

METHODS FOR TESTING DISCRIMINANT VALIDITY

Professor PhD Adriana ZAIȚ

University "A. I. Cuza", Iași, Romania

Email: azait@uaic.ro

PhD Student Patricea Elena BERTEA

University "A. I. Cuza", Iași, Romania

Email: patricia.bertea@yahoo.com

Abstract:

The study presents three methods which can be used to assess discriminant validity for multi-item scales. Q-sorting is presented as a method that can be used in early stages of research, being more exploratory, while the chi-square difference test and the average variance extracted analysis are recommended for the confirmatory stages of research. The paper describes briefly the three methods and presents evidence from two surveys that aimed to develop a scale for measuring perceived risk in e-commerce.

Keywords: validity, discriminant validity, Q-sorting, confirmatory factorial analysis

Introduction

Scale development represents an important area of research in Marketing. Since we deal with latent variables which are not observable we have to create instruments in order to measure them. Variables such as personality or perceived risk are measured through multi-item scales. When developing such a scale researchers generally look at one important criteria which is the level of reliability given by alpha Cronbach values. A value of more than 0,7 for alpha Cronbach is considered acceptable (Nunnally, 1967). Scale reliability is influenced by several factors of research design (Bertea, 2010), this is why it is important to apply other methods in order to be sure that the instrument presents reliability as well as validity.

Construct validity refers more to the measurement of the variable. The issue is that the items chosen to build up a construct interact in such manner that allows the researcher to capture the essence of the latent variable that has to be measured.

It is important to make the distinction between internal validity and construct validity. The first one refers to assuring a methodology that enables the research to rule out alternative explanations for the dependent variables, while construct validity is more concerned with the choice of the instrument and its ability to capture the latent variable. Internal validity becomes a problem in experimental studies, where each experimental group has to follow the same methodology in order to be able to correctly isolate the effect. Construct validity has three components: convergent, discriminant and nomological validity. Discriminant validity assumes that items should correlate higher among them than they correlate with other items from other constructs that are theoretically supposed not to correlate.

Testing for discriminant validity can be done using one of the following methods: O-sorting, chi-square difference test and the average variance extracted analysis.

Q-sorting

The Q-sorting procedure aims to separate items in a multi-dimensional construct according to their specific domain. There are two ways that it can be done (Storey, et al., 1997):

- *Exploratory*, when respondents are given the items and asked to group and identify category labels for each group of items.

- *Confirmatory*, when the categories are already labeled and respondents are asked to classify each item in one category.

Q-sorting is applied on experts and other persons of interest for the research. It helps eliminate items that do not discriminate well between categories of items.

For the confirmatory procedure, the analysis can be made by calculating a percent for correct classification of each item from a construct. When this percent has low values, that means we have items with problems that do not discriminate well in relation with other items that form a different construct.

Chi-square difference test

Another method that can be used to assess discriminant validity is to do a chi-square difference test (Segars, 1997) that allows the researcher to compare two models, one in which the constructs are correlated and one in which they are not. When the test is significant the constructs present discriminant validity. In order to do that the constructs are analyzed using

Confirmatory Factor Analysis (CFA) which is commonly used for validity issues. The measurement models should be reflective and should be introduced in analysis in pairs of two. So, we will compare each time two constructs that we suspect to have problems with items discriminating among them. The first model analyzed through CFA will be a model where the two constructs are not correlated, while the second will be the one where we will allow for correlation. Each model will present a value for Chi-square (χ^2) and degrees of freedom (df). After doing the difference between the values of the two models we can see if the test is significant or not.

Average variance extracted analysis

In order to establish discriminant validity there is need for an appropriate AVE (Average Variance Extracted) analysis. In an AVE analysis, we test to see if the square root of every AVE value belonging to each latent construct is much larger than any correlation among any pair of latent constructs. AVE measures the explained variance of the construct. When comparing AVE with the correlation coefficient we actually want to see if the items of the construct explain more variance than do the items of the other constructs.

AVE, which is a test of discriminant validity, is calculated as:

$$AVE = \frac{\sum[\lambda_i^2]}{\sum[\lambda_i^2] + \sum[\text{Var}(\varepsilon_i)]},$$

where λ_i is the loading of each measurement item on its corresponding construct and ε_i is the error measurement.

The rule says that the square root of the AVE of each construct should be much larger than the correlation of the specific construct with any of the other

constructs. The value of AVE for each construct should be at least 0.50 (Fornell and Larcker, 1981).

Perceived risk in e-commerce

Marketing literature talks about perceived risk as a multi-dimensional construct. That means each dimension represents a construct in itself and is measured through multiple-items. In traditional commerce there were defined six dimensions of perceived risk: financial, physical, functional or performance, psychological, social (Jacoby & Kaplan, 1972) and a time risk or convenience risk (Roselius, 1971). However, in the context of e-commerce it is important to notice that there are changes as far as the risk concepts are concerned. Crespo et al. (2009) study the multi-dimensional perceived risk in relation with a certain product bought through the Internet. Nevertheless, it

makes sense to analyze perceived risk also in relation with the shopping channel. Featherman and Pavlou (2003) talk about financial, social, psychological, time, privacy risk and performance risk. They refer to the risk of the shopping channel, not of the product. The authors argue that adopting e-services involves a much higher risk than e-commerce adoption as users are to engage in a long term relationship. In analyzing the influence of perceived risk on e-services adoption intention, Featherman and Pavlou (2003) used the basic TAM (Technology Acceptance Model, Davis, 1989) and the multi-dimensional perceived risk approach (fig. 1). Each dimension of perceived risk was measured through multiple-items.

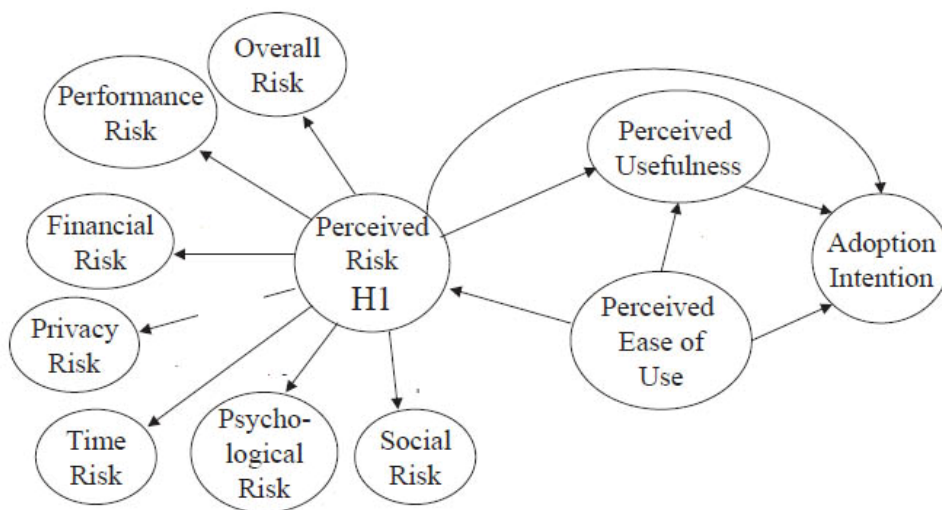


Figure 1. Featherman and Pavlou (2003) research model

Source: Featherman, M. S. & Pavlou, P. A. (2003), 'Predicting e-services adoption: a perceived risk facets perspective', *International Journal of Human-Computer Studies* 59(4), p.457

The use of multi-item scales for measuring perceived risk was recommended by Mitchell (1999) who considered that it is a better way to go inside the consumer's mind and to find out what really defines his behavior.

Research Methodology

The present study has used perceived risk as a multi-dimensional construct, having six dimensions redefined in the context of e-commerce (table 1).

Table 1

Perceived risk dimensions		
Type of risk	Definition	Number of items
Financial	The risk of losing money when buying online.	5
Product	The risk of getting a product that it is not what presented on the website.	6
Delivery	The risk of having a delayed delivery.	4
Security	The risk that the personal data is stolen and used for identity theft.	3
Social	The risk that the social group does not agree with e-commerce.	4
Psychological	The risk of feeling anxiety when shopping online.	4

For the Q-sorting study we developed a questionnaire were we included all items measuring perceived risk without showing which item belongs to which type of perceived risk.

Respondents had to classify items into 6 categories: social, psychological, financial, security, product and delivery risk (table 2).

Table 2

Q-sorting questionnaire example

Risk Item	Chek the risk type appropriate for the item	
Online shopping gives me a state of stress because it does not fit with my self-image.	Social	
	Financial	
	Psychological	√
	Security	
	Delivery	
	Product	

The questionnaire was applied on a sample formed of 23 students. The small sample was due to the fact that the research was in exploratory stage. As a quantitative indicator of the Q-sorting procedure we used the correct classification percent, which describes the percent of respondents that have correctly classified an item (Straub, et al., 2004).

For the Chi-square difference test and the AVE analysis we applied a questionnaire that aimed to measure perceived risk in e-commerce on a sample of 481 students. The larger sample was necessary since the research stage was confirmatory. The questionnaire had items that used a 7 point Likert scale, items which were either taken from the literature (Featherman & Pavlou, 2003; Pires et

al., 2004; Forsythe et al., 2006; Crespo et al., 2009) either from in-depth interviews. The in-depth interview was used as a qualitative method in order to obtain information that is common to Romanian consumers, users of Internet. The motivation is that the other scales were developed on different populations which are characterized by different cultures and levels of economic development.

For the Chi-square difference test we performed a confirmatory factor analysis using AMOS 19, while for the AVE we did the correlation matrix for the types of risk and calculated the AVE values for each type of risk.

Results

Since this study aims to discuss methods for assessing discriminant validity we will present only results that were obtained by applying the three methods previously mentioned.

Q-sorting

In order to calculate the percent of correct classification, we identified the frequency of respondents that checked the correct category for each item. We had items that obtained a 100% correct classification – 3 items, items that had

percents higher than 70% – 22 items, but also items with lower percents -4 items. We considered items with a low classification percent those who were below 60% (table 3).

Taking into account that more than 80% of all 26 items were correctly classified, we can consider that the scale has a good level of discriminant validity. However, it is important to further analyze those items that were not correctly recognized as belonging to a certain category of risk.

Table 3

Q-sorting results (items with low classification)

Risk type	Item	Percent
Psychological	Online shopping does not fit my self-image.	0.52
Security/privacy	There is high chance that hackers take over my personal account from a e-shop.	0.59
Time/delivery	If I do my shopping online, There is a high risk that I receive a different product that the one I ordered.	0.52
	When I buy online I am sure that I will receive exactly the product I ordered.	0.22

Chi-square difference test

We will exemplify the chi-square difference test on two constructs that had items which were suspected to produce confusion among respondents. The two constructs are: product risk and delivery risk. In order to test for discriminant validity we followed Segars (1997) recommendations:

- Create a model in which the two constructs do not correlate and perform CFA (fig. 2)

- Create a model in which the two constructs correlate and perform CFA (fig. 2)

- Do the chi-square difference test and if the test is significant than discriminant validity exists.

We introduced the two models into AMOS and performed the analysis. We set correlation to 0 for the first model (left side of figure 2) and for the second model we allowed free correlation (right side of figure 2).

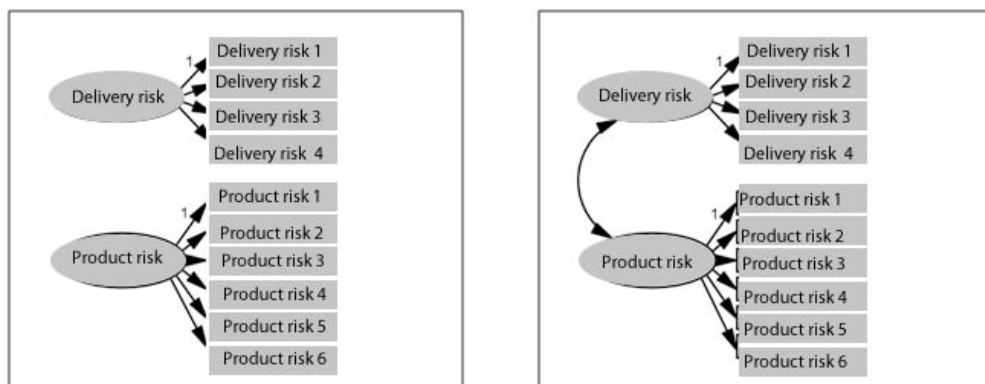


Figure 2 – Product risk versus delivery risk

Afterwards we calculated the chi-square difference test to see if it is significant or not (table 4).

Table 4

CFA results	
Model 1	Model 2
Chi-square = 400.58 Degrees of freedom =35 Probability level = 0.000	Chi-square = 133.632 Degrees of freedom =34 Probability level = 0.000
$\chi_1 - \chi_2 = 266.948$	
$df_1 - df_2 = 1$	

The difference test result was significant ($p=0 < 0,05$) which means that the two constructs present discriminant validity.

AVE analysis

As said before, the AVE values calculated according to the formula presented must be compared with the

correlation coefficients of each construct with the other constructs. So, first of all it is necessary to obtain a matrix where we can see the correlation of each type of risk with the other types. Afterwards on the diagonal we insert the AVE value in order to compare it with the other correlation coefficient (table 5).

Table 5

AVE analysis						
	Delivery risk	Financial risk	Product risk	Psychological risk	Social risk	Security risk
Delivery risk	0,707					
Financial risk	,539**	0,707				
Product risk	,652**	,551**	0,707			
Psychological risk	,357**	,325**	,559**	0,707		
Social risk	,159**	,201**	,257**	,517**	0,707	
Security risk	,541**	,616**	,507**	,243**	,133**	0,707

Table 5 shows results of the AVE analysis. It can easily be seen that the AVE values are above 0.5 and, moreover, are above the correlation coefficients for each type of perceived risk.

Conclusions

Conclusions refer to the use of assessment methods for discriminant validity. Since the study aimed to present the methods and exemplify their use, conclusions will not regard whether perceived risk's dimensions proved or not discriminant validity. Thus, this section will be concerned more on recommendations regarding the methods.

Therefore, we strongly recommend the Q-sorting procedure should be used in early stages of research, especially in the exploratory stage because it can reveal those items that generate problems for discriminant validity. If a small percent of respondents can correctly classify an item to its rightful category, than the researcher should consider reformulating the item or rejecting it from the construct. Q-sorting

is an easy to apply procedure, however its weak point is that the sample used should be formed mainly by experts and sometimes experts are not available or their number is very low.

The other two procedures should be used in the confirmatory stage of research. It is not necessary to apply both methods in one research, because both of them are strong and give valid results. Nevertheless, it seems that the AVE analysis is more popular because it offers a more parsimonious procedure, having all constructs grouped in one matrix. The Chi-square difference test is time consuming since we have to take 2 constructs each time and in the case of perceived risk's dimensions we would have performed 15 tests for verifying discriminant validity of each construct.

In conclusion, the Q-sorting procedure should be use in the phase of exploratory research when developing a scale for measuring latent variables, while the AVE analysis and the Chi-square difference test must be used in the confirmatory stage.

REFERENCES

- Bertea, P. (2010), 'Scales For Measuring Perceived Risk In E-Commerce-Testing Influences On Reliability', *Management and Marketing Journal Craiova* 8(S1), S81--S92.
- Crespo, Á. H.; del Bosque, I. R. & de los Salmones Sánchez, M. M. G. (2009), "The Influence Of Perceived Risk On Internet Shopping Behavior: A Multidimensional Perspective", *Journal of Risk Research* 12(2), 259–277.
- Davis, F. (1989), 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS quarterly*, 319–340.
- Featherman, M. S. & Pavlou, P. A. (2003), 'Predicting e-services adoption: a perceived risk facets perspective', *International Journal of Human-Computer Studies* 59(4), 451 - 474.
- Forsythe, S.; Liu, C.; Shannon, D. & Gardner, L. C. (2006), 'Development Of A Scale To Measure The Perceived Benefits And Risks Of Online Shopping', *Journal Of Interactive Marketing* 20(2).

- Jacoby, J. and L. B. Kaplan (1972), "The Components Of Perceived Risk", in *Proceedings of the Third Annual Conference of the Association for Consumer Research*, 1972.
- Mitchell, V.-W. (1999), 'Consumer Perceived Risk: Conceptualisations And Models', *European Journal of Marketing* 33, 163-195(33).
- Nunnally, J. (1967), *Psychometric theory*, Tata McGraw-Hill.
- Pires, G.; Stanton, J. & Eckford, A. (2004), 'Influences on the Perceived Risk of Purchasing Online', *Journal of Consumer Behaviour* 4, 118-131(14).
- Roselius, T. (1971), "Consumer Rankings of Risk Reduction Methods", *The Journal of Marketing*, 35(1), 56-61.
- Segars, A. (1997), 'Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research', *Omega* 25(1), 107--121.
- Straub, D.; Boudreau, M. & Gefen, D. (2004), 'Validation guidelines for IS positivist research', *Communications of the Association for Information Systems* 13(24), 380--427.