Department of Statistical Sciences
University of Padua
Italy

Latent class analysis for evaluating a multi-item scale to measure customer satisfaction with reference to a shopping good: a pair of branded jeans

Francesca Bassi
Department of Statistical Sciences
University of Padua
Italy

Luca Volini
Department of Statistical Sciences
University of Padua
Italy

Abstract: In the field of marketing many objects of interest exist that are not directly observable, nevertheless they can be measured through multi-item measurement scales. As a consequence, this kind of instruments are extremely useful and their importance requires an accurate development and validation procedure. The traditional marketing literature highlights specific protocols along with statistical instruments and techniques to be used for achieving this goal. For example, correlation coefficients, univariate and multivariate analysis of variance and factorial analysis are widely employed with this purpose. However, these kind of statistical tools are usually suited for metric variables but they are adopted even when the nature of the observed variables is different, as it often occurs, since in many cases the variables measured by the items of which the scale is made up are ordinal. On the contrary, latent class analysis takes explicitly into account the ordinal nature of the observed variables and also the fact that the object of interest, that has to be measured, is unobservable. The aim of this paper is showing how latent class analysis can improve the procedures for developing and validating a multi-item measurement scale for measuring customer satisfaction with reference to a shopping good that is a good characterized by a high level of involvement and an emotional learning, linked to the lifestyle of the customer. This latent class approach explicitly considers both the ordinal nature of the observed variables and the fact that the construct to be measured is not directly observable. Especially, applying appropriate latent class models, important features such as scale dimensionality, criterion and construct validity can be better assessed while evaluating the scale.

Keywords: measurement scale, satisfaction, shopping goods, validity, reliability

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Final Version (2015-06-22)

Contents

Latent class analysis for evaluating a multi-item scale to measure customer satisfaction with reference to a shopping good: a pair of branded jeans

Francesca Bassi
Department of Statistical Sciences
University of Padua
Italy

Luca Volini
Department of Statistical Sciences
University of Padua
Italy

Abstract: In the field of marketing many objects of interest exist that are not directly observable, nevertheless they can be measured through multi-item measurement scales. As a consequence, this kind of instruments are extremely useful and their importance requires an accurate development and validation procedure. The traditional marketing literature highlights specific protocols along with statistical instruments and techniques to be used for achieving this goal. For example, correlation coefficients, univariate and multivariate analysis of variance and factorial analysis are widely employed with this purpose. However, these kind of statistical tools are usually suited for metric variables but they are adopted even when the nature of the observed variables is different, as it often occurs, since in many cases the variables measured by the items of which the scale is made up are ordinal. On the contrary, latent class analysis takes explicitly into account the ordinal nature of the observed variables and also the fact that the object of interest, that has to be measured, is unobservable. The aim of this paper is showing how latent class analysis can improve the procedures for developing and validating a multi-item measurement scale for measuring customer satisfaction with reference to a shopping good that is a good characterized by a high level of involvement and an emotional learning, linked to the lifestyle of the customer. This latent class approach explicitly considers both the ordinal nature of the observed variables and the fact that the construct to be measured is not directly observable. Especially, applying appropriate latent class models, important features such as scale dimensionality, criterion and construct validity can be better assessed while evaluating the scale.

Keywords: measurement scale, satisfaction, shopping goods, validity, reliability

1. Introduction

One of the most important constructs of interest in marketing research is customer satisfaction because firms build a wide part of their competitive advantage on it; nevertheless, the concept is multidimensional and not directly observable and, like many other relevant objects in the marketing field, it has to be measured through multi-item scales.

Usually, this kind of scales are developed following traditional protocols and the statistical methodology outlined in the literature often does not take explicitly into account the actual nature of the variables involved. For example, many of them are suited for metric variables while the observed variables are often ordinal. In this paper, we show how latent class analysis (McCutcheon, 1987) can improve the development and validation procedures of a measurement scale with reference to a shopping good, that is a good characterized by a moderate purchase frequency, mid/high-level price and it is linked to the lifestyle of the customer who feels strong involvement and, as a consequence, evaluate goods belonging to this category more often than the goods belonging to the others (Bagozzi & Ruvio, 2011). In particular, latent class analysis considers explicitly both these aspects, the fact that the construct is multidimensional and not directly observable and, in addition, the fact that customer satisfaction is usually measured with ordinal items. A previous work on this topic (Bassi, 2011) reveals that latent class analysis brings to different results from those obtained with the traditional protocol, when applied for assessing validity and reliability properties of a scale for measuring customer satisfaction with reference to an experiential good, like a movie seen at the cinema. Experiential goods distinguish themselves, compared to shopping goods, because they raise weak involvement in the customer even if both experience and shopping goods are characterized by emotional learning. Starting from this evidence, we want to study if such result can occur even when evaluating a multi-item scale related to a shopping good. The measurement scale considered in this paper was designed to measure customer satisfaction with reference to a pair of branded jeans.

The paper is organized as follows. It starts with a brief description about products' classification followed by the description of the multi-item measurement scale considered here; the following section reports the results obtained following the traditional protocols for validation and the third one is committed to the latent class approach. In this section, latent class models for evaluating scale's validity and reliability will be shown as well as the outcomes of the analyses carried out with latent class models' support. The paper finishes with some concluding remarks.

