatelli, N., Cragg, P., and Cavaye, A., (1997), *Personal computing acceptance factors in small firms: A structured equation model,* MIS Quarterly, September

Jones, M., and Alony, I., (2011), *Guiding the Use of Grounded Theory in Doctoral Studies – An Example from the Australian Film Industry,* International Journal of Doctoral Studies, Volume 6.

Kwon, T.H., and Zmud, R.W., (1987), *Unifying the fragmented models of information systems implementation,* in critical issues in information systems research, R.J. Boland and R.A. Hirschheim (eds.), Wiley, New York

Lawrence, J.E., (2002), The Use of Internet in Small to Medium-Sized Enterprises, PhD thesis, Information Systems Institute, University of Salford, UK.

Lofland, J., and Lofland, L.H., (1984), *Analyzing social settings: A guide to qualitative observation and analysis*, (2nd edition), Wadsworth Pub. Co., Belmont, California

Martin, P.Y., and Turner, B.A., (1986), *Grounded theory and organisational research,* Journal of Applied behavioural science, vol.22 (2), pp. 141-157

Miles, M. B., and Huberman, A. M., (1994), *An Expanded Sourcebook: Qualitative Data Analysis,* Sage Publications Inc., Thousand Oaks, California

Miles, M. B., and Huberman, A. M., (1984), *Qualitative Data Analysis: A sourcebook of new methods,* Sage Publications Inc., Thousand Oaks, California.

Myers, M. D. (1997), *Qualitative Research in Information Systems*, Management Information Systems Quarterly, vol. 21, no. 2, pp. 221-242.

Oliphant, J. and Blockley, D. I. (1991), *Knowledge-Based System: Advisor on the Earth Retaining Structures*, Computers and Structures, vol. 40, no. 1, pp.173-183.

Orlikowski, W.J., (1993), *CASE tools as organisational change: Investigating incremental and radical changes in systems development,* MIS Quarterly, vol. 17, pp. 309-340

Orlikowski, W.J., (1991), *Integrated information environment or matrix of control? The contradictory implications of information technology,* Accounting, management and information technologies, vol. 1, no. 1, pp. 9-42

Orlikowski, W.J., and Baroudi, J.J., (1991), *Studying Information Technology in Organisations: Research Approaches and Assumptions,* Information Systems Research, 2:1, pp. 1-28.

Pettigrew, A., (1990), *Longitudinal field research on change: theory and practice,* Organisational science, 1:3, pp. 267-292

Pidgeon, N.F., Turner, B.A., and Blockley, D.I., (1991), *The use of grounded theory for conceptual analysis in knowledge elicitation,* International journal of Man-machine studies, vol.35 (2), pp. 151-173

Pries-Heje, J., (1992), *Three barriers for continuing use of computer-based tools in information systems development: a grounded theory approach,* Scandinavian journal of information systems, vol. 4, pp. 119-136

Spradley, J.P., (1979), *The Ethnographic Interview,* Holt, Rinehart and Winston, New York

Strauss, A.L, and Corbin, J., (1990, 1998), *Basics of Qualitative Research: Techniques and Procedures for developing Grounded Theory,* Sage Publications Ltd, London

Strauss, A.L, (1987), *Qualitative Analysis for Social Scientists,* Cambridge University Press, Cambridge

Taylor, S. J., and Bogdan, R., (1984), *Introduction to qualitative research methods: The search for meaning*, (2nd edition), New York, NY: John Wiley & Sons.

Torasker, K., (1991), *How managerial users evaluate their decision support: a grounded theory approach,* In Nissen, H.E., Klein, H., Hirschheim, R., (eds.), Information systems research: Contemporary approaches and emergent traditions, Proceedings of the IFIP WG 8.2 Working Conference, Copenhagen, 14-16 December, North-Holland, Amsterdam

Turner, B.A., (1983), *The use of grounded theory for the qualitative analysis of organisational behaviour,* Journal of management studies, vol. 20, no.3, pp. 333-348

Trauth, E. M. (2000), *The Culture of an Information Economy: Influences and Impacts in the Republic of Ireland*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Urquhart, C. (2001), *An Encounter with Grounded Theory: Tackling the Practical and Philosophical Issues*. In Trauth, E. (Ed.) Qualitative Research in Information Systems: Issues and Trends, Idea Group Publishing, London.

Walsham, G., (1993), *Interpreting Information Systems in Organisation,* John Wiley & Sons Ltd, Chichester, England.

Yin, R.K., (1984), *Case Study Research: Design and Methods,* Sage publications, Beverly Hills.

Zuboff, S., (1988), *In the age of smart machine,* Basic books, New York.

Using Photo-Elicitation to Explore Place Attachment in a Remote Setting

Joanna Tonge[1], Susan Moore[1], Maria Ryan[2] and Lynnath Beckley[1]
[1]School of Environmental Science, Murdoch University, Perth, Australia
[2]School of Marketing, Tourism and Leisure, Edith Cowan University, Perth, Australia
j.tonge@murdoch.edu.au
s.moore@murdoch.edu.au
m.ryan@ecu.edu.au
l.beckley@murdoch.edu.au

Abstract: People are often attracted to unique natural environments, but what makes them continually return to these locations, especially when considerable time and effort are required to get there? This paper discusses the methods and findings of a research project aimed at identifying and exploring how visitors develop an attachment to the remote Ningaloo Marine Park in north-western Australia. This Marine Park attracts a high percentage of repeat visitors (55%) and in order to determine the complex aspects contributing to this attachment, photo-elicitation was employed. Photo-elicitation is a qualitative technique where participants are asked to take photographs relating to the concept under study, and these are then used as triggers for underlying memories and feelings during a subsequent interview. For this study, participants were provided with digital cameras to take photographs of why they like visiting the Ningaloo Reef and what it was that made them return. Given this remote location and the inability to get photographs developed in reasonable timeframes, digital cameras were used instead of the disposable cameras more commonly used in this type of study. After a few days, the cameras were returned, and photographs uploaded on the researcher's laptop computer with interviews conducted while viewing the photographs. Over a period of four weeks, during the peak visitor period, 30 participants took over 200 photographs and provided over 15 hours of interview recordings. Key aspects contributing to place attachment included the beauty of the physical environment, reef and marine based activities, social bonding with family and friends and enjoying a challenging though rewarding experience. By using a technique familiar to people on holidays, i.e. taking photographs, a method was invoked that people could engage with easily without the research impinging on their holiday experience.

Keywords: photo-elicitation, photographs, interviews, marine, place attachment, remote location

1. Introduction

This paper reports on a novel methodology to examine and describe the relationships between people and place in a remote marine park in north-western Australia. Ningaloo Marine Park is located over 1,000km from the capital of Western Australia, but is still a popular destination for marine-based activities such as swimming, surfing, snorkelling and diving with many visitors returning year after year. Therefore, as the primary research population was people enjoying their leisure time, a method that was as unobtrusive as possible on visitors' experiences, while still providing information, was needed. Photo-elicitation using digital photographs taken by visitors to the Marine Park and subsequent interviews onsite proved to be a suitable technique, especially given the constraints of remoteness and unobtrusiveness while at the same time providing comprehensive information on peoples' relationships with Ningaloo Marine Park. The first part of this paper provides a rationale for examining place concepts in a remote marine setting, and is followed by an explanation of why photo-elicitation was a suitable technique. Included are details on why the technique was used, despite the tyranny of distance, and how this issue was overcome through the use of digital cameras and onsite interviews. Key aspects contributing to peoples' relationship with the Marine Park are also examined. The paper concludes with a discussion of the appropriateness and effectiveness of this technique in remote setting where knowledge is urgently required on the relationships between people and place.

2. The choice of photo-elicitation for Ningaloo Marine Park case study

Ningaloo Marine Park contains one of the largest fringing coral reef systems in the world and provides visitors with the opportunity to visit a unique marine environment relatively easily with only a shallow lagoon separating the reef from mainland Australia (Cassata & Collins 2008). Although the Marine Park is located approximately 1,200km north of Perth, the capital of Western Australia (Figure 1), approximately 200,000 people visit annually to participate in a range of nature-based activities including swimming, snorkelling,

fishing, boating and diving (Beckley, Smallwood, Moore & Kobryn 2010). A recent comprehensive survey of how people use the entire 300km length of the Marine Park identified that 55.2% of survey respondents had visited on a previous occasion and of these, 43.8% always stayed at the same location (Beckley et al 2010). But what is it that makes these visitors continually return to this remote Marine Park and even the same site time after time? One possible explanation explored in this study is place attachment, a concept that has emerged as a promising focus for exploring the relationships people have with their surrounding environment (Koons Trentelman 2009).

In exploring the attachment between visitors and their environment in this remote location, a number of research logistics were salient. As place attachment considers the relationship people have with specific locations, on-site rather than off-site methods were more suitable. In addition, the method chosen had to be able to provide the in-depth responses required to fully articulate people's attachment. Finally, visitors to the Ningaloo Reef region have been subject to numerous surveys as part of a national scientific research focus on the region (Beckley et al 2010) and there was a sense that visitors were being 'over researched'. Therefore the technique employed needed to ensure the detailed information through engaging visitors was obtained without them feeling their holiday experience was being impinged upon.

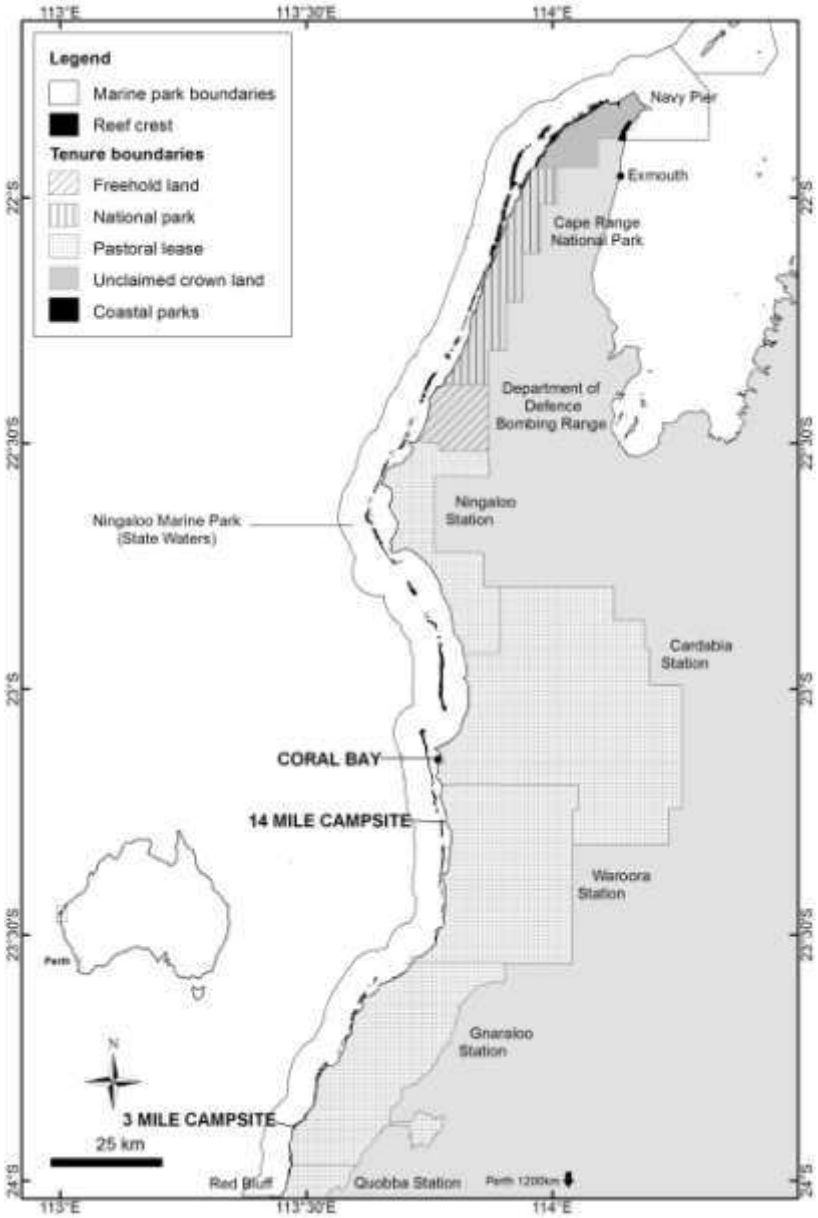

Figure 1: Map of Australia showing location of Ningaloo Marine Park and Perth, capital of Western Australia and Ningaloo Marine Park showing study site locations

2.1 Place attachment

Place attachment studies have been informed by a number of research traditions including philosophy, psychology, geography, sociology and natural resource management. For natural resource management, place attachment seeks to understand what type of experiences make people value places, what they seek when going or indeed returning to particular places and why some people may visit a particular place and others do not (Manzo 2008). A place can be described as a spatial setting that has been given meaning based on the experiences that people have there. These symbolic meanings underpin place attachment whereby people attribute meanings to a location and in turn people become attached to these locations (Stedman 2003; Stedman, Beckley, Wallace & Ambard 2004). As such, place attachment reflects the strong bonds which form between a person and a location (Smaldone, Harris & Sanyal 2008; Stedman et al. 2004).

Much of the natural resource management place literature has arisen from attempts to convince managers of the need to include the human, meaning-orientated dimensions of place in management decisions regarding recreational settings on publically-managed land (Koons Trentleman 2009). The emotional attachment that forms between a person and place usually involves strong sentiments that see them unwilling to substitute their 'place' for another and often results in heightened concerns about how the places are managed (Gunderson & Watson 2007; Williams 2008). It is important to understand these complex relationships to places as these insights assist in the effective and sensitive management of natural resources (Manzo 2008).

Qualitative research methods have been successful in exploring the make-up of people's relationships with places. Such methods are more sensitive to the intricate distinctions in the meanings ascribed to places than quantitative measures (Manzo 2008). However, simply asking someone to outline "why are you attached to this place?" may not reveal the extent of the experiences and information that went into forming their attachment (Beckley, Stedman, Wallace & Ambard 2007). This has led some place researchers to experiment with visual methods to help articulate what constitutes peoples' attachment to places. For example, Beckley et al. (2007) used resident-employed photography, a form of photo-elicitation, to help residents communicate what formed their attachment to their communities.

Visual research methods have been used as a research strategy for studying the inter-related aspects of people's lived experiences by allowing them to express themselves in diverse ways (Dennis, Gaulocher, Carpiano & Brown 2009). There are a variety of approaches in visual research methods ranging from those that have a high level of researcher intervention (e.g. researchers choose images) to those with a high level of participant input (e.g. participant creates or collects images). The method employed in this study was photo-elicitation, which is based on the simple premise of inserting a photograph into an interview (Harper 2002; Loeffler 2004; Rose 2007).

2.2 Photo-elicitation

First named by John Collier in the 1950s (Harper 2002; Loeffler 2004; Rose 2007), photo-elicitation was devised to illustrate categories of housing to study participants in his study on mental health in changing communities (Harper 2002). Since then it has been used to conduct research in anthropology, education, community health, psychology, sociology and natural resource management (Loeffler 2004). In most cases, participants are provided with a camera and asked to take photographs to illustrate their interpretation of the subject or theme of interest to the researcher (Garrod 2008; Rose 2007). The photographs are then used as stimuli in a subsequent interview to evoke or elicit responses of a greater emotional depth than those that may have been prompted by more traditional interview methods (Jacobsen 2007).

The very act of framing the photograph assists participants to see their everyday experiences in new ways (Dennis et al. 2009). They reflect the participants' point of view, their biases and experiences and can also act as a memory trigger by preserving a moment in time. This helps the participants to sharpen reflections on their experiences and resultant discussion (Loeffler 2004). Photographs also capture a greater amount of detail than participants can remember on their own. Images can evoke deeper elements of people's experience than words alone to provide a greater understanding of the concept under study (Loeffler 2004). The meaning or the intent behind the images is not always apparent simply by looking at them (Stedman et al 2004) and often it is only through the accompanying interview and discussion that the information inherent in the images can be obtained by the researcher (Rose 2007). In addition, discussing the images can often prompt further

conversation on other related issues (Rose 2007) or enable respondents to express feelings they may have found hard to verbalise through traditional interview methods (Garrod 2008).

The success of the photo-elicitation method can be attributed to a number of factors, including that the research process is often an enjoyable experience for both the participants and the researchers (Beckley et al. 2007). For the participant, the research task of taking photographs is an enjoyable and familiar activity (Stedman et al. 2004) and, being task-orientated, participants engage deeper into research about their own experiences (Dennis et al. 2009). This can increase their initial willingness to participate and their subsequent engagement with the research process (Garrod 2008; Stedman et al. 2004). In addition, the interview focuses on the photographs rather than the participant, which can remove some of the pressure of being interviewed from the participant. It also provides an opportunity for participants to feel valued in a non-evaluative and non-judgemental environment (Dennis et al. 2009). Finally, the researcher becomes the listener while the participant interprets the images (Loeffler 2004).

3. Applying photo-elicitation to Ningaloo Marine Park case study

3.1 Study sites

Three study sites in the southern section of Ningaloo Marine Park (Figure 1) were selected as fewer studies have been undertaken in this area because it is more difficult to access, with two of the sites requiring a four-wheel drive vehicle. The first study site, Coral Bay, is a small township with a major focus on providing tourism experiences including glass-bottom boat tours, fishing charters and four-wheel drive tours (WAPC 2004). The two other study sites are located on pastoral (sheep) stations which abut the coastline boundary of the Marine Park, 3 Mile Campsite on Gnaraloo Station and the 14 Mile Campsite on Warroora Station. The 3 Mile Campsite offers beachside camping with unpowered sites and simple facilities including bore-water showers and toilets with basic laundry and cleaning facilities. The 14 Mile Campsite on Warroora Station provides coastal camping at its most basic. All visitors must be self-sufficient in relation to water, food and power as no facilities are provided.

3.2 Data collection

The photo-elicitation research process was undertaken on-site over a four week period in July 2009, with approximately a week to ten days spent at each of the three study sites. July was chosen to coincide with the peak visitor period (Beckley et al 2010). Adult visitors at each of the three sites were approached and asked if they had stayed at the site at least twice previously, as attachment to a place starts to develop after one visit (Gunderson & Watson 2007). If the visitor met this criterion, and agreed to participate in the study, they were provided with a camera and asked to take six to eight photographs of what makes the place special for them. An instruction sheet outlining what was required of them as well as contact details for the researchers was provided at the time of camera handover. A general interview time (e.g. day, morning or afternoon) was also agreed upon at this time.

The interviews were conducted in a semi-structured manner, with a list of questions to act as probes determined *a priori* but not strictly adhered to in terms of the order or exact wording. The interviews began with a brief description of the participant's socio-demographic details (gender, life-cycle stage etc) and their travel group type (family, friends, etc). An in-depth discussion of the photographs followed to capture the intent and meanings behind the images. Further questions were asked to clarify responses or to follow up on interesting aspects. Interviews were digitally recorded for transcription upon return from the field. At the conclusion of the interview, participants were provided with a CD containing their photographs.

3.3 Analysis

The photographs taken by the participants were coded to describe their content. The researcher consulted the interview transcripts for the participant's description of the photograph, e.g. "beach fishing" or "family sitting around the campfire". Once initial coding of all photographs was complete, these were then rolled into the overarching categories adapted from the place attachment literature - recreational activities, social situations and the physical environment. Then each was categorised as being marine or terrestrial in nature. To illustrate, a photograph of fishing rods in the sand on the beach (as per Figure 2) ended up being coded as "recreational activity – marine". However, due to the nature of some the photographs, they were given two codes. For

example, people on a boat was described as "marine – recreational activity" and "marine – social situation". The researcher did not see the photographs before the interviews as they were loaded onto the laptop computer at the time of the interview. Coding of the photographs was undertaken prior to coding of the interview transcripts.

The interviews were transcribed verbatim and were then read individually to allow the researcher to obtain a general understanding of the content before undertaking coding. Coding, the classification of individual pieces of data (Babbie, 2005), identified key blocks of text which could then be conceptually placed in categories or themes (Garst, Williams, & Roggenbuck, 2010). Once key blocks of text and categories were identified, each interview was then re-read to ensure consistency across all transcripts and to expand or condense categories as required. This was undertaken using the QSR N'Vivo (version 2.0) software program. The coding of the photographs informed the coding of the interview transcripts in addition to the natural resource management literature associated with place attachment and place meanings.

3.4 Studying place attachment in a remote setting

Given the remote location and difficulty in accessing the study sites, it was not feasible to use disposal film cameras, as has been the approach in other photo-elicitation place studies. A considerable amount of time would have been lost in travelling to the nearest towns (at least 100km away) to get the films developed for use in the subsequent interviews. Staying in remote locations is costly; therefore time management was a crucial part of the research design. As such, digital cameras helped overcome this issue. Such cameras allowed the photos to be uploaded onto the researcher's laptop computer for viewing during the interview process.

As all digital cameras are different in the way they operate, tags were attached providing instructions in how to take, view and delete photographs. The researcher's name and telephone number were also included, with the researcher staying on-site to assist with any technical difficulties or queries (e.g. changing batteries, etc). Digital cameras may also help to overcome the sampling bias proposed by Jacobsen (2007) in relation to film cameras. As film cameras only have a fixed number of exposures, participants may not take all the photos that represent their experience. They may take too many photographs of one particular aspect and not have any exposures left to capture some other important aspect at a later time. Alternatively, they may not take photos of some aspects of their experience, concerned that they may not have an exposure left for the perfect shot later on. Digital cameras allow participants to take as many photographs as they wish as there are no fixed number of exposures, they can then choose from these those best representing their experience.

4. Results

A total of 30 participants across the three study sites contributed to this study. The participants were part of three different types of travel groups – family (47%), couples (33%) and with friends (20%). Over half of the participants (57%) were from Perth, 23% from regional Western Australia and the remainder from interstate and overseas. Most participants (70%) visited at least once per year, followed by more than once per year (23%) and once every two years (7%).

4.1 Overview from the photographs

Participants took 207 photographs with the most provided by one participant being 16, the least, three. The types of photographs included landscapes, seascapes, beach scenes, family gatherings, fishing rods, surfboards and sunsets. Following coding of the photographs, three overarching categories were evident – the physical environment, recreational activities and social situations, examples of which are provided in Figure 2.

The marine aspects of the physical environment dominated the photographs, with 82% being marine-based (Table 1). There was a more equal split between marine and terrestrial for recreational activity and social situations (Table 1) which suggests the adjacent hinterlands are important for these participants as they use both environments.

Analysis of the photographs and interviews together followed, as the meaning of photographs is elaborated and built upon in the accompanying stories (Garrod 2008, Loeffler 2004). This integrated analysis provided a fourth category – emotional connections with Ningaloo Reef – additional to the previously identified physical environment, recreational activities and social situations.

Figure 2: Examples of participants' photographs illustrating the three overarching categories of the physical environment, recreational activities and social situations.

Table 1: Categorisation of participants' photographs

Photograph category	Description	Total*	Marine-based (% of Total)	Non marine-based (% of Total)
Physical environment	Contains aspects of environment, can include people but predominant feature is physical environment	94	82	18
Recreational activities	People participating in a recreational activity OR containing equipment to undertake activity	78	41	59
Social situations	Groups of people interacting, not based around a recreational activity	102	43	57

* - Photos were placed in more than one category, as such sums to 274 rather than 207

4.2 Physical environment: Beauty in remoteness

The remoteness of the location and coastline free from people provided an escapist feeling. With the surrounding environment being so expansive, they enjoyed the impression of a wild environment, untamed by humans:

> *"That is in the lagoon (Figure 3), wild place, with this light in the background, with just wilderness..."*

Participants enjoyed being able to visit a location, which through geography and lack of human presence, gave the feeling of isolation and escape from their busy everyday lives. Being able to put psychological and geographical distance between themselves and their normal everyday life (Korpela & Hartig, 1996) and being in a natural setting provided a reduction in visitors' stress and anxiety (Garst et al 2010). The naturalness of the setting helps to give these places an "escape" feel rather than more developed or populated places that lend themselves to a "social" feel (Stedman, 2003).

Figure 3: Participant's photograph showing the lagoon

4.3 Recreational activities: Reef-based activities

Participants enjoyed being able to undertake their favourite reef and marine based activities, such as surfing, fishing, snorkelling and diving in a convenient location so close to shore.

> *"We like the snorkelling. It is an amazing reef, you just go out there a couple of 100 meters and you have got an amazing coral reef. You can't do that anywhere else." (Coral Bay participant)*

Some of these marine-based activities, such as surfing (Figure 4), formed their main reason for visiting – *"we are here to surf" (3 Mile participant)*. One participant even indicated that if the surfing conditions were not right and their friends did not come, they would not visit the Ningaloo Reef region at that time:

> *"The surf as well, but if you said why wouldn't you come back, it would be onshore (wind) and none of the friends would turn up, if we knew that the surf was not going to be on and none of our friends would be here, that would be our criteria for nah, no, not going…" (3 Mile participant)*

Figure 4: Surfing is an important aspect of some participants' experiences

Recreational activities are an important component in the development of place attachment (Eisenhauer, Krannich & Blahna 2000). The convenience, accessibility and ability of the place to facilitate desired activities can help to build visitors' attachment. Often visitors participating in activities requiring specific types of places, such as good surfing locations as per this study, may exhibit greater levels of attachment than an indifferent visitor (Farnum, Hall & Kruger 2005).

4.4 Social situations: Maintaining family bonds

Many participants commented that they saw holidays as an important opportunity for them to spend time with their children, especially when the everyday stresses of life made it hard to do this in their home environment. There were also a number of activities that they could participate in as a family, which helped to maintain and reaffirm bonds:

> *"No, no it is not the fishing, it is the mix of things. Sort of being a family away from the emails and all of that sort of thing, the diversity of things, snorkelling on the reef would be quite high on the list and that is something I can do with anybody in the family" (3 Mile participant)*

After travelling to the same location year after year, traditions can begin to form. What started off as a family holiday with relations developed to the second and third generations as the 'original' children grew up and started having families of their own.

> *"Researcher (R): And the families all come up at the same time of year do they?*
>
> *Participant (P): Yeah, it is second and third generations coming up now*
>
> *R: So this is sort of a real family catch-up for you*
>
> *P: Yes, cousins, aunties, uncles…*
>
> *R: Did you originally start the tradition?*
>
> *P: We started it many years ago with my brother-in-law and my husband's sister and two sisters, their husbands and their kids. And then their kids married…" (Coral Bay participant)*

Travelling and holidays are often part of a family tradition (Lee, 2001), and as families continually visit the same site, positive interactions and memories become associated with it, forming significant meanings of place (Brooks, Wallace, & Williams, 2006).

4.5 Emotional connection: Challenging but rewarding experience

Another strong aspect related to the effort and planning required coming to the Ningaloo Reef region and having a holiday – *"holidays here are hard work, getting here and all the rest of it" (14 Mile participant)*. Rather than detracting from the experience, this seemed to add to the enjoyment due to the challenge it provided – *"that probably adds to the experience, that you have to plan ahead" (Coral Bay participant).* Having to work for the experience was also seen as a mechanism that protected the participants' experiences from being over-run by busloads of tourists or outsiders, spoiling the isolated and remote character of the locations:

> *"but I think to seriously enjoy it you have got to go to a bit of effort, if the need for that effort is taken away, then we are going to see busloads [of tourists]" (3 Mile participant).*

Overcoming obstacles and challenges to stay at these sites, such as large travelling distances and bringing enough food, water and shelter, provided a sense of fulfilment for participants. The effort put in to planning and preparing for the trip resulted in the reward of having wonderful experiences in these remote locations. The essence of a place can be expressed and constrained through the effects of the physical parameters (Sampson & Goodrich, 2009), such as a place that is hard to get to or a place that is remote with few or no facilities. Places can be characterised by the opportunities they provide as well as by the limits they impose on human behaviour and social activity (Sampson & Goodrich, 2009).

5. Review of photo-elicitation research method

5.1 Place attachment

This work shares a number of similarities with other natural resource management studies (e.g. Eisenhauer et. al (2000), Smaldone et al. (2008)) in the aspects contributing to place attachment, i.e. the physical environment, recreational activities, social situations and emotional connections. Visitors to natural

environments are drawn by and depend on the physical environment (Manzo 2008) and one which allows them to undertake specific or desired activities (Jacobsen 2007). Additionally the physical environment can instil a sense of escape, relaxation or achievement in overcoming limitations in visitors. And, the spatial setting provides the opportunity to reaffirm social relationships and reconnect with family and friends (Garst et al. 2010).

5.2 Photo-elicitation as a method

Visitors in this study were very amenable to taking photographs, possibly as this is often an activity undertaken on holidays. Scheduling interviews a few days after the cameras were distributed appeared to keep the purpose of the study at the forefront of participants' minds and provided lively and in-depth discussion of the photographs taken. From the review of the photographs taken, it became apparent that the participants valued the physical environment, particularly marine, as an attraction. Also apparent was the significance of the adjacent beach and hinterlands as a locale for participating in activities and facilitating social situations. Participants obviously value both the marine and terrestrial settings.

As mentioned in the Methods, the researcher conducting the interviews did not see the photographs beforehand. Such an approach minimises any inherent bias resulting from the researcher having developed preconceived ideas about the photographs and participants' association intentions prior to the interviews. This also added to the enjoyment of the participant and the researcher as the photographs were viewed together for the first time on the laptop computer. Participants were also encouraged to provide as much discussion of the photographs as they wished, generating in-depth and emotive responses from the participants.

Photo-elicitation does, however, have limitations. Aspects that are difficult to photograph such as sounds, smells or even the emotions felt at the time cannot be adequately represented (Dorwart, Moore & Leung 2006). Interpretation of photographs also relies heavily on what is seen or captured in the photograph, but what about those aspects that are not captured? Indeed, what events, aspects would go un-photographed because of the type of setting, limitations of equipment or the nature of the activity (Loeffler 2004). In this study, the cameras provided were not water-proof; so participants could not take photographs underwater. To try and overcome this limitation, an interview question was included asking participants to indicate whether there was anything they couldn't photograph. However, this question may not always accurately capture the meanings inherent in such uncaptured images.

Finally, while using digital cameras overcame a number of limitations associated with conducting research in a remote setting, their use may not be applicable in every situation. Procuring the cameras, memory cards and batteries is costly. There is also the probability of not receiving all of the cameras back once participants have taken their photographs. Thankfully, this was not an issue for this study which may have been influenced by the researcher staying on-site and being highly visible around the campsites. This may not be possible in a more 'open' study location or in more highly populated campsites.

6. Conclusion

Photo-elicitation is an effective tool for investigating people-place related constructs as often the elements participants are asked to photograph are things they may have been inclined to photograph anyway (Beckley et al. 2007). It has been proven successful in investigating how people make sense of and comprehend the environment around them (Dennis et al. 2009) by empowering the participants to take some aspect of control of the research process resulting in richer and more meaningful data on people-place constructs (Dorwart et al. 2006). The use of modern technology, digital cameras and laptop computers, allowed for this under-utilised technique to be applied in a remote setting. This study provides another example of the successful application of visual methods contributing to the continued advancement of people-place theory.

References

Babbie, E.(2005) *The Basics of Social Research*, 3rd ed. Thomson Wadsworth, Belmont, CA.

Beckley, L.E., Smallwood, C.B., Moore S.A. and Kobryn, H.T. (2010) "Ningaloo Collaboration Cluster: Human use of Ningaloo Marine Park". Ningaloo Collaboration Cluster Final Report No. 2. 166 pp.

Beckley, T., Stedman, R.C., Wallace, S., and Ambard, M. (2007) "Snapshots of What Matters Most: Using Resident Employed Photography to Articulate Attachment to Place", *Society & Natural Resources,* Vol. 20, no. 10, pp 913-29.

Brooks, J.J., Wallace, J.J., and Williams, D.R. (2006) "Place as a Relationship Partner: An Alternative Metaphor for Understanding the Quality of Visitor Experience in a Backcountry Setting," *Leisure Sciences,* Vol. 28, no. 4, pp 331-49.

Cassata, L., and Collins, L.B. (2008) "Coral Reef Communities, Habitats, and Substrates in and near Sanctuary Zones of Ningaloo Marine Park," *Journal of Coastal Research,* Vol. 24, no. 1, pp 139-51.

Dennis, S. F., Gaulocher, S., Carpiano, R.M., and Brown, D. (2009) "Participatory Photo Mapping (Ppm): Exploring an Integrated Method for Health and Place Research with Young People", *Health and Place,* Vol. 15, no. 2, pp 466-73.

Dorwart, C. E., Moore, R. L., & Leung, Y.-F. (2006) "Visitor Employed Photography: Its Potential and Use in Evaluating Visitors' Perceptions of Resource Impacts in Trail and Park Settings", Paper presented at the 2006 Northeastern Recreation Research Symposium, Bolton Landing, NewYork.

Eisenhauer, B. W., Krannich, R. S., & Blahna, D. J. (2000) "Attachments to Special Places on Public Lands: An Analysis of Activities, Reason for Attachments, and Community Connections", *Society & Natural Resources,* Vol 13, No. 5, pp421-441.

Farnum, J., Hall, T., & Kruger, L. E. (2005) *Sense of Place in Natural Resource Recreation and Tourism: An Evaluation and Assessment of Research Findings* (Report No. PNW-GTR-660), US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon.

Garrod, B. (2008) "Exploring Place Perceptions: A Photo-Based Analysis", *Annals of Tourism Research,* Vol. 35, no. 2, pp 381-401.

Garst, B.A., Williams, D.R. and Roggenbuck, J.W. (2010) "Exploring Early Twenty-First Century Developed Forest Camping Experiences and Meanings", *Leisure Sciences,* Vol. 32, no. 1, pp 90-107.

Gunderson, K. and Watson, A.E. (2007) "Understanding Place Meanings on the Bitterroot National Forest, Montana", *Society & Natural Resources,* Vol. 20, no. 8, pp 705-21.

Harper, D. (2002) "Talking About Pictures: The Case for Photo Elicitation", *Visual Studies,* Vol. 17, no. 1, pp 13-26.

Jacobsen, J. K.S. (2007) "Use of Landscape Perception Methods in Tourism Studies: A Review of Photo-Based Research Approaches", *Tourism Geographies,* Vol. 9, no. 3, pp 234-53.

Koons Trentelman, C. (2009) "Place Attachment and Community Attachment: A Primer Grounded in the Lived Experience of a Community Sociologist", *Society & Natural Resources,* Vol. 22, no. 3, pp 191-210.

Korpela, K. and Hartig, T. (1996) "Restorative Qualities of Favorite Places", *Journal of Environmental Psychology,* Vol. 16, no. 3, pp 221-33.

Lee, C.C. (2001) "Predicting Tourist Attachment to Destinations", *Annals of Tourism Research,* Vol. 28, no. 1, pp 229-32.

Loeffler, T. A. (2004) "A Photo Elicitation Study of the Meanings of Outdoor Adventure Experiences", *Journal of Leisure Research,* Vol. 36, no. 4, pp 536-56.

Manzo, L.C. (2008) "Chapter 7 - Understanding Human Relationships to Place and Their Significance for Outdoor Recreation and Tourism", In *Understanding Concepts of Place in Recreation Research and Management*, edited by Linda E Kruger, Troy Hall and Maria Stiefel, US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon.

Rose, G. (2007) "Chapter 11: Marking Photographs as Part of a Research Project - Photo-Elicitation, Photo-Documentation and Other Uses of Photos", In *Visual Methodologies: An Introduction to the Interpretation of Visual Materials,* Sage Publications, London.

Sampson, K.A., and Goodrich, C.G. (2009) "Making Place: Identity Construction and Community Formation through "Sense of Place" in Westland, New Zealand", *Society & Natural Resources,* Vol. 22, no. 10, pp 901-15.

Smaldone, D., Harris, C. and Sanyal, N. (2008) "The Role of Time in Developing Place Meanings", *Journal of Leisure Research,* Vol. 40, no. 4, pp 479-504.

Stedman, R.C. (2003) "Is It Really Just a Social Construction?: The Contribution of the Physical Environment to Sense of Place", *Society & Natural Resources,* Vol. 16, no. 8, pp 671-85.

Stedman, R.C., Beckley, T., Wallace, S. and Ambard, M. (2004) "A Picture *and* 1000 Words: Using Resident-Employed Photography to Understand Attachment to High Amenity Places", *Journal of Leisure Research,* Vol. 36, no. 4, pp 580-606.

Western Australian Planning Commission (WAPC).(2004) "Ningaloo Coast Regional Strategy: Carnarvon to Exmouth", Western Australian Planning Commission, Perth.

Williams, D.R. (2008) "Chapter 2 - Pluralities of Place: A User's Guide to Place Concepts, Theories and Philosophies in Natural Resource Management", In *Understanding Concepts of Place in Recreation Research and Management*, edited by Linda E Kruger, Troy Hall and Maria Stiefel, US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon.