THE IMPACT OF THE NEUTRAL POINT IN THE PAIRED COMPARISONS OF STATEMENTS MODEL

(Dedicated to the memory of Prof. Roberto Corradetti)

Roberto Furlan, Graham Turner

Kantar Health, Epsom, UK

Abstract. *In recent years, Paired Comparisons of Statements (PCS), useful for collecting and scaling preference measurements through a structured research questionnaire, has gained a significant increase in popularity. The researcher defines a set of items and assumes that there is an underlying subjective dimension, such as importance or preference. Survey respondents are repetitively shown pairs of the possible items and the ultimate aim is to measure the location or position of the items on that dimension. With growing popularity there is a clear need to better understand the potentialities and limitations of PCS. While some preliminary work has already been done, there are still several unexplored areas. In this study we investigate to what extent the inclusion of a neutral point in the scale impacts on the accuracy of results.*

Dedication: *This paper is dedicated to the memory of Prof. Roberto Corradetti, with deep sadness at the loss, not only of a very distinguished statistician, but also of a very close personal friend whose gentleness was matched only by his wisdom. His great dedication and enthusiasm to research and innovation, and his even greater dedication to young researchers and students will be deeply missed. Without Roberto's knowledge and love of statistics which he so generously shared, this work would not have been possible.*

Keywords*: Paired Comparisons of Statements, PCS, neutral point, comparison, hierarchical bayes.*

1. INTRODUCTION

Companies constantly seek to enhance customer satisfaction and retention by improving the overall quality of a product or service. To do so, managers must focus on enhancing particular attributes of the product or service, those with the potential greatest impact on customer satisfaction. However, identifying such key characteristics can be challenging and a key step is determining the value customers attach to the different features. The market researcher has several tools in the arsenal to assess such value. Among these, the most popular metrics are traditional approaches such as ratings, rankings, and constant sum. However, in the last decade trade-off approaches such as Maximum Difference Scaling (MDS) (Louviere and Woodworth, 1990; Finn and Louviere, 1992) and Paired Comparisons of Statements (PCS) (David, 1988; Corradetti and Furlan, 2006) have become rather popular among market researchers due to their advantages over the more traditional techniques. In the literature, one can find plenty of theoretical and empirical studies dedicated to these individual approaches and also some works involving a comparison of different methodologies. For instance, Chrzan and Golovashkina (2006) conducted a study to test six different types of importance metrics including traditional approaches (i.e., ratings and constant sum), MDS, and three other less popular methodologies; Jaeger et al., (2008) compared MDS to preference ratings; Madansky (2010) considered PCS, MDS, and preference rankings. All these studies were based on empirical results.

PCS is a discrete choice model that has its roots in the law of comparative judgment presented by Thurstone (1927) and that has been extensively described by David (1988) and more recently by Corradetti and Furlan (2006) and Furlan and Turner (2011). To date, it is widely used to collect and scale preference measurements through a structured research questionnaire. The researcher defines a set of items (usually statements, messages, product features, service characteristics, options in a decision, etc.) and assumes that there is an underlying subjective dimension, such as extent of preference, degree of importance, degree of credibility, extent of appeal, impact on prescription (for medical products), impact on purchasing, etc. In the PCS approach, the ultimate aim is to measure the location or position of the set of items on that dimension. These locations are estimated through an algorithm that provides a set of utilities, with one utility score associated to each item.

In a PCS exercise, survey respondents are repetitively shown subsets of size two of the possible items (each subset is also referred to as a PCS task). In its simplest setup, referred to as *short paired comparison of statements*, the respondent is asked to choose the preferred item (or the most credible, important, appealing, etc.) from each subset. As the resulting data are quite poor from both a psychological and a statistical points of view, the researcher often prefers to ask the respondent to also indicate the intensity of the preference in what is called *graded paired comparison of statements* model. In the graded version, the two items are usually presented horizontally and a scale is presented underneath. In both the short and graded version the researcher might decide to include a *neutral* (or *indifference*) point for indicating ‘no preference’, useful when one does not want to force respondents to make a choice towards one of the two items.

To some extent PCS is a valid and popular alternative to self-explicated models. In this class of models, respondents would directly rate or rank the elements or allocate a number of points among them. With a rating approach, survey respondents are presented the features individually and asked for their evaluations. While this exercise is straightforward and requires little time and effort, it does not explicitly capture priorities and results might suffer from lack of differentiation (e.g., everything emerges as being important); in addition, the scale suffers from scalar inequivalence issue (i.e., due to response style and cultural and personal background differences there might be differences across respondents in the usage of the scale - Louviere and Flynn, 2011; Sawtooth Software, 2007). All these drawbacks might compromise the correct interpretation of the results and thus the actionability. The ranking approach would not present these issues, however rank evaluations imply an ordinal scale, while some researchers prefer to work with interval or ratio scales because of their statistical properties. Similarly to the ranking approach, the popular constant-sum allocation, an approach requiring respondents to divide a limited amount of resources across a number of elements, captures priorities quite well and the scale is not affected by the inequivalence issue. However, with a large number of elements (e.g., ten or more), it becomes very difficult for the respondent to effectively allocate scores among all of them, thus limiting the applicability of this approach to only the smallest batteries of elements (Srinivasan and Wyner, 2009).

In this context, PCS represents a valid approach to collect preference measurements, as it is based on a trade-off approach rather than a self-explicated one. It is a rather simple exercise, usually requiring an acceptable effort from respondents, it is simple to execute, it can handle many elements, it provides results that are empirically consistent with more complex ordering tasks, and produces reasonable differentiation in the results which appear to be on a convenient ratio scale. Probably, the most important property is that it measures all the items on a common scale, thus addressing the scalar inequivalence problem characterizing the way respondents use rating scales, arising mostly from differences in response styles and cultural differences (Cohen and Neira, 2003; Steenkamp and Hofstede, 2002).

The simplest way to analyse PCS data is through a logit model (Corradetti and Furlan, 2006). Let P be the set of items in the experimental design and T the set of PCS tasks to be evaluated. Each task $t \in T$ is assessed through a preference score assigned to one of the two presented items. The evaluation for task t is stored onto $\mathbf{y}_t$, an interval-scaled variable that can assume values in the range $[-s, +s]$, where s is a positive integer set by the researcher. A negative value for $\mathbf{y}_t$ indicates a preference for the first/left item in the task while a positive one indicates a preference for the second/right item. The indifference for either of the two items is expressed by $\mathbf{y}_t = 0$ and it can be made available or not to respondents. While in the short paired comparison setup $s = 1$, in the graded paired comparison s usually ranges from 2 to 4. The larger the absolute value of $\mathbf{y}_t$, the stronger the preference for the associated item. The PCS logit model is specified by a generalized linear model with a logit link function: the stochastic component of the model is based on the preference $\mathbf{y}_t$ suitably recoded on a 0:1 scale, while the systematic component is based on a design $\mathbf{X}$ matrix with P columns describing the PCS tasks:

$$E\left(f(\mathbf{y}_t)\right) = \boldsymbol{\mu}, \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad logit(\boldsymbol{\mu}) = \boldsymbol{\eta} \tag{1}$$

The P-dimensional vector of parameters β s represents the items' utilities and they can be estimated by the maximum likelihood method. This approach is particularly indicated for the graded PCS model, as it can model the strength of competition within each set. The analysis is usually carried out for the full sample or for major groups of respondents. However, given a large enough number of tasks with respect to the number of items to be assessed, individual-level analysis can be carried out. McCullagh and Nelder (1989) have provided exhaustive information about the estimation algorithm and asymptotic properties of the parameter estimates.

Another popular approach to estimate individual-level PCS scores is hierarchical Bayes (HB) analysis. HB is particularly indicated to estimate PCS individual utilities given only a few tasks assessed by respondents. This is accomplished by borrowing information from population information describing the preferences of other respondents. HB models estimate preference coefficients for a given respondent based on his or her responses as well as on responses of similar respondents. Consequently, more information is used in estimating individual utilities, thus it is possible to estimate a larger number of parameters or the same number with greater precision than other approaches allow. HB estimates tend to be robust to mistakes or inappropriate answers due, for example, to tiredness. HB approach was first adopted for conjoint analysis where, as for PCS analysis, usually there are many heterogeneous units of analysis (the respondents) but for each unit only little information is available (tasks evaluations).

While until a decade ago researchers could only run basic analysis on PCS data allowing only aggregate-level logit estimation for studies investigating many items, nowadays software packages offer comprehensive analytical capabilities, and HB is probably the typical choice for PCS utility estimation as it allows individual-level analysis.

2. THE NEED FOR MORE INFORMATION

In recent years, PCS has gained a significant increase in popularity among market researchers, due to its potentiality and design and analysis simplicity. Currently, it is a widely adopted model in many different research areas including automotive, FMCG, healthcare, transport economics, etc. The launch and diffusion of commercial software for the analysis of PCS data has surely contributed to the recent success of this approach, by increasing user accessibility and thus making it available to non-statisticians. Sawtooth Software *MaxDiff* (Sawtooth Software, 2007) is probably the most popular package to analyze PCS data (due to their cross-selling strategy – most of their customers approach Sawtooth Software for their wide conjoint offering), however the software seems to handle only the short model. A more complete package is *Demia R-sw Tradeoff* (Demia Studio Associato, 2014) which has been specifically designed to handle both the short and the graded PCS models. Both packages support HB analysis for PCS data and they also handle MDS analysis.

With growing popularity more and more researchers need to better understand the potentialities and limitations of PCS, especially considering that PCS results are often not just presented to the final user in their raw form, but they might be used for additional statistical analyses, such as feeding a segmentation model (Dillon et al., 1993).

To date, it is not very clear the role played by the different PCS exercise elements with respect to the results accuracy. There are several elements to be considered in a PCS study, and all play a potentially key role in the accuracy of PCS results, although their role has not been properly quantified yet:

- number of items considered in the exercise;
- short model versus graded model;
- choice of the scale for the graded model;
- inclusion of a neutral/indifference point;
- number of times each item is presented to each respondent;
- number of PCS tasks in the questionnaire;
- type and number of design versions (blocks);
- number of respondents;
- preference homogeneity among respondents;
- preference homogeneity among items.

There is very little work done in this area, as most of the PCS literature focus on alternative analysis algorithms, on comparisons against other popular approaches such as ratings, rankings or MDS, or on practical applications of the approach.

Corradetti and Furlan (2006) carried out an analysis of the impact of the number of PCS tasks on the quality of the results. In their work, they considered 12 items assessed through a graded PCS model with a 7-point scale inclusive of an indifference point. They varied the number of tasks from 1 to 8 and they found out that there is a linear loss of quality as the number of tasks decreases, and no evident threshold could be identified. More recently, Furlan and Turner (2011) carried out a more comprehensive study based on 15 items to assess the impact on results of the type of PCS model (short, graded with 5 points, graded with 7 points; all models were inclusive of an indifference point), of the number of tasks, and of the number of design versions. They showed that (1) administering at least 5 design versions considerably improves accuracy of results; (2) increasing the number of tasks has an important effect on accuracy; (3) the type of the PCS model adopted for the exercise has an effect on accuracy, with more complex evaluation frameworks providing more accurate results.

With our work, we intended to explore to what extent the inclusion of the neutral point impacts on the results as we could not find any relevant information in the literature, in order to provide some actionable insight for researchers involved in designing PCS exercises. In order to meet this objective, we decided to base our analysis on 15 items as, based on our experience, most PCS projects require the analysis of 12 to 18 items. We have also decided to always have 5 design versions (blocks) to be assigned to the respondents sample because Furlan and Turner (2011) showed that this approach improves accuracy of results and it is also likely to reduce potential context bias and order effects which might have a negative effect on the quality of responses. Finally, we chose to focus on the graded PCS model with a 5-point scale as this represents a good compromise between simplicity (limitation of respondents confusion and fatigue) and estimation accuracy (Furlan and Turner, 2011). We created a number of design combinations by varying the following three elements:

- PCS model: indifference point available/not available (see Figure 1);
- number of tasks to be assessed by each respondent: 12, 15, 18;
- two possible error terms, 15 and 30, to represent respectively high and low respondents’ accuracy (e.g., reflecting two levels of engagement).

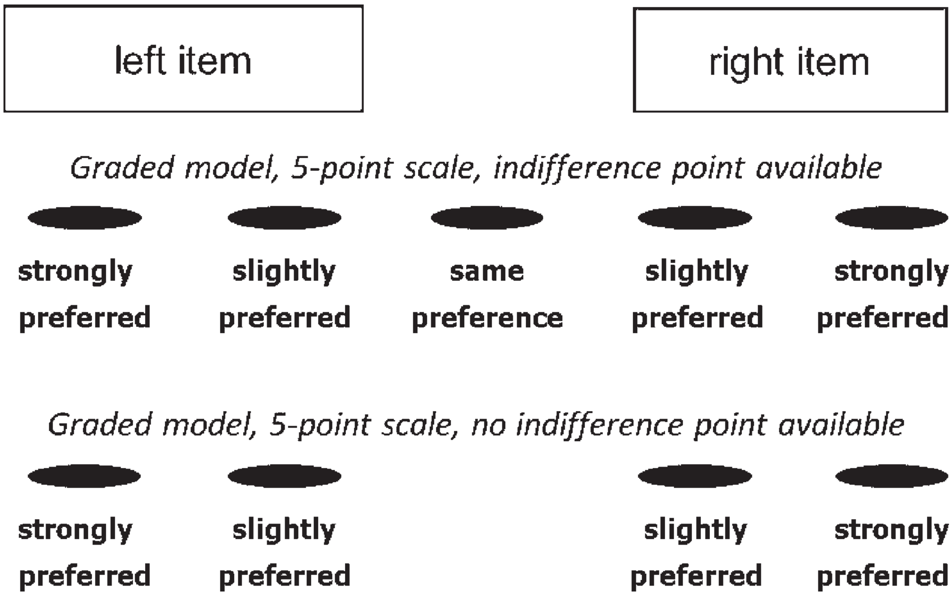

Figure 1: The PCS models explored in this paper.

These three elements (PCS model, number of tasks, and respondents accuracy) characterize a 2x3x2 full factorial design (12 combinations).

In addition to the 12, 15, or 18 tasks assessed by the respondents, we also included some *holdout tasks* for validation purposes. Holdout tasks are scenarios that are "held out" or set aside during the estimation procedure. After estimating the model parameters, it is possible to determine how well the model predicts the holdout observations. Usually, just a couple of holdout tasks are included in a PCS exercise to keep it manageable and keep additional fatigue to a minimum. However, as reported in the next section, we used simulated respondents, thus we could include far more validation tasks (15 were presented to each respondent).

3. THE SIMULATION

In order to assess the impact of the PCS model, of the number of tasks, and of the respondents' accuracy, we could not use results from real surveys, as they would inevitably be based on only one specific combination of these elements. Theoretically speaking, we could have administered alternative designs to the same sample, but this would not have been practical and we might also have risked introducing bias due to the fact that the same respondents would have already been exposed to a similar exercise a number of times. The only practical solution was simulating data, which consisted of two steps:

(1) simulating respondents' preferences (true utilities) for each statement;

(2) exposing these 'computerized' respondents to the alternative PCS design versions to obtain PCS data. These respondents would 'choose' the most preferred item and express the strength of the preference from each pair according to their preferences simulated at (1).

In order to simulate the respondents' preferences, we looked at a number of previous PCS studies analysed with a HB model in order to assess what could be a reasonable distribution for each item. We noticed that the average of PCS preferences across the sample tends to be between 10 and 90 for most items (considering a scale 0:100). The distribution of these preference scores is asymmetric except, as one would expect, for items with an average around 50, with the asymmetry being the largest for the items whose average is closer to 10 (positive skewness) or closer to 90 (negative skewness). We fitted a beta distribution to each PCS item in every available project to assess potential beta coefficients for the PCS preference scores.

Based on this analysis of past studies, we generated preference scores for 15 items and 200 simulated respondents through a two-stage process. We chose this specific sample size as, based on our PCS projects review, this appeared the most common one, a good compromise between robustness and affordability. First, we randomly assigned an average preference score to each item between 10 and 90. Second, we generated scores for every item for each respondent by the addition of a beta distributed random variable with appropriate coefficients. The resulting generated scores were asymmetric with their distributions mirroring those seen in previous PCS studies.

As a second step, we had to give the various PCS design (i.e., 12 to 18 tasks) as well as the 15 holdout tasks to this set of 'computerized' respondents. For each task and respondent, the preference scores associated to the two items within a task were identified and thus transformed into an expressed choice based on an algorithm assumed to closely mirror the choice behaviour in the real world. This algorithm is based on the difference in preference score between the two items in the task and its structure depends on both the model considered and on the error term associated to the respondent, as shown in Figure 2. In case of non-availability of the indifference point, if the difference of the preference scores associated to the two items within the task is smaller than the error term, the two items are equally likely to be selected (slightly preferred left item or slightly preferred right item). The resulting choice data have the correct format to be analyzed by the package R-sw Tradeoff (Demia Studio Associato, 2014) without further recoding.

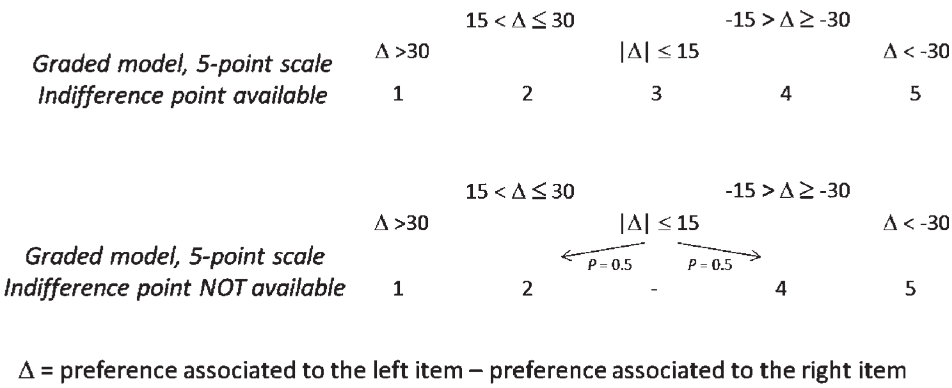

Figure 2: Rationale to convert preference scores into respondents choices (error term = 15)

It is worth noting that no matter how well constructed the 'computerized' respondents are, a simulation is not able to fully mirror the choice behaviour in the real world. In fact, in any trade-off exercise there is a certain amount of response error that might lead, for instance, to an item with higher utility not being preferred to an item with lower utility. However, although this limitation exists, we are confident that the simulation mirrors sufficiently well the actual choice behaviour in terms of order, context, and layout effects. These can be largely reduced and sometimes completely eliminated (e.g., when there are no prohibited combinations) with an accurate experimental design with an excellent one-way, two-way, positional, and within-block balance. One element, however, that could have an effect on the realism of the simulation is the number of tasks seen by each respondent and thus the length of the exercise. This is an effect that has not been well studied in relation to PCS projects and there appears to be the opportunity for further research. However, there is some evidence available for other trade-off models (e.g., conjoint analysis and discrete choice modelling) to suggest that this effect is hardly controllable, as it depends on many elements such as the target respondents, the complexity of the task, the respondent's level of engagement, etc. For this reason, we have decided not to introduce any adjustment coefficients in the simulation.

The design creation, the preference scores generation, the subsequent identification of the preference scores associated to the various items, the choice of the most preferred item, as well as the analyses described in the next section, were performed 100 times for each design combination in order to obtain accurate estimates for the various outcomes of our analysis. Figure 3 illustrates the key steps of the process.

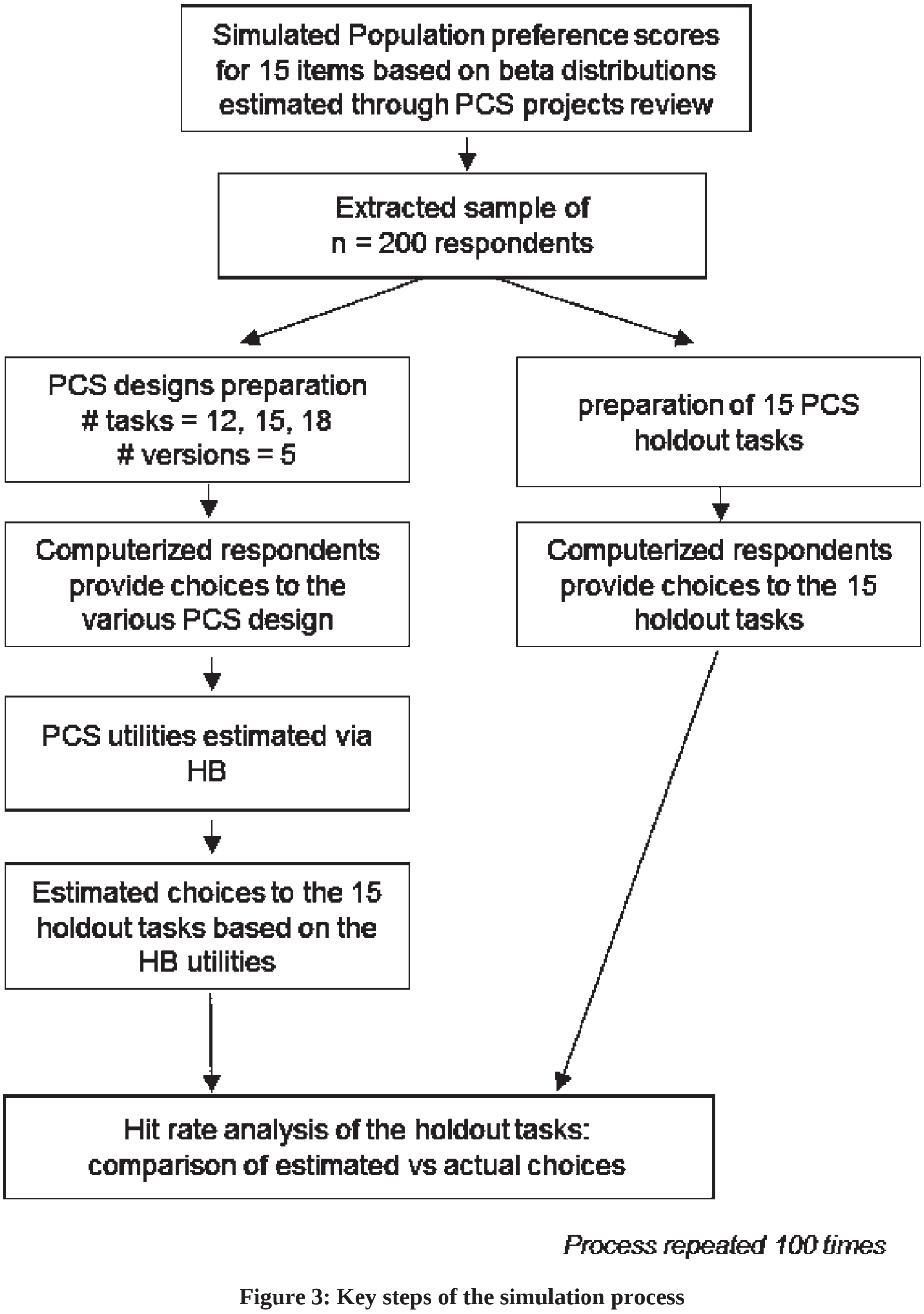

Figure 3: Key steps of the simulation process

4. THE ANALYSIS AND DATA VALIDATION

Once the above simulated datasets had been obtained, we could proceed with the analysis. We decided to analyse the data using HB modelling as this is the typical choice to obtain individual-level results; this approach works well even when there are many items to be estimated with respect to the number of tasks assessed by each respondent.

Only the data associated to the PCS designs have been considered for the analysis, and not also the holdout tasks, which were considered later on for validation purposes only. For the HB analysis, we chose the *estimate.PCS.HB* function available in the package R-sw Tradeoff (Demia Studio Associato, 2014). This choice was dictated by the fact that this is the only commercial software that we are aware of that has been specifically designed to handle the graded PCS model. This is a very flexible and convenient package; it has been possible to prepare an appropriate script to analyse all simulated datasets without repetitive and tedious manual intervention from us. For each alternative design and each respondents set we obtained a full set of PCS individual scores or utilities (a score for each item and respondent).

As a final but important step, we had to choose and adopt an appropriate approach to validate the quality of these sets of utilities against the original simulated preference scores. Thanks to our simulation framework, we had the assessment of 15 holdout tasks for each respondent, thus we could use a hit rate approach. We can say there is a hit when the PCS utility associated to the preferred item in the holdout task is larger than the PCS utility associated to the other item appearing on the task. Therefore, we defined as hit rate the percent of times that the HB model 'guesses' the preferred item. This analysis is based on all sample respondents and all holdout tasks they have been exposed to for which a preference was given either to the left or to the right item. The hit rate index for each design combination, averaged over the 100 iterations, is presented in Figure 4.

It is important to mention that, for the sake of an appropriate interpretation of results, the hit rate score for the model under investigation (i.e., the one based on the PCS individual scores) should be compared against the hit rate score of a random model (i.e., a model based on absolute randomness of choices which is obtained by the reciprocal of the number of items presented in the various tasks). If the hit rate for the model under investigation is significantly higher than the one for a random model (in our case 1/2=50%), then it is possible to say that the model is, to some extent, satisfactory. The hit rate score of a random model represents a lower limit and is used to put the hit rate score into context.

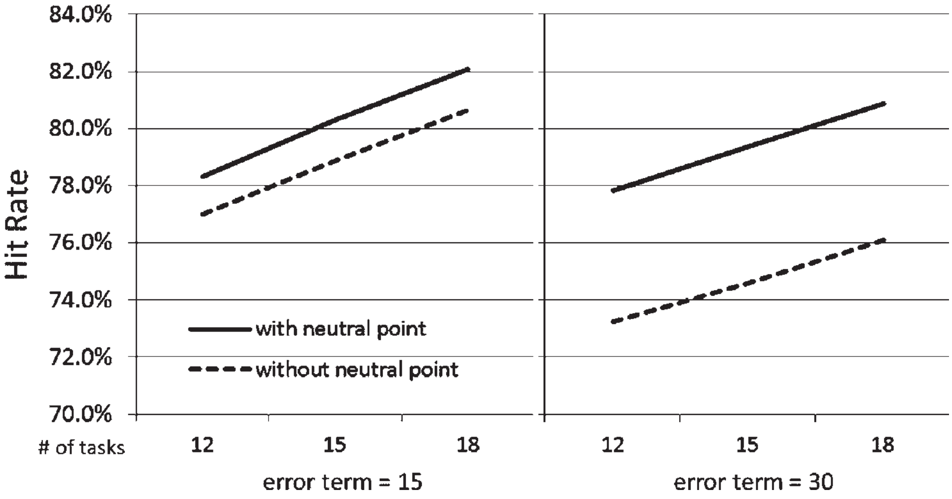

Figure 4: Effects of each design combination based on hit rates

From Figure 4 it is evident that the three design elements we considered in our simulation project all have a significant impact on the results. The magnitude of the impact of the number of PCS tasks was expected as this is usually the main element, along with the number of items to be assessed, considered by the statistician when designing a PCS exercise. This impact seems to be almost linear and this is consistent with previous findings (Corradetti and Furlan, 2006; Furlan and Turner, 2011).

The inclusion of the indifference point has a positive impact on the accuracy of results and this benefit is similar for any number of tasks.

The impact of respondents' accuracy is somehow in line with expectations, although this seems to be much lower when the indifference point is included (Figure 4). Including this point seems to mitigate the negative effects of a larger respondents' inaccuracy.

In addition, it is worth noting that including the indifference point might have a significant and positive impact on the accuracy in an indirect way. For instance, as the indifference point provides respondents with an escape route from having to think, its exclusion is likely to increase respondents' efforts/fatigue, with consequent potential lower quality (larger error term) and a longer survey time. Therefore, we conclude that it seems to be beneficial to have the indifference point in a PCS exercise, especially when we anticipate low respondents' accuracy (e.g., large number of PCS tasks, low engagement, long/complex labels for the PCS elements). The positive effect of the inclusion of the indifference point is extremely valuable for researchers, as it is easy to accomplish, economical, and practical.

5. CONCLUSIONS

With this study we explored how some key elements in a PCS design, in particular the inclusion of the indifference point, impact on the accuracy of results. Our findings are consistent with, and complement well, previous research conducted in this area. Among the various potential elements we could focus on, we chose the PCS model (indifference point available/not available), the number of tasks, and respondents' accuracy.

The main result is that our study indicates the inclusion of the indifference point has a positive direct impact on the accuracy of results, especially when we anticipate low respondents' accuracy (e.g., large number of PCS tasks, low engagement, long/complex labels for the PCS elements). Moreover, the inclusion of the indifference point might have a significant and positive impact on the accuracy also in an indirect way. For instance, as the indifference point provides respondents with an escape route from having to think, its exclusion is likely to increase respondents' efforts/fatigue, with consequent potential lower quality (larger error term) and a longer survey time. Therefore, we conclude that it seems to be beneficial to have the indifference point in a PCS exercise, and this is particularly true when respondents need to be exposed to a large number of PCS scenarios, when the exercise is otherwise potentially challenging or time consuming (e.g., long/complex labels for the PCS elements), or when the questionnaire is very long (there are usually other sections before the PCS exercise).

We also showed that increasing the number of tasks has an important effect on accuracy, well in line with previous research, thus the researcher should include as many tasks as practical in the PCS exercise, but not so many to introduce elements of fatigue and confusion among respondents. In fact, fatigue and confusion have a detrimental effect on accuracy, as our study shows (by varying the error term).

It is important to highlight that the results obtained in this study are valid for a project with 15 items and 12 to 18 tasks included in the questionnaire, and they might be slightly different with a different project setup. Moreover, our findings are valid only for the simulation model we adopted. Results could have been different if another model were appropriate, for instance if average preference scores differed by larger or smaller amounts, their variability was different, or if they followed a lognormal, a gamma, or some other distribution. Further research is needed to assess to what extent the results are affected by the simulation model.

Further research is also required to assess different project setups and to investigate elements that have not been considered in this or previous studies, such as preference homogeneity among respondents and among items. Some additional research is also required to assess the impact of PCS elements in relation to various respondent types (e.g., busy professionals, young or old respondents) and in different fields (FMCG, B2B, durables, etc).

REFERENCES

Chrzan, K. and Golovashkina, N. (2006). An empirical test of six stated importance measures. *International Journal of Market Research*, 48(6): 717-740.

Cohen, S.H. and Neira, L. (2003). Measuring preference for product benefits across countries. Paper presented at the *Sawtooth Software Conference 2003*, San Antonio, TX.

Corradetti, R. and Furlan, R. (2006). A trade-off model for eliciting physicians' or patients' preferences in healthcare research projects. *Memorie della Accademia delle Scienze di Torino* – Classe di Scienze Fisiche, Matematiche e Naturali – Applicazioni della Matematica, Classe V, 30: pp. 55-89.

David, H.A. (1988). *The method of paired comparisons (second edition).* Oxford University Press, New York.

Demia Studio Associato (2014). *R-sw tradeoff: a package for advanced trade-off analysis*. http://www.demia.it. Last access 03/09/2014.

Dillon, W.R., Kumar, A., Smith de Borrero, M. (1993). Capturing individual differences in paired comparisons: an extended BTL model incorporating descriptor variables. *Journal of Marketing Research,* 30: 42-51.

Finn, A. and Louviere, J.J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy and Marketing*, 11: 12-25.

Furlan, R. and Turner, G. (2011). Paired comparisons of statements: exploring the impact of design elements on results. *Statistica Applicata – Italian Journal of Applied Statistics (IJAS)*, special issue on "Customer surveys and market forecasting", 23(3): 405-424.

Jaeger, S.R., Jorgensen, A.S., Aaslyng, M.D., Bredie, W.L.P. (2008). Best-worst scaling: an introduction and initial comparison with monadic rating for preference elicitation with food products. *Food Quality and Preference*, 19: 579-588.

Louviere, J.J. and Flynn, T. (2011). Advances in best-worst scaling (BWS) and choice-based measurement methods. Paper presented at the *AREE 2011 Pre-Conference Workshop*: Recent Developments in the Design and Implementation of Discrete Choice Experiments, Seattle, Washington (June).

Louviere, J.J. and Woodworth, G.G. (1990). *Best-worst scaling: a model for largest difference judgments*. Working paper, Faculty of Business, University of Alberta.

Madansky, A. (2010). Data Use: Evaluating paired comparisons, maximum difference and traditional ranking. *Quirk's Marketing Research Review*, 24(10): 22-27.

Mccullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. London: Chapman and Hall.

Sawtooth Software (2007). The maxdiff/web system technical paper. *Sawtooth Software Technical Paper Series*, http://www.sawtoothsoftware.com/download/ techpap/maxdifftech.pdf. Last access: 03/09/2013.

Srinivasan, V.S. and Wyner, G.A. (2009). An improved method for the quantitative assessment of customer priorities. *Stanford University Graduate School of Business*, Research Paper No. 2028.

Steenkamp, J.B.E.M. and Hofstede, F.T. (2002). International market segmentation: issues and outlook. *International Journal of Research in Marketing*, 19: 185-213.

Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Review*, 34: 273–286.