2. Product classification

Usually, firms, in order to make potential customers choose their own products, use marketing-mix strategies that can be interpreted like stimuli producing different kind of responses by potential customers, depending on the particular physical and intellectual activity they provoke.

Since in the marketing field it is important to define the process that leads to a particular response in order to evaluate the congruence of the efforts undertaken, customer responses are split into a sequence of three stages: cognitive, emotional and behavioral (Tyagi & Kumar, 2004). The first one involves the knowledge and the information held by a customer, the second one is linked with customer's own preferences and subjective evaluation, the last stage describes the purchase and post-purchase behavior. However, this sequence may differ if two more features are taken into account, that is the degree of involvement (weak or strong) and the type of learning (intellectual or emotional). The intellectual learning is based on rationality while the emotional one is based on emotions and insight. Both these aspects are almost always present at the same time but, depending on the kind of product, they have different weight.

Considering the degree of involvement and the type of learning together, it is possible to define several response paths and, as a consequence, a scheme for product classification (Ferber & Holton, 1958). When there is a strong involvement along with an intellectual learning the response sequence is knowledge-evaluation-action and it is suited for durable goods having a high price, called specialty goods. If the involvement is still strong but there is an emotional learning, the response path becomes evaluation-knowledge-action or even evaluation-action-knowledge. These response paths are suited for goods, referred to as shopping goods, linked with the lifestyle of the customers who choose them because they reflect both their values and the image they want to show, like the pair of branded jeans considered in this paper. When, on the contrary, the degree of involvement is weak and the type of learning is intellectual, we have the following response path, that is action-knowledge-evaluation. This sequence characterizes goods purchased frequently and having a low price, referred to as convenience goods. Finally there are goods characterized by a low degree of involvement and an emotional learning that produce response as a sequence like action-evaluation-knowledge, that are products linked with the hedonistic sphere called experience goods.

The degree of involvement (Zaichkowsky, 1985) and the type of learning as well as other aspects such as the degree of distinction between brands, experience, purchase frequency and perceived risk, that are not obviously the same for each category of goods, determine different type of purchase processes in terms of different importance and duration of each of their stages. Indeed, a purchase process can be represented by a sequence of steps that spread from need recognition to post-purchase evaluation (Wilkie, 1990). This sequence of steps describes the entire consumption experience and it is useful in order to develop effective marketing strategies. The steps are the following: need recognition during which the customer perceives a need that must be satisfied usually as a consequence of a gap between the actual condition and the desired one; information search that is the step when information about possible alternatives are collected; evaluation of the alternatives, linked with the evaluation of the available products in order to choose the one that fits better for satisfying customer's need; purchase decision, that is the act of purchasing; and post-purchase evaluation during which the chosen product is evaluated taking into account the entire consumption experience; this last stage is really important because it can have a strong impact on firm's competitive advantage.

3. A scale to measure customer satisfaction with reference to a shopping good

The scale considered in this paper aims at measuring customer satisfaction with reference to a shopping good represented by a pair of branded jeans. This kind of goods are characterized by a moderate purchase frequency since they are purchased just occasionally, and have a mid/high-level price. Moreover, the purchase of these goods is preceded by weighting and selection, since customers compare available alternatives on the basis of price level, style and convenience. The degree of involvement is strong and the way of learning is emotional. Because of their peculiar nature, goods belonging to this category are evaluated more often than the other kinds of goods (Bettman, Johnson & Payne, 1991).

The scale considered here is made up of 23 items referring to all phases composing the consumption experience. The paradigm used here for defining customer satisfaction can be seen as an extension of the traditional disconfirmation one. In particular, it treats customer satisfaction as the positive result of comparing expectations with the entire consumption experience, and not only with product performance as perceived by customers. This means that the comparative term is extended to include all aspects of consumption experience, not merely product performance (Guido, Bassi & Peluso, 2010).

Scale's items can be grouped considering each different phase of consumption experience, this leads to five sets of items. Items named E1-E2 (see Appendix) relate to the initial phase of consumption experience when the customer recognizes to have a specific need to be satisfied and explores aspects of major influence on it. Items R1-R6 regard the phase of collecting information and the ways through which this is achieved, pointing out parameters considered for evaluating information themselves, such as clearness, reliability, accessibility and so on. Items V1-V4 are linked with the third phase of comparing different options and examining standards used by customers for selecting between them. Items U1-U5 regard purchase evaluation focusing on what makes customers buy a specific product. Finally, items P1-P6 refer to post-purchase evaluation. In addition, in order to evaluate criterion and construct validity four more items were included. Item S1 intends to measure customer satisfaction with reference to the entire consumption experience. Items C1, C2 and C3 concern repurchase intention, positive word of mouth and absence of complaints. Respondents were asked to express their judgement on each item on a seven-point Likert scale where 1 means "completely not satisfied" and 7 "completely satisfied". Data were collected on a (non-probabilistic) sample of 300 units. Many questionnaires (250) were administered by an interviewer out of retail stores, while the remainders 50 ones were sent by e-mail.

At the beginning, scale properties were evaluated using traditional protocols, focusing on scale reliability and (criterion and construct) validity (De Vellis, 1991). With reference to scale reliability, traditional factor analysis highlighted the presence of one latent factor capable of explaining about 32% of the variance between items with factor loadings higher than the threshold equal to 0.35. This result led to conclude that the construct to be measured was unidimensional; furthermore, item-to-total correlation coefficients were higher than 0.30. Cronbach's alpha was equal to 0.893

and split-half indexes of internal consistency like, Split-half R, Spearman-Brown Y and Guttman G, took on the following values 0.674, 0.806 and 0.805 respectively, showing up that the scale could be considered reliable. Further analyses based on split-half sample procedures led to the same conclusions. Indeed, comparing values assumed by reliability coefficients above mentioned, reported in Table 1, it could be seen they don't significantly differ between the two subsamples obtained splitting the starting set of 300 units; moreover, the means of the items and of the total scale score were not statistically different in the two subsamples since the p-values associated with the t-statistic were higher than 0.05.

	Cronbach α	Split-half R	Spearman-Brown Y	Guttman G
Subsample 1	0.886	0.689	0.816	0.815
Subsample 2	0.900	0.666	0.800	0.799

Table 1 – Reliability coefficients computed within each subsample in split-half sample procedure

In order to assess criterion validity, an additional item (S1) was introduced asking respondents to express their satisfaction with the entire consumption experience. Both correlation analysis and analysis of variance were carried out. On one hand, the correlation coefficient between the average scale score and the criterion variable was equal to 0.721. On the other hand, the analysis of variance suggested that the average scale scores within groups defined by the levels of the criterion variable were statistically different from one another due to the high F-statistic value ($F = 65.949$, $p < 0.001$), confirming the property of criterion validity.

For evaluating construct validity, three additional items were included in the questionnaire. Each of them aimed at measuring constructs theoretically linked with customer satisfaction; in particular, repurchase intention, positive word of mouth and absence of complaints. Like criterion validity, even construct validity was assessed using correlation analysis and analysis of variance. The results obtained carrying out these two kinds of analyses, defined within traditional protocols, were the following. Correlation coefficients between the total scale score and the additional items C1, C2 and C3 were 0.628, 0.700 and 0.602, respectively; while the analysis of variance's outcome showed that different levels of satisfaction had a statistically significant effect on control variables. Our total scale score was classified into three categories: low (total score ≤ 99), medium ($100 - 122$) and high (≥ 123), according to the quartiles of the distribution. Furthermore, post-hoc tests led to conclude that the average scores on additional items increased significantly when satisfaction level becomes higher, concluding that construct validity was confirmed.

The goal of this paper is discussing these results showing how latent class analysis can improve the evaluation of a scale for measuring customer satisfaction with reference to a shopping good. This approach differs from the traditional one just described since it considers explicitly the ordinal nature of the observed variables and the fact the object to be measured that is customer satisfaction, is not directly observable.

4. Latent class models

Latent class (LC) analysis provides models that consider explicitly the fact that one or more latent variables exist which are not directly observable when studying relationships between observed variables, and take into account the categorical nature of these variables. Since items which made up a measurement scale often generate ordinal variables and the construct to be measured is not directly observable, these models seem to fit well in order to develop and validate a multi-item scale in the field of marketing (Bassi, 2011). Traditional methods and statistical tools widely used to assess measurement scale properties do not reflect the real nature of the variables involved; consequently they might produce misleading results. For example, in a previous work (Bassi, 2011), considering a scale for measuring customer satisfaction with reference to an experiential good, a film seen at the cinema, the results obtained using latent class analysis showed that traditional protocols were not robust enough. Considering these evidences, we want to study what happens when evaluating a scale for measuring customer satisfaction with reference to a different kind of good, such as a shopping one, like a pair of branded jeans.

Latent class models were introduced by Lazarsfeld and Henry (1968) to express latent attitudinal variables from dichotomous survey items, then they were extended to nominal variables by Goodman (1974a, 1974b), who also developed the maximum likelihood algorithm for estimating latent class models that serves as the basis for many software with this purpose. Later, these models were further extended to include observable variables of mixed scale type, like ordinal, continuous and counts.

Latent class models described in this paper are the latent class cluster model, the latent class factor model and the latent class regression model.

A traditional latent class cluster model, with one latent variable and four nominal indicators, for example, can be expressed with the following equation (1):

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}, \qquad (1)$$

where π_{ijklt}^{ABCDX} is the proportion of units in the five-way contingency table; π_t^X is the probability of being in latent class $t = 1, \dots, T$ of variable X; $\pi_{it}^{A|X}$ is the conditional probability of obtaining the ith, $i = 1,2, \dots, I$, response to item A from members of latent class t; $\pi_{jt}^{B|X}, \pi_{kt}^{C|X}, \pi_{lt}^{D|X}$, $j = 1,2, \dots, J$, $k = 1,2, \dots, K$, $l = 1,2, \dots, L$, are the conditional probabilities of item B, C, D respectively. An important assumption is that of local independence, that is, given a latent class, the indicators are independent from one another.

Haberman (1979) demonstrated that the model just described is equivalent to a hierarchical log-linear model with the following form (2):

$$\ln F_{ijklt}^{ABCDX} = \lambda + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{AX} + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX}, \qquad (2)$$

where F_{ijklt}^{ABCDX} is the absolute frequency in the generic cell of the five-way contingency table; $\lambda_t^X, \lambda_i^A, \lambda_j^B, \lambda_k^C$ and λ_l^D are the first-order effects and $\lambda_{it}^{AX}, \lambda_{jt}^{BX}, \lambda_{kt}^{CX}$ and λ_{lt}^{DX} are the second-order or interaction effects. The link between the parameters of these two representations of the same model can be expressed as follows (Haberman, 1979; Heinen, 1993):

$$\pi_{it}^{A|X} = \frac{\exp(\eta_{i|t}^{A})}{\sum_{i'=1}^{I} \exp(\eta_{i'|t}^{A})}, \qquad (3)$$

with

$$\eta_{i|t}^{A} = \lambda_i^A + \lambda_{it}^{AX}.$$

The same holds for the other indicators B, C and D. If the observed variables are nominal there is no need for further restrictions except for dummy or effect coding constraints in order to let the parameters be identifiable. On the contrary, if the observed variables are ordinal this aspect is taken into account restricting the two-variable log-linear parameters appearing in the logistic form of $\pi_{it}^{A|X}$ using the category scores y_i^A, that is the score y assigned to the ith response to item A, in the following way $\lambda_{it}^{AX} = \lambda_t^X y_i^A$.

Rejection of a traditional T-class latent class cluster model because it doesn't fit well, means that the local independence assumption does not hold with T classes. In such cases, a model with $T + 1$ classes is fitted to the data; however different model-fitting strategies may be adopted in order to obtain a model that fits better, for example increasing the number of latent variables rather than latent classes. This leads to an important extension of traditional latent class cluster model that is the latent class factor model (Magidson & Vermunt, 2001). Traditional latent class cluster models containing four or more classes can be interpreted in terms of two or more component latent variables by treating those components as a joint variable. For example a latent variable X consisting of $T = 4$ classes can be re-expressed in terms of two dichotomous latent variables $V = \{1,2\}$, $W = \{1,2\}$ using the following correspondences: $X = 1$ corresponds with $V = 1$ and $W = 1$; $X = 2$ with $V = 1$ and $W = 2$; $X = 3$ with $V = 2$ and $W = 1$; $X = 4$ with $V = 2$ and $W = 2$. Formally, for four nominal variables, the four-class latent class cluster model can be reparameterized as an unrestricted latent class factor model with two dichotomous latent variables as follows (4):

$$\pi_{ijklrs}^{ABCDVW} = \pi_{rs}^{VW}\pi_{ijklrs}^{ABCD|VW} = \pi_{rs}^{VW}\pi_{irs}^{A|VW}\pi_{jrs}^{B|VW}\pi_{krs}^{C|VW}\pi_{lrs}^{D|VW}. \tag{4}$$

Again, there is an equivalent hierarchical log-linear representation of this model, which is (5);

$$\begin{aligned}\ln F_{ijklrs}^{ABCDVW} &= \lambda + \lambda_r^V + \lambda_s^W + \lambda_{rs}^{VW} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ir}^{AV} + \lambda_{jr}^{BV} + \lambda_{kr}^{CV} + \lambda_{lr}^{DV} + \lambda_{is}^{AW} + \lambda_{js}^{BW} \\ &\quad + \lambda_{ks}^{CW} + \lambda_{ls}^{DW} + \lambda_{irs}^{AVW} + \lambda_{jrs}^{BVW} + \lambda_{krs}^{CVW} + \lambda_{lrs}^{DVW}. \end{aligned} \tag{5}$$

When the discrete factors are dichotomous, independent from one another and three-variable terms are set to zero, the latent class factor model becomes a basic latent class factor model (Magidson and Vermunt, 2001) and the equation (5) becomes:

$$\begin{aligned}\ln F_{ijklrs}^{ABCDVW} &= \lambda + \lambda_r^V + \lambda_s^W + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ir}^{AV} + \lambda_{jr}^{BV} + \lambda_{kr}^{CV} + \lambda_{lr}^{DV} + \lambda_{is}^{AW} + \lambda_{js}^{BW} + \lambda_{ks}^{CW} \\ &\quad + \lambda_{ls}^{DW}. \end{aligned} \tag{6}$$

The main advantage of this basic latent class factor model is a consequence of the following result: it turns out that the number of distinct parameters of a basic latent class factor model including R factors is the same as an LC cluster model with just $R + 1$ classes; so it allows a specification of a 2^R-class model with the same number of parameters as a traditional latent class cluster model with only $R + 1$ classes. This offers a great advantage in parsimony over traditional latent class cluster models and let the parameters be identifiable even when traditional latent class cluster model parameters are not.

To take into account the fact that the latent factors are dichotomous or ordinal, conditional response probabilities, for example $\pi_{irs}^{A|VW}$, are restricted by means of logit models with linear terms:

$$\eta_{i|r,s}^{A} = \lambda_i^A + \lambda_{iV}^A x_r^V + \lambda_{iW}^A x_s^W. \tag{7}$$

As it can be seen , the two-variable terms λ_{ir}^{AV} and λ_{is}^{AW} are restricted using the category scores x_r^V, x_s^W, that is the scores x assigned to the rth and the sth category of factor V and W, respectively.

Another kind of non-traditional latent class model is the latent class regression model (see, for example, Agresti, 2002; Vermunt & Van Dijk, 2001; Wedel & DeSarbo 1994; Wedel & Kamakura, 1998).

The main difference between this model and the other two described above is that the latent class regression model has just one dependent variable which may be measured repeatedly on a single unit. Another difference regards the distinction between the two types of exogenous variables which may be included in the model. The ones affecting the latent variable, called covariates, and the ones affecting the dependent variable, called predictors. This model differs from traditional regression models because it allows for different causal relationships between observed variables among latent classes, since it considers explicitly the presence of a latent variable interacting with the dependent one.

The most general probability structure of a latent class regression model takes on the form (8):

$$f\left(\mathbf{y}_i \middle| \mathbf{z}_i^{cov}, \mathbf{z}_i^{pred}\right) = \sum_{x=1}^{K} P\left(x \middle| \mathbf{z}_i^{cov}\right) \prod_{t=1}^{T_i} f\left(y_{it} \middle| x, \mathbf{z}_{it}^{pred}\right), \tag{8}$$

where y_{it} is the value of the dependent variable observed on unit i at occasion t; T_i is the number of observations on unit i; $\mathbf{z}_i^{cov}$ is a vector of covariates and $\mathbf{z}_i^{pred}$ is a vector of predictors.
A special case of this model is when we have just one replication for each case and there are no covariates interacting with the latent variable. Such a model is the one used here for studying construct validity and it is described by the equation (9):

$$f\left(y_i \middle| \mathbf{z}_i^{pred}\right) = \sum_{x=1}^{K} P(x) f\left(y_i \middle| x, \mathbf{z}_i^{pred}\right). \tag{9}$$

5. Scale evaluation

The purpose of this paper is discussing the results obtained following traditional protocols for developing and validating a multi-item measurement scale with reference to a shopping good that is a pair of branded jeans, taking into account that traditional statistical tools employed are suited for metric variables and may not be adequate when items generate ordinal variables. Moreover, they don't consider explicitly the unobservable nature of the latent variable that is customer satisfaction. Consequently, a different approach based on latent class analysis may improve scale evaluation since it considers both these aspects, and lead to different outcomes revealing that traditional methods might not be adequate enough to carry out this kind of analyses. This is what happens when considering a multi-item measurement scale with reference to an experiential good, a film seen at the cinema (Bassi, 2011). Here, we want to show that such result occurs even for a shopping good, that is a good characterized by a stronger involvement than the ones belonging to the experiential category, and consequently leads to a different type of consumption experience.
The aspects considered in this paper in order to evaluate the scale adopted are internal consistency along with scale dimensionality and criterion and construct validity. All these features are important scale properties and are assessed here using latent class models. In particular, latent class factor models are used in order to evaluate scale dimensionality (if a scale is multidimensional internal consistency should be assessed for each of construct dimensions, Churchill, 1979); latent class cluster models are employed to evaluate criterion validity; finally, latent class regression models are involved for studying construct validity[1].

5.1. Scale dimensionality

The first feature studied with the support of latent class analysis is scale dimensionality. In order to determine the number of dimensions underlying the construct to be measured, several latent class factor models were estimated including an increasing number of factors. Looking at Table 2, according to the p-values associated with the L^2-statistic, indicating the amount of association between observed variables which remains unexplained after estimating the model, the two-factor and three-factor models were selected. Besides this, L^2 value decreases significantly when the number of latent factors changes from two to three and even the BIC index leads us to conclude that

[1] All results presented were obtained with the software Latent Gold 5.0 (Vermunt & Magidson, 2013)

the model with three latent factors is the one that fits better, because it takes on the lowest value among the models which show an adequate fit ($p > 0.05$).

	LL	$\mathrm{BIC_{LL}}$	N. of par.	L^2	p-value
1 factor	-10,874.405	22,644.304	157	18,362.585	0.098
2 factors	-10,605.576	22,243.537	181	17,824.926	0.120
3 factors	-10,421.964	22,013.204	205	17,457.703	0.106
4 factors	-10,270.417	21,847.001	229	17,154.609	0.012
5 factors	-10,162.079	21,767.214	253	16,937.932	0.030

Table 2 – Log-likelihood (LL), BIC index, number of parameters, L^2-statistic and p-value for each of the estimated latent class factor models

Looking at the factor loadings in Table 3 and taking into account the content of each item, the first factor is linked to items E1 and R3 and can be interpreted as the capability of advertising to involve customers and catch their attention; the second one, linked to items E2, R2, V2, V3, V4, U1, U4 and P1, refers to wearability and image communicated through the product itself; finally the third factor that is linked to items R1, R4, R5, R6, V1, U2, U3, U5, P2, P3, P4, P5 and P6, represents the quality of the good even in relation with its price.

			Latent factors		
Stage	**Item description**	**Item**	Factor 1	Factor 2	Factor 3
Need recognition	Involvement	E1	**-0.607**	0.360	-0.151
	Product's style	E2	-0.052	**0.671**	-0.197
Information search	Manufacturing	R1	0.028	0.100	**-0.325**
	Color/shape	R2	0.003	**0.349**	-0.265
	Catch attention	R3	**-0.620**	0.268	-0.167
	Personnel's competence	R4	0.007	0.047	**-0.508**
	Clearness	R5	0.049	-0.035	**-0.380**
	Quality	R6	0.159	0.277	**-0.447**
Evaluation of alternatives	Perceived quality	V1	0.126	0.271	**-0.549**
	Wanted features	V2	0.099	**0.567**	-0.301
	Notoriety	V3	-0.108	**0.358**	-0.329
	Wearability	V4	0.209	**0.509**	-0.332
Purchase decision	Outlet's features	U1	-0.044	**0.394**	-0.338
	Personnel's being willing	U2	0.132	0.093	**-0.469**
	Price/quality ratio	U3	0.277	0.162	**-0.455**
	Image communicated	U4	-0.029	**0.548**	-0.397
	Price paid	U5	0.096	0.335	**-0.416**
Post-purchase evaluation	Overall performance	P1	0.280	**0.500**	-0.399
	Confirmed information	P2	0.298	0.383	**-0.423**
	Reliability	P3	0.378	0.307	**-0.479**
	Keep color/shape	P4	0.334	0.203	**-0.444**
	Convenience	P5	0.315	0.226	**-0.410**
	Quality certification	P6	0.121	0.175	**-0.606**

Table 3 – Factor loadings for the estimated latent class factor model including three factors

This outcome is quite different compared to the one obtained previously following traditional protocols. Indeed, traditional factor analysis which is suited for metric variables suggested the presence of one single factor underlying the construct to be measured and the scale seemed to be unidimensional. On the contrary, using a different approach based on latent class analysis it is obtained that customer satisfaction is multidimensional. This evidence suggests scale reliability should be assessed for each one of these three dimensions in order to avoid misleading results. Cronbach's alpha coefficients calculated separately for each dimensions took on the values 0.800, 0.840, 0.855, respectively, so it can be concluded the scale has the property of being reliable.

5.2. Criterion validity

A different approach than the traditional one based on statistical tools like correlation coefficients and analysis of variance, both suited for metric variables, was followed to assess criterion validity. Again, the new approach is based on latent class analysis. Taking into account the ordinal nature of the observed variables, several latent class cluster models were estimated for characterizing the latent variable, which was then related to item S1, the criterion variable which measures customer satisfaction with reference to the entire consumption experience. This approach lets us to consider explicitly that customer satisfaction is not directly observable.

	LL	BIC_{LL}	N. of par.	L^2	p-value
1 class	-11,493.602	23,745.807	133	19,600.978	0.042
2 classes	-10,874.405	22,644.304	157	18,362.585	0.088
3 classes	-10,676.984	22,386.352	181	17,967.741	0.114
4 classes	-10,545.795	22,260.866	205	17,705.365	0.132
5 classes	-10,437.151	22,180.467	229	17,488.075	0.082
6 classes	-10,372.127	22,187.311	253	17,358.028	0.060
7 classes	-10,319.534	22,219.016	277	17,252.842	0.054

Table 4 – Log-likelihood (LL), BIC index, number of parameters, L^2-statistic and p-value for each of the estimated latent class cluster models

As above said, a set of latent class cluster models with an increasing number of classes, representing customers with different levels of satisfaction, were estimated. Looking at Table 4, according to the p-values associated with the L^2-statistic, two models fit better than the others, the model with three and the model with four latent classes. The amount of association between the observed variables remaining unexplained after estimating the model decreases significantly when the number of classes changes from three to four, so the latter model fits best to the data. Even the BIC value leads us to conclude that the four class model is the one with the best fit.

Consequently, the latent variable can be described by four different classes of customers with different satisfaction levels, each one of these classes is large enough to be considered relevant for the purpose of the analyses and the profile of customers who belong to them is quite different. In particular, see Table 5, the largest class is composed of 44.4% of the sample and individuals belonging to it have a medium level of satisfaction (486). There is a class which includes just 8.2% of the sample with an average satisfaction level equal to 4.76. These customers are particularly unsatisfied with the capability of advertising to involve them and to catch their attention, but at same time, they are really satisfied about the quality of the good, even in relation with its price. These peaks are absent when we consider customers belonging to the largest class, thus these first two clusters are quite different. In addition, we have two classes with opposite satisfaction levels since the first, of size equal to 24.6%, is described by the highest satisfaction level (5.57) and the latter, composed of 22.8% of the sample, includes respondents characterized by the lowest level of satisfaction, equal to 3.69. Another interesting result is that all items contribute in a significant way towards the ability to discriminate between clusters, since the p-values associated with the Wald statistic, used for testing the null hypothesis stating that all the effects associated with each indicator equal to zero, are always less than 1%.

	Classes			
	Class 1	Class 2	Class 3	Class 4
Class size	44.4%	24.6%	22.8%	8.2%
Item				
E1	3.82	3.98	2.88	1.37
E2	5.13	6.39	3.82	4.52
R1	3.58	4.03	2.77	4.12
R2	4.42	5.11	3.49	4.44
R3	4.61	4.54	3.41	1.35
R4	4.59	4.97	3.50	3.83
R5	4.51	4.71	3.64	4.45
R6	4.96	5.75	3.68	5.10
V1	5.08	5.92	3.72	5.15
V2	5.09	6.31	4.08	5.42
V3	5.38	5.99	4.11	4.05
V4	5.57	6.66	4.30	6.25
U1	5.04	6.00	3.84	4.39
U2	4.80	5.46	3.68	4.64
U3	5.02	5.46	3.59	6.32
U4	5.23	6.29	3.66	4.16
U5	4.83	5.67	3.73	4.87
P1	5.41	6.39	4.10	6.04
P2	4.92	5.86	3.83	5.80
P3	5.22	6.10	4.01	6.36
P4	5.34	5.82	3.98	6.33
P5	4.80	5.31	3.54	6.07
P6	4.55	5.45	3.41	4.51
Overall mean	4.86	5.57	3.69	4.76

Table 5 – Class sizes and conditional means of each indicator for the estimated latent class cluster model including four classes

The latent variable just described was then studied in relation with the criterion variable by means of the Pearson Chi-squared test and the Goodman and Kruskal Gamma index. Both these tools are suited for ordinal variables and show a significant association between them, the latent variable and criterion variable (item S1). On one hand, the Pearson Chi-squared test statistic is equal to 181,585 with an associated p-value which takes on a value lower than 0.001; on the other hand, Goodman and Kruskal Gamma is equal to 0.665, confirming in both cases criterion validity property for our scale.

5.3. Construct validity

The last feature taken into account here in order to study measurement scale's properties is construct validity. For improving measurement scale evaluation a different procedure based on latent class regression model was adopted. The main difference between this kind of regression model and traditional ones, is mainly that the first allows for different causal relationships between observed variables among latent classes. The purpose of the analyses is to study if there are any differences in causal relationships between the total scale score and the control variables generated by the additional items, C1, C2 and C3, given a specific latent class. Moreover, latent class regression models let us consider the ordinal nature of the dependent variables generated by the additional items, thus, as a consequence, latent class analysis is still more adequate for studying this sort of relationships compared to traditional correlation coefficients and analysis of variance.

To achieve this goal, a set of latent class regression models were estimated including an increasing number of latent classes representing customers with different satisfaction levels. According to the BIC index, for all three additional items C1, C2 and C3, models with just one latent class show the best fit, this means that the causal relationship between the total scale score and each one of these three variables is the same for the whole sample. Furthermore these relationships are positive and statistically significant, as it can be seen in Table 6 which reports regression coefficients and associated z-values.

	Regression coefficient	z-value
C1	0.039	8.657
C2	0.053	9.240
C3	0.038	8.584

Table 6 – Regression coefficients describing the causal relationship between the total scale score and the control variables along with z-values

This procedure replaces the traditional approach based on statistical instruments such as correlation coefficients and analysis of variance which are suited for metric variables and not take explicitly into account the fact that the latent variable is unobservable.

Another proof of association between control variables and customer satisfaction was obtained computing Goodman and Kruskal Gamma cograduation coefficients. They were equal to 0.603, 0.695 e 0.522 for items C1, C2 and C3, respectively; thus, there is cograduation between the latent variable defined previously with the support of traditional latent class cluster model and each one of these items. Gathering these outcomes, it looks like that even construct validity is confirmed.

6. Conclusions

The aim of this paper is showing how latent class analysis can improve multi-item measurement scale evaluation when we consider a scale for measuring customer satisfaction with reference to a shopping good, a kind of good characterized by a strong involvement and an emotional learning especially due to the mid/high-level price and the fact these goods are linked with the lifestyle of the customer. Such an evidence arose in a previous work about a scale with reference to an experiential good, a film seen at the cinema, characterized by a weaker degree of involvement than the shopping good considered here, determining a different kind of consumption experience. However, this occurs even when considering a pair of branded jeans, a product that belongs to the shopping category, as we do in this paper. The assumptions that latent class analysis makes reflect more accurately the nature of the observed variables taking into account the fact they are ordinal and let us consider explicitly that the construct to be measured, that is customer satisfaction, is a latent variable which is not directly observable. These are the main differences between the latent class approach and procedures defined within traditional protocols, based on statistical tools better suited for metric variables which do not often consider explicitly that customer satisfaction is a construct not directly observable. As a consequence, latent class analysis is more adequate for scale evaluation and development and sometimes leads to different conclusions compared with outcomes of traditional analyses.

The data used here were obtained administering a scale for measuring customer satisfaction with reference to a branded pair of jeans to a sample of 300 customers. The scale considers all phases of which consumption experience is made up.

Within the new approach based on latent class analysis, latent class factor models were used for studying scale dimensionality, latent class cluster models for assessing criterion validity and latent class regressions model in order to evaluate construct validity.

The outcomes of the analyses, as above mentioned, do not always confirm what was obtained following traditional protocols. In particular, a scale judged unidimensional was multidimensional instead, thus reliability issue should be assessed for each dimension separately. Furthermore, this new approach provided additional information about traits like customers' profile and relationships between customer satisfaction and other variables theoretically liked with it, like repurchase intention, positive word of mouth and absence of complaints. In any way the scale was judged valid and reliable even adopting a latent class approach. Concluding, the new procedure based on latent class analysis' disclosed its usefulness and potential for evaluating and developing multi-item measurement scales, suggesting its application in this field even when considering a shopping good.

Appendix: Final questionnaire

Questionnaire:
A scale to measure customer satisfaction with reference to shopping goods

Questionnaire N° ______________________

Date ________________________________

Screening question:

Think of a purchase experience with reference to a pair of branded jeans with a strong advertising campaign.

WARNING: If the respondent does not have this kind of experience, thank him/her and close the interview. Otherwise proceed with it.

First of all we want to thank you for your kind cooperation.
We are carrying on a research about customer satisfaction with reference to consumption experiences.

We inform you that all your answers will be completely anonymous and data collected about your personal information will be just used for statistical purposes.

We ask you to answer *all questions* honestly (don't omit any question) as we consider your opinions really important.

A SCALE TO MEASURE CUSTOMER SATISFACTION WITH REFERENCE TO SHOPPING GOODS

Code	How much am I satisfied about…	Completely NOT satisfied … Completely satisfied
E1	… the way the strong advertising campaign involved me?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
E2	… product's style: the degree of adherence to new fashions and trends?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
R1	… information search through business sources regarding product's manufacturing?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
R2	… information search through business sources regarding product's aesthetic features (color and shape)?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
R3	… strong advertising campaign's capability to catch my attention?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐

Code	How much am I satisfied about…	Completely NOT satisfied … Completely satisfied
R4	… outlet's personnel's competence in describing product's features?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
R5	… clearness of information included on the label?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
R6	… information I gathered with reference to brand image in terms of quality?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
V1	… product's perceived quality compared with that of the alternatives available on the market?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
V2	… the presence of wanted features in the product compared with the alternatives available on the market?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
V3	… the notoriety of the chosen branded jeans compared with that of the other branded jeans available on the market?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
V4	… the wearability of the chosen branded jeans compared with that of the other branded jeans available on the market?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
U1	… the outlet's being modern and comfortable?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
U2	… the outlet's personnel's being willing?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
U3	… product's price/quality ratio?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
U4	… the image communicated through the jeans of the chosen brand?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
U5	… the price paid with relation to the offer (that is not only considering the product itself but also the warranty, brand image and so on)?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P1	… the overall performance (wearability) of the chosen branded jeans I actually perceived in their using?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P2	… the degree to which collected information regarding the chosen branded jeans were confirmed?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P3	… product reliability I actually perceived in its using?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P4	… product's capability to keep its features like color, shape and dimensions as they are?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P5	… product's convenience?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐
P6	… quality certification's validity provided by manufacturer?	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐

Ultimately:

Code	How much am I satisfied about…	Completely NOT satisfied						Completely satisfied
S1	… my own entire consumption experience with relation to the jeans of the chosen brand?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Code	**How much do I agree with the following statements?**	**I fully DON'T agree**						**I fully agree**
C1	I'm going to purchase the product again	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
C2	I will speak well about the consumption experience I had with the product	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
C3	I do not have any complaints about any aspects of the consumption experience I had with the product	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

For classifying previous data, finally answer the following questions:

Code	Personal information		
SEX	Sex	1 ☐	Male
		2 ☐	Female
AGE	Age	*Fill in your age (in years): ____________*	

The questionnaire is over. Thank you for your kind cooperation again.

References

Agresti, A. (2002). *Categorical Data Analysis*. Second Edition, New York: Wiley.

Bagozzi, R., & Ruvio, A. (2011). *Encyclopedia of Marketing, Vol. 3, Consumer Behavior*. New York: Wiley.

Bassi, F. (2011). Latent class analysis for marketing scale development. *International Journal of Market Research*, 53, 2, 211-232.

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1991). Consumer decision making. *Handbook of Consumer Behavior*. Robertson, T. S., & Kassarjian, H. H., Englewood Cliffs, N. J.: Prentice-Hall, 50-84.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16, 1, pp. 64-73.

De Vellis, R. F. (1991). *Scale Development. Theory and Applications.* Newbury Park: Sage.

Ferber, R., & Holton, R. H. (1958). The distinction between convenience goods, shopping goods, and specialty goods. *Journal of Marketing*, 23, 1, 53-56.

Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I. A modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.

Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

Guido, G., Bassi, F., & Peluso, A. M. (2010). *La soddisfazione del consumatore. La misura della customer satisfaction nelle esperienze di consumo*. Franco Angeli, Milano, 2010.

Haberman, S. J. (1979). *Analysis of qualitative data, Vol. 2, New developments*. New York: Academic Press.

Heinen, A. (1993). *Discrete latent variables models*. Tilburg: Tilburg University Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. Sociological Methodology, 31, 223-264.

McCutcheon, A. L. (1987). *Latent class analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-064. Newbury Park, CA: Sage.

Tyagi, C. L., & Kumar, A. (2004). *Consumer Behaviour*. Atlantic Publishers and Distributors, New Delhi, 2004.

Vermunt, J. K., & Magidson, J. (2013). Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. R. P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, 352-388, Cambridge: Blackwell Publishers.

Wedel, M., & Kamakura, W. A. (1998). Market Segmentation: Concepts and Methodological Foundations. Boston: Kluwer Academic Publishers.

Wilkie, W. L. (1990). *Consumer Behavior*. Wiley, New York, 1990.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research*, 12, 3, 341-352.

Acknowledgements

Research for this paper was supported by grant CPDA121180/12 financed by the University of Padova for the project with title "Statistical and econometric approach to marketing: applications and developments to customer satisfaction and market segmentation".

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order copies of the working papers from by emailing to wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy