

Adaptive Hybrid Methods for Choice-Based Conjoint Analysis: A Comparative Study

Robert Bauer¹, Klaus Menrad¹ & Thomas Decker¹

¹Chair of Marketing and Management of Biogenic Resources, Straubing Center of Science, Straubing, Germany
Correspondence: Robert Bauer, Straubing Center of Science, Schulgasse 16, 94315 Straubing, Germany. Tel: 49-157-7831-0427. E-mail: bauer.robert.j@gmx.de

Received: December 5, 2014 Accepted: December 27, 2014 Online Published: January 26, 2015
doi:10.5539/ijms.v7n1p1 URL: <http://dx.doi.org/10.5539/ijms.v7n1p1>

Abstract

Adaptive choice-based conjoint analysis (ACBC) and hybrid individualized two-level choice-based conjoint analysis (HIT-CBC) were developed to improve standard choice-based conjoint analysis through additional interviewing techniques. Both methods have demonstrated their applicability in comparison to standard choice-based conjoint methods. The objective of our study was a direct comparison of the two adaptive hybrid methods ACBC and HIT-CBC. Therefore, we analysed the previous comparative literature on the methods and used the results to conduct both a Monte Carlo simulation study and an empirical study for validity comparisons. The simulation study confirms the vulnerability of HIT-CBC to produce incorrect ratings of respondents in the last part of the questionnaire. The empirical findings reveal an advantage of ACBC in comparison to the current version of HIT-CBC. We conclude that the rating tasks in the last section of HIT-CBC questionnaires reduce the predictive validity of the method and suggest an improvement to HIT-CBC.

Keywords: conjoint analysis, market research, simulation study, CBC, ACBC, HIT-CBC

1. Introduction

Conjoint analysis is widely used (Hauser et al., 2006). However, there are many different conjoint analysis methods and selecting the best method for a (research) problem is a difficult task for researchers and practitioners. Several new methods have arisen in the field of choice-based conjoint analysis (CBC) in recent years. In the present article, we evaluate two adaptive hybrid methods that were developed for product concepts with many attributes and levels.

In classical CBC choice tasks, respondents choose their preferred (product) concept from a set of two or more concepts (Louviere & Woodworth, 1983; Batsell & Louviere, 1991). Because such choice decisions are similar to (purchase) decisions in reality, it is argued that CBC performs better than other methods for preference measurement (Louviere & Woodworth, 1983; Toubia et al., 2004), a claim supported by several studies (Louviere et al., 2004; Toubia et al., 2004; Eggers & Sattler, 2009). In recent years, online surveys have garnered increasing popularity and the rising online processing power allows for adapting questions on the basis of prior responses (Toubia et al., 2004). These advances led to methods that create respondent-specific choice tasks during the survey (e.g., Toubia et al., 2004 and Toubia et al., 2007 suggested (probabilistic) polyhedral methods and Yu et al., 2011 suggested Bayesian methods for the adaptive design generation).

Adaptive hybrid approaches use additional interview techniques to improve the efficiency of the choice tasks. This approach is known from adaptive conjoint analysis (ACA) introduced in the 1980s (Johnson, 1987). Adaptive hybrid methods for CBC were suggested by Johnson and Orme (2007) and Eggers and Sattler (2009). Eggers and Sattler (2009) introduced the hybrid individualized two-level choice-based conjoint approach (HIT-CBC), while Johnson and Orme (2007) developed adaptive choice-based conjoint (ACBC). It is claimed that due to the more elaborate interview technique, the new adaptive hybrid methods can handle product concepts with many attributes and levels more efficiently (Johnson & Orme, 2007; Eggers & Sattler, 2009).

In previous studies, ACBC and HIT-CBC demonstrated their applicability in comparison to the standard CBC approach, which does not include additional interviewing techniques. Our study intends to offer the first direct comparison of these two adaptive hybrid methods. We particularly aimed to compare the performance of those

methods in the case of complex products with many attributes and levels. Therefore, we conducted a Monte Carlo simulation study and an empirical study on car preferences for our validity comparisons.

In section 2, we provide the methodological background of ACBC and HIT-CBC. In section 3, we give an overview of previous comparative studies on either ACBC or HIT-CBC. In section 4, we present the results of the Monte Carlo simulation followed by the results of our empirical study in section 5. The discussion of the results and the conclusion follow in the final section.

2. Methodological Background

2.1 Adaptive Choice-Based Conjoint (ACBC)

ACBC was established by Johnson and Orme (2007) and several scientists have applied ACBC in recent years. For example: Chapman et al. (2009) conducted a study on consumer electronics. Jarvis et al. (2012) studied sour cream. Boesch and Weber (2012) studied the industrial milk, potato and wheat markets. Heinze et al. (2013) carried out a study on real estate investments.

The core of ACBC is the CBC portion of the questionnaire, but, due to the adaptive approach, it is possible to reduce irrelevant information and to make tasks more engaging for the respondents. The researcher defines the attributes and attribute levels before starting data collection in an empirical study. Thereby the questionnaire can be constructed with the Sawtooth Software package. In general, ACBC questionnaires can be divided into four main sections, which are illustrated in appendix B.

In the configurator section, respondents indicate the best level for each attribute. When the analyst assumes a general preference order for specific attributes, the evaluation can be omitted for those attributes. Different product concepts are shown in the screening section. Thereby respondents are not asked to make final choices, but they should indicate whether or not concepts are a possible choice. The design of the concepts is regulated so that mainly the individually preferred levels from the configurator section are presented in the screening section.

If potential non-compensatory screening rules can be detected from previous questions, respondents have to state if an attribute level is a must-have or an unacceptable level. If several of such levels are detected, respondents must choose the most desired or least acceptable. In the choice task section, respondents evaluate the remaining concepts from the screening section. The “winning” concept in each choice set advances to the next stage until a final winner is identified.

The design algorithm of ACBC generates random designs that are not optimal regarding the design efficiency measures. Nevertheless, the set of the algorithm ensures near-orthogonality and high design efficiency values.

In ACBC, individual part-worth utility values are estimated with the Hierarchical Bayes algorithm. Apart from the data from the choice task section, further information can be used for utility estimation. Every choice from the Configurator section may be coded as a choice task where respondents choose 1 of K levels. Information from the screener section can be coded as binary choices where the respondent is assumed to compare the utility of the concept to the utility of a constant threshold prior to making a choice. Part-worth thresholds for the none-option can be estimated via the dropped concepts from the screening section.

Additionally, several more optional components are available in ACBC. For example, a calibration section at the end of the survey can be used to estimate a further part-worth threshold for the none-option. Furthermore, the summed prices approach enables the analyst to incorporate conditional prices.

2.2 Hybrid Individualized Two-Level Choice-Based Conjoint Approach (HIT-CBC)

HIT-CBC was established in Eggers and Sattler (2009). In addition to the initial application of Eggers and Sattler, we found two further published studies using this methodology (Kaltenborn, 2012; Zenker et al., 2013). Eggers and Sattler (2009) conducted a study on flights, Kaltenborn (2012) on the German telecommunication market and Zenker et al. (2013) applied HIT-CBC to residence location choice.

In HIT-CBC, the core of the questionnaire is the CBC portion. The identification of best and worst levels in the initial section renders it possible to simplify the choice design to a fixed 2^k design, which is design-efficient and Pareto optimal.

HIT-CBC surveys can be divided into the following three main stages, which are also illustrated in appendix C. In the first stage, respondents identify the best and worst levels for every attribute. When the analyst assumes a general preference order for specific attributes, the evaluation of the best and worst levels can be omitted for those attributes. Only the best and the worst levels of each attribute are included in the choice task section, in which a none-option can be incorporated optionally. In the last stage, respondents evaluate the remaining levels on a rating scale with the identified best and worst levels as the maximum and minimum scaling points.

HIT-CBC uses a fixed design approach. Due to the fact that only the best and worst levels identified in the first section appear in the choicetasks, every design reduces to a 2^k factorial. A shifting procedure algorithm is applied for the generation of the 2^k design (Burgess & Street, 2003). The algorithm is easy to implement and ensures design efficiency and orthogonality. Due to the knowledge of best and worst levels, dominated choicesets can be excluded a priori. As a consequence, the remaining choicesets are Paretooptimal.

An advantage of the 2^k design is that the number-of-levels effect (NLE) cannot occur in HIT-CBC. The NLE can occur when the numbers of levels are not equally distributed among the attributes. It biases the results since attributes with a higher number of levels tend to have an artificially higher attribute importance (Wilde et al., 2008)

Individual part-worth utilities are estimated with the Hierarchical Bayes algorithm. Because of the 2^k design, only parameters for the best and worst levels must be estimated with the algorithm. Utilities for the remaining levels are adjusted according to the ratings of the interpolation section.

An additional section (WTP elicitation) can be incorporated after the identification of the best and worst levels. In that stage, respondents state their willingness to pay for two concepts. One consists of the best levels for every attribute; another, of the worst levels. Consequently, prices for every other concept must lie in between these two stated prices.

3. Previous Research on the Validity of ACBC and HIT-CBC

3.1 ACBC-Related Validity Comparisons

We identified four studies dealing with the validity of ACBC. These studies all compared the validations of standard choice-based conjoint models to ACBC. Validity was evaluated in terms of hit rates and (market) share predictions for holdout tasks. Three of these studies showed a predominance of ACBC: Johnson and Orme (2007) evaluated validity and several qualitative criteria in two studies. The first study was on laptops (ten attributes). ACBC showed comparable market share predictions (in terms of mean absolute error) and significantly greater hit rates than standard CBC. Even though the median interviewing time in ACBC was almost twice the time in standard CBC, the majority of the respondents preferred ACBC. The second study of Johnson and Orme (2007) was on recreational equipment (eight attributes) and produced comparable results. However, a significant difference in hit rates was not observable. Orme and Johnson (2008) studied home purchases (ten attributes). They compared three different setups of ACBC and standard CBC. With regard to hit rates, ACBC yielded favourable results, although differences between ACBC and CBC were not significant. In terms of market share predictions, descriptive analysis of mean absolute error implied a predominance of ACBC. The qualitative results (e.g., respondent's enjoyment of the tasks in the questionnaire) demonstrated that respondents preferred ACBC.

Chapman et al. (2009) studied a consumer electronics product with eight attributes. In terms of within-subject preference (hit rates in holdout tasks), CBC performed slightly better. However, a significant difference could not be observed. Additionally, the authors analysed the performance of the methods with regard to observed market data with better results in ACBC.

3.2 HIT-CBC-Related Validity Comparisons

We identified two studies dealing with the validity of HIT-CBC. Eggers and Sattler (2009) compared HIT-CBC to standard CBC. They conducted a simulation study and an empirical study on European flights (six attributes). The simulation study showed a predominance of HIT-CBC in two different scenarios. Variances in utility estimates were smaller, and medians of the utility estimates were closer to the theoretical values. The predominance of HIT-CBC was clear, especially in the case of heterogeneity among "respondents." In the empirical study, no significant difference between the methods was observed with regard to hit rates for holdout tasks. Market share predictions led to similar results. Additionally, they found that the overall enjoyment of the respondents was similar to CBC.

Kaltenborn (2012) studied mobile phone contracts (seven attributes) and compared the validity of HIT-CBC, Graded Paired Comparison and standard CBC. On the one hand, he found significantly greater hit rates for CBC compared to Graded Paired Comparisons and HIT-CBC. On the other hand, market share predictions for HIT-CBC yielded the best results. Therefore, a clear predominance of one method was not observable in this study.

3.3 Summary

Few scientific studies have compared ACBC and HIT-CBC to other methods. The ACBC-related studies identified in section 3.1 compare the validity of ACBC to standard CBC. The results imply a predominance of the adaptive hybrid method compared to standard CBC. For HIT-CBC, we identified only two comparative studies. The study of Kaltborn (2012) analysed hit rates and market shares leading to inconsistent results. The study of Eggers and Sattler (2009) compared HIT-CBC to standard CBC. The empirical study revealed similar results for HIT-CBC and CBC. Their simulation study revealed a predominance of HIT-CBC when ratings for the intermediate level utilities (levels between best and worst levels) of HIT-CBC were correct. In practice, this is only true if the ratings in section three of the HIT-CBC questionnaire are carried out correctly by the respondents. In the case of incorrect ratings, the utilities are biased and the results may differ. However, Eggers and Sattler (2009) did not study the effect a posteriori and not as a part of the simulation study. In our simulation study, we investigate the effect of incorrect ratings in the HIT-CBC questionnaire directly (section 4).

4. Simulation Study

Previous research on the adaptive hybrid methods ACBC and HIT-CBC compared these methods to standard CBC (and in one case to Graded Paired Comparisons). In our inquiry into the literature, we did not identify a study with a direct comparison between ACBC and HIT-CBC neither in terms of simulations nor empirical studies. Therefore, we compare ACBC, HIT-CBC and standard CBC in our simulation study. Within that study, we defined eight attributes: three attributes having five levels, two attributes having four levels, two attributes having three levels and one attribute having two levels. Theoretical part-worth utilities have been constructed according to the suggestions of Arora and Huber (2001) and Eggers and Sattler (2009) and can be found in appendix A. In order to account for the effect of preference heterogeneity, we varied the theoretical utility values among three groups. Every group consisted of 50 simulated respondents leading to a sample size of $n = 150$ simulated respondents. Choices in the different choice sets were simulated in the following way: we varied the individual utilities by adding a Gaussian distributed error value with mean zero and standard deviation 0.5 to the theoretical utility value. Then we predicted choices according to the first choice rule. For ACBC and HIT-CBC, we assumed that best and worst levels were identified correctly.

For the rating tasks in section three of HIT-CBC, we defined three different scenarios. In the first scenario, we assumed that the ratings were correct. In order to take into account the possible vulnerability of HIT-CBC to incorrect ratings, we defined two further scenarios. In scenario 2, we varied the ratings by adding an error value with mean zero and standard deviation one; in scenario 3, by adding an error value with mean zero and standard deviation two.

Table 1. Mean absolute errors in the simulation study

Method	Mean absolute error (MAE)
ACBC	0.35
HIT-CBC (Scenario 1)	0.30
HIT-CBC (Scenario 2)	0.35
HIT-CBC (Scenario 3)	0.42
CBC	0.46

For the estimation of individual part-worths, we applied the Hierarchical Bayes (HB) algorithm from Sawtooth Software. The HB estimations exhibited the same settings for all methods. For the evaluation of the methods, we evaluated the mean absolute error over all utility estimates and respondents (Table 1). The adaptive hybrid methods yielded significantly ($p < 0.01$) lower mean absolute errors than standard CBC with one exception. When the ratings in section 3 were assumed to be biased with a mean deviation of 2 (scenario 3) rating points on an eleven point rating scale, HIT-CBC did not perform significantly better than standard CBC ($\alpha = 0.05$). Comparing ACBC and HIT-CBC produced similar results when the rating is biased with a mean deviation of 1 rating point (scenario 2).

5. Empirical Study

The aim of the empirical study was to compare the validity of ACBC and HIT-CBC in a real-world application with a complex product. Standard CBC was not superior to ACBC or HIT-CBC in previous studies nor in our simulation study. Therefore, we decided to abstain from standard CBC in the empirical study and rather to

concentrate on a direct comparison of the two adaptive hybrid approaches. For this purpose, we conducted an online study aiming to measure preferences for cars in the German market.

5.1 Research Design

The adaptive hybrid methods ACBC and HIT-CBC were suggested for the measurement of preferences in the case of complex product concepts (i.e., many attributes and levels) (Eggers & Sattler, 2009; Sawtooth Software, 2009a). In order to compare the methods in a meaningful application, we chose cars as an example of a complex product with a total of eight attributes and 31 attribute levels (Table 2).

Preferences for cars in the German market were assessed in the choice experiment. Our discrete choice experiment was hypothetical, and the aim of our study was to compare the validity of ACBC and HIT-CBC. Traditionally, the car market consists of several vehicle types that are often differentiated by means of vehicle size. Since several car attributes vary heavily among vehicles types, we concentrated on one vehicle type in our study, namely the compact class type. The compact class was chosen because vehicles of that class were mostly sold during recent years. The market share of compact class vehicles was 23.8% in 2012 (German Federal Motor Transport Authority, 2013).

Table 2. Explored attributes and attribute levels for cars

Attribute	Attribute levels
Brand	BMW, Kia, Renault, Toyota, VW
Drive system	diesel, electric, gasoline, hybrid, natural gas
Engine performance	63 KW, 77 KW, 105 KW
Design	classical, futuristic, inconspicuous, sporty
Vehicle equipment	grassroots-equipment, premium
Driving range	150 km, 450 km, 900 km
Energy consumption	4.50 €, 7 €, 9.50 €, 14.50 €
Price	16,990 €, 21,240 €, 26,490 €, 31,990 €, 36,990 €

In developing a choice experiment, the first step is to decide attributes and attributes levels. This determination influences all results and is consequently very important for a preference study (Helm et al., 2008). We developed the attributes and corresponding levels for our study in the following way. We first reviewed scientific and commercial literature to detect possible items (e.g., Bunch et al., 1993; Moore, 2004; Eggers & Eggers, 2011; Rust, 2011). Afterwards, we ran expert interviews and a pretest to determine the attributes and attribute levels listed in table 2. We found that alternative drive systems are an important issue for car purchases, and we identified battery electric vehicles, natural gas vehicles and hybrids as the most important alternative drive systems. In our study, we collectively termed range extended electric vehicles (REEV), plug-in hybrid electric vehicles (PHEV) and hybrid electric vehicles (HEV) as “hybrids.” Because of their very early market stage and the lacking filling station infrastructure in Germany, fuel cell vehicles were not used in the study. First-generation biofuels were excluded from the study as well. In Germany, biofuels are mostly used as an admixture in petrol and are not regarded as an actual alternative to conventional drive systems.

The attribute “vehicle equipment” was explicitly explained in the online questionnaire in the following way: “Grassrootsequipment includes air conditioning, car radio with CD and USB, electric windows, six airbags, ESP and TCS. Premium equipment includes in addition to grassrootsequipment heated seats, car computer with navigation system, alarm, parking sensors, lane departure warning system and bending light.”

For the evaluation of cars’ eco-friendliness, several authors studied the attribute CO₂ emissions (Bhat et al., 2009; Achtnicht, 2012; Ziegler, 2012; Olson, 2013). However, expert interviews and the pretest revealed that this attribute does not belong to the most important attributes and we consequently excluded it from the study.

The levels of the brand attribute facilitate diversity in prices ranging from low-priced (Kia) to premium brands (BMW). We also included brands that, at the time of our study, were already successful in selling alternative fuel vehicles (Toyota hybrids, Renault battery electrics and VW natural gas vehicles) on the German market. Purchase and energy prices represent typical German market prices from the compact class segment at the time of our study. We did not exclude any combinations of attribute levels that were unrealistic at the time of our study (e.g., battery electrics with high driving ranges), as respondents were explicitly asked to make their choices in a hypothetical situation. This setup allowed us to anticipate the technological innovations in the German car market in the next years ahead.

For measuring preferences, it is important that potential respondents have a basic knowledge of the car decision problem in general and alternative drive systems in particular (Helm et al., 2008). Therefore, we selected students who acquire technical knowledge concerning cars during their vocational training. Every respondent answered the HIT-CBC and the ACBC tasks in a randomly determined order in the online survey. This randomized within-subjects design allowed us to compare the methods for each respondent individually and to preclude order effects. For a discussion of the within-subjects design, see Charness et al. (2012). Additionally, respondents evaluated both methods in terms of closeness to reality, task simplicity, enjoyment and their overall pleasure. The online survey was open from November 2012 until January 2013. Altogether, 423 respondents completed the survey.

The HIT-CBC part was arranged in the following way. In the first part, respondents selected the best and worst levels for all attributes including price. In order to reduce complexity, we treated price as a "usual" attribute, i.e., we refrained from the WTP elicitation. Afterwards, respondents considered twelve choice sets, which were presented in a randomly determined order. These choice sets consisted of three alternatives each. A none-option was not given. In the last part of our study, the respondents evaluated the remaining levels on an eleven point scale.

The ACBC portion was created according to the suggestions of SawtoothSoftware (2009a). We set the number of screening tasks to 7; the number of concepts per screening task to 4; the minimum of attributes to vary from BYO Selections to 2 and the maximum to 3; the number of unacceptables was set to 4 and the number of must-haves to 3. The maximum number of products brought into choice tournament was 16 and the number of concepts per choice task 3. No calibration concepts were included in the survey. Again price was treated as "usual" attribute, i.e., we refrained from the summed pricing approach in order to reduce the complexity of the tasks.

To measure the internal predictive validity, respondents additionally evaluated three holdout choice sets, which were not used for estimation of individual utility values. Furthermore, a holdout group of 66 respondents (Group 2) served as an external prediction benchmark to assess cross-sample predictive validity. This group only evaluated the three holdout choice sets mentioned above.

For data analysis the software packages from Sawtooth Software were applied. For assessing validity measures and further statistical test R was used.

5.2 Empirical Results

In order to compare the validity of ACBC and HIT-CBC, we applied the following assessment criteria:

- Goodness of fit
- Predictive validity (regarding hit rates and RMSE)
- Qualitative criteria (respondents' evaluation of the methods, interviewing time analysis)

For the estimation of individual part-worths, which were the basis for most of the assessment criteria, we applied the Hierarchical Bayes algorithm (Sawtooth Software, 2009b) to both methods.

5.2.1 Goodness of Fit

The goodness of fit of the model indicates the internal information processing within a method. In our empirical comparison, we assess the goodness of fit through the pseudo- R^2 measure, which is a typical measure for logit models (McFadden, 1974; Hauser, 1978; Hensher et al., 2007). As mentioned above, we applied Sawtooth Software's Hierarchical Bayes algorithm for the estimation of the individual part-worths in ACBC and HIT-CBC. The model output contains the percent certainty measure, which is equivalent to the pseudo- R^2 measure (Hauser, 1978; Sawtooth Software, 2009b).

The pseudo- R^2 of a choice model is not exactly the same as the R^2 of a linear regression model. However, there exists a direct empirical relationship between the two measures, and pseudo- R^2 values between the range of 0.3 and 0.5 can be translated as an R^2 of between 0.6 and 0.9 for the linear model equivalent (Domencich & McFadden, 1975; Hensher et al., 2007). Therefore, a model with a pseudo- R^2 of 0.3 can be interpreted as a decent model (Hensher et al., 2007). In our empirical study, both methods revealed similar pseudo- R^2 values and both methods led to decent model fits (ACBC: pseudo- $R^2=0.529$; HIT-CBC: pseudo- $R^2=0.476$).

5.2.2 Predictive Validity

Predictive validity indicators assess the ability of a method to predict real choices (Helm et al., 2008). Following the published scientific literature (Toubia et al., 2004; Helm et al., 2008; Sawtooth Software, 2009a, Chapman et

al., 2009; Eggers & Sattler, 2009; Meißner et al., 2011), we analysed hit rates and market share predictions for measuring internal predictive validity. Hit rates for the three holdout choice tasks were computed from expected part-worthsthrough the first-choice rule, and market share predictions were evaluated with mean absolute error (MAE). Both measures were tabulated as mean values across the three holdout choice tasks. Accordingly, external MAE was derived for the holdout group (Group 2).

Table 3. Predictive validity assessment in terms of hit rates and MAE

Method	Hit rate (%)	Internal MAE (%)	External MAE (%)
ACBC	64.22	5.44	5.45
HIT-CBC	55.79	6.66	7.37

Overall, the predictive validity indicators demonstrated that both methods were able to predict choices (table 3). The majority of the choice predictions were successful for both methods. In particular, both hit rates were significantly (McNemar's test; $p < 0.001$) higher than the random hit rate of 25% (as we presented four stimuli in the holdout sets). Comparing the two analysed methods, ACBC performed better than HIT-CBC in the three holdout tasks (McNemar's test $p < 0.05$). ACBC also performed better in market share predictions as shown in the internal and external MAE (table 3).

5.2.3 Qualitative Criteria

We asked respondents to evaluate the two methods in terms of closeness to reality, task simplicity, enjoyment and their overall pleasure using a seven point scale from 7=high to 1=low. The assessment of these subjective criteria is important, since tasks that are too difficult or too monotonous may lead to less carefully determined trade-offs and consequently to less validity (Hartmann & Sattler, 2004).

In every category, ACBC was preferred by the respondents (t-Test, $p < 0.05$) (table 4). As mentioned in section 5.1, the presentation order of the methods was randomly determined for each respondent. Therefore, the differences cannot be explained by order effects. In HIT-CBC, task simplicity was evaluated poorly. Since sections one and two of HIT-CBC are quite similar to the tasks in ACBC, we suppose that the rating tasks in section three are crucial for the poor evaluation of the task simplicity criterion. We therefore assume that the rating of the intermediate levels is complicated for respondents.

Table 4. Respondents' evaluation of qualitative criteria for ACBC and HIT-CBC

Criterion	ACBC	HIT-CBC
Closeness-to-reality	4.92*	4.31
Task simplicity	4.06*	3.48
Enjoyment	4.64*	4.10
Overall pleasure	5.22*	4.48

* Statistically different at $p < 0.05$ (t-Test).

Another important criterion is the time needed to complete the survey. Very time-consuming questionnaires may tire or bore respondents, leading to an increasing threat to biased evaluations. As longer survey times usually incur higher costs, time is also a cost factor for commercial studies (Helm et al., 2008). Regarding the time respondents spent on the methods, we did not find a significant difference (t-Test, $\alpha = 0.05$). The average time used per method was approximately 9 minutes (ACBC 551s; HIT-CBC 535s).

Additionally, we regarded the number-of-levels effect. While ACBC is vulnerable to the effect, it cannot occur in HIT-CBC because of the 2^k design in the choice sets. Previous research suggests that number-of-level effects are most likely for respondents that do not have well defined preferences (Wittink et al., 1989; Wittink et al., 1992; Hair et al., 2010). The inclusion of alternative fuel vehicles in our car preference study allows us to account for a degree of preference uncertainty. Battery electrics, hybrids and natural gas vehicles are new products on the German market and their market share is very small. In 2012, the market share of alternative fuel vehicles was only 0.96% in new vehicle registrations (German Federal Motor Transport Authority, 2013). Thus we assume that most respondents do not have well defined preferences for alternative fuel vehicles.

In order to study the number-of-levels effect, we analysed individual attribute importances of the attributes with the most attribute levels: brand, drive system and price, each of which had five levels. For brand and drive

system, the individual attribute importance was not significantly higher in ACBC than in HIT-CBC (t-Test; $\alpha = 0.05$). For the attribute price, we observed a significantly higher importance for ACBC (t-Test; $p < 0.001$). Therefore, clear evidence for the occurrence of the number-of-levels effect in ACBC could not be observed in our empirical study. In the recent literature, we could not identify any scientific studies analysing the number-of-levels effect for ACBC. This is an opportunity for further research.

6. Discussion and Conclusion

Recent adaptive hybrid methods refine the concept of CBC by introducing respondent-specific choice tasks. The aim of our work was to compare the validity of two adaptive hybrid methods, namely ACBC and HIT-CBC, which were suggested for the measurement of preferences in the case of complex product concepts. For this purpose, we reviewed scientific literature for previous comparative studies on the methods. Based on the results of the literature review, we conducted a Monte Carlo simulation study and an online study on the German car market for empirical validity comparisons. In order to compare the methods in a meaningful empirical application, we chose cars as an example. We defined eight attributes and a total number of 31 attribute levels. The simulation study revealed advantages of the adaptive hybrid methods in comparison to standard CBC. However, it also revealed the vulnerability of HIT-CBC to incorrect ratings of intermediate attribute levels. The empirical study showed that both methods were able to apply to the complex study design. Regarding predictive validity measures (hit rates and market share predictions) and several qualitative criteria, we found that both methods were able to predict choices. The direct comparison of the two adaptive hybrid methods ACBC and HIT-CBC demonstrated that especially in case of hit rates and the qualitative criteria, ACBC performed significantly better than HIT-CBC.

The analysis of qualitative criteria (closeness to reality, task simplicity, enjoyment and overall pleasure) is important, since questionnaires that are too difficult or too monotonous may reduce a method's validity (Helm et al., 2008). The results showed that the complex study design did not lead to dissatisfaction of the respondents and consequently showed the applicability of both methods to the complex study design. This finding is in line with former studies on ACBC (Johnson & Orme, 2007; Orme & Johnson, 2008; Sawtooth Software, 2009a) and HIT-CBC (Eggers & Sattler, 2009). More specifically, we found that ACBC outperformed HIT-CBC with regard to the four measured qualitative criteria. The criterion with the worst rating was the task simplicity of HIT-CBC.

Eggers and Sattler (2009) posited that HIT-CBC might be vulnerable to incorrect ratings of intermediate levels. Their simulation study revealed that incorrect ratings bias aggregate level analyses substantially. However, they did not study the effect directly, i.e., they included a rating error a posteriori and not during simulations of individual responses. Therefore, one focus of our simulation study was the actual simulation of incorrect ratings. We found that the performance of HIT-CBC depends to a great extent on the ratings in the last section of the HIT-CBC questionnaire. With that result in mind, our comparative empirical study yields a further important outcome. Firstly, we found that the predictive validity of HIT-CBC was worse than ACBC, even though the simulation study showed that both methods should lead to similar results if the rating error in section 3 of HIT-CBC is small. Secondly, respondents evaluated the task simplicity of ACBC significantly higher than for HIT-CBC. Since the first sections of HIT-CBC are quite similar to the analogous tasks in ACBC, we suppose that the rating tasks in section 3 are crucial to the poor evaluation of the task simplicity criterion of HIT-CBC. We therefore assume that respondents have difficulties in these rating tasks leading to incorrect ratings and consequently less predictive validity of HIT-CBC. We suggest a methodological improvement of HIT-CBC in the rating section that could be tested in future research. A possible solution might be to create rating tasks that are more appealing graphically. Additionally, it might be beneficial to combine the first section (identification of the best and worst levels) and the rating section of HIT-CBC.

In our empirical study, we did not observe a significant impact of the number-of-levels effect on the results. However, the main focus of our study was validity comparisons rather than an analysis of the number-of-levels effect. As we could not identify any scientific studies analysing the number-of-levels effect for ACBC in recent literature, our results may provide a first reference for future studies on the number-of-levels effect in ACBC applications.

When interpreting our findings, some limitations of our study should be taken into account. Firstly, the attribute price was treated as an "ordinary" attribute. However, both methods also provide opportunities for a deeper study of the price attribute and consequential of WTP figures. One question ripe for further research is how the inclusion of additional pricing components such as the summed pricing approach in ACBC and the WTP elicitation in HIT-CBC influences the empirical results and the validity of the methods in different applications.

The second limitation of our study concerns the measurement of predictive validity. For this purpose, we included holdout tasks in the questionnaire. Additionally, a holdout group served as an external prediction benchmark. In order to overcome the hypothetical bias of such an approach, several authors suggest observing real purchase decisions for validity assessments. Wertenbroch and Skiera (2002) and Ding (2007) evaluated real purchase decisions directly in the survey. Chapman et al. (2009) compared the predicted market shares from internal holdout tasks to observed actual market shares. However, for big-ticket items and very innovative product concepts (for which no physical products are on the market so far), the inclusion of real purchase decisions is quite challenging, although it would be an interesting topic for further research.

In the simulation study we mainly investigated the impact of incorrect rating in HIT-CBC. Of course, there are several topics, which also would be interesting to study (e.g., number of attributes, number of attribute levels, number of simulated consumer groups). This is a direction for further research.

The findings of our study provide implications for practitioners. As our results imply a superiority of ACBC, we suggest this method for practical applications. However, since ACBC might be vulnerable to the number-of-levels effect, we advise that those who would use it be alert to the possible occurrence of the effect, especially in studies where respondents do not have well defined preferences (Wittink et al., 1989; Wittink et al., 1992; Hair et al., 2010). We do not recommend the application of HIT-CBC in its current version. The vulnerability of the method to incorrect ratings in the last section of the questionnaire should be remedied before practical application.

For the analyst, the choice of the method may also depend on practicability. Sawtooth Software provides a software package for ACBC that enables the analyst to create the questionnaire and, after fielding, to derive utility estimates (Sawtooth Software, 2009a). Such a software package is not yet available for HIT-CBC. While the experimental design is fixed and easy to create for HIT-CBC, this should not strictly prevent application of this method because as well Sawtooth Software's CBC/HB package as the R package "bayesm" can be used for utility estimations (Rossi, 2013).

Towards the promising introduction of ACBC and HIT-CBC, this first comparative study on both methods confirmed their validity in the simulation study as well as in the empirical study. The findings reveal a predominance of ACBC in comparison to the current version of HIT-CBC. Therefore, we suggest an improvement of HIT-CBC in the rating section to improve the predictive validity of the method.

References

- Achtnicht, M. (2012). German car buyers' willingness to pay to reduce CO2 emissions. *Climatic Change*, 113(3-4), 679-697. <http://dx.doi.org/10.1007/s10584-011-0362-8>
- Arora, N., & Huber, J. (2001). Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments. *Journal of Consumer Research*, 28(2), 273-283. <http://dx.doi.org/10.1086/322902>
- Batsell, R. R., & Louviere, J. J. (1991). Experimental Analysis of Choice. *Marketing Letters*, 2(3), 199-214. <http://dx.doi.org/10.1007/BF00554126>
- Bhat, C. R., Sen, S., & Eluru, N. (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B: Methodological*, 43(1), 1-18. <http://dx.doi.org/10.1016/j.trb.2008.06.009>
- Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies*, 27(2), 87-101. <http://dx.doi.org/10.1111/j.1745-459X.2012.00370.x>
- Boesch, I., & Weber, M. (2012). Processor's Preferences and Basic Differentiation Strategies for Potatoes, Milk, and Wheat in Switzerland. *Journal of Agricultural & Food Industrial Organization*, 10(1). <http://dx.doi.org/10.1515/1542-0485.1377>
- Bunch, D. S., Bradley, M., Golob, T. F., Kitamura, R., & Occkiuizzo, R. (1993). Demand for clean-fuel vehicles in California: A discrete-choice stated preference pilot project. *Transportation Research Part A*, 27A(3), 237-253. [http://dx.doi.org/10.1016/0965-8564\(93\)90062-P](http://dx.doi.org/10.1016/0965-8564(93)90062-P)
- Burgess, L., & Street, D. J. (2003). Optimal Designs for 2 k Choice Experiments. *Communications in Statistics - Theory and Methods*, 32(11), 2185-2206. <http://dx.doi.org/10.1081/STA-120024475>

- Chapman, C. N., Alford, J. L., Johnson, C., Weidemann, R., & Lahav, M. (2009). CBC vs. ACBC: Comparing Results with Real Product Selection. Retrieved from <http://www.sawtoothsoftware.com/download/techpap/chapman09.pdf>
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1-8. <http://dx.doi.org/10.1016/j.jebo.2011.08.009>
- Ding, M. (2007). An Incentive-Aligned Mechanism for Conjoint Analysis. *Journal of Marketing Research*, (24), 214-223. <http://dx.doi.org/10.1509/jmkr.44.2.214>
- Domencich, T., & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis* (1st ed.). Amsterdam and Oxford: North-Holland Publishing Co.
- Eggers, F., & Eggers, F. (2011). Where have all the flowers gone? Forecasting green trends in the automobile industry with a choice-based conjoint adoption model. *Technological Forecasting and Social Change*, 78(1), 51-62. <http://dx.doi.org/10.1016/j.techfore.2010.06.014>
- Eggers, F., & Sattler, H. (2009). Hybrid individualized two-level choice-based conjoint (HIT-CBC): A new method for measuring preference structures with many attribute levels. *International Journal of Research in Marketing*, 26(2), 108-118. <http://dx.doi.org/10.1016/j.ijresmar.2009.01.002>
- Frederick, S., & Fischhoff, B. (1998). Scope (in)sensitivity in elicited valuations. *Risk Decision and Policy*, 3, 109-123.
- German Federal Motor Transport Authority. (2013). Fahrzeugzulassungen (FZ): Neuzulassungen von Kraftfahrzeugen nach Umwelt-Merkmalen. Jahr 2012. Retrieved from http://www.kba.de/cln_031/nn_191064/SharedDocs/Publikationen/FZ/2012/fz14_2012__pdf,templateId=raw,property=publicationFile.pdf/fz14_2012__pdf.pdf
- Greenwald, A. G. (1976). Within-Subjects Designs: To Use or Not To Use? *Psychological Bulletin*, 83(2), 314-320. <http://dx.doi.org/10.1037/0033-2909.83.2.314>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River [etc.]: Pearson.
- Hartmann, A., & Sattler, H. (2004). Wie robust sind Methoden zur Präferenzmessung? (in German) *Zeitschrift für betriebswirtschaftliche Forschung*, 3-22.
- Hauser, J. R. (1978). Testing and Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach. *Operations Research*, 26(May-June), 406-421.
- Hauser, J. R., Tellis, G. J., & Griffin, A. (2006). Research on Innovation: A Review and Agenda for Marketing Science. *Marketing Science*, 25(6), 687-717.
- Heinzle, S. L., Boey Ying Yip, A., & Low Yu Xing, M. (2013). The Influence of Green Building Certification Schemes on Real Estate Investor Behaviour: Evidence from Singapore. *Urban Studies*, 50(4), 1-18. <http://dx.doi.org/10.1177/0042098013477693>
- Helm, R., Steiner, M., Scholl, A., & Manthey, L. (2008). A Comparative empirical study on common methods for measuring preferences. *International Journal of Management and Decision Making*, 9(3), 242-265. <http://dx.doi.org/10.1504/IJMDM.2008.017408>
- Hensher, D. A., Rose, J. M., & Greene, W. H. (2007). *Applied choice analysis* (3rd ed.). Cambridge: Cambridge University Press.
- Hsee, C., & Zhang, J. (2004). Distinction bias: misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology*, 86, 680-695. <http://dx.doi.org/10.1037/0022-3514.86.5.680>
- Jervis, S. M., Ennis, J. M., & Drake, M. A. (2012). A Comparison of Adaptive Choice Based Conjoint and Choice Based Conjoint to Determine Key Choice Attributes of Sour Cream with Limited Sample Size. *Journal of Sensory Studies*, 27, 451-462. <http://dx.doi.org/10.1111/joss.12009>
- Johnson, R. M. (1987). Adaptive Conjoint Analysis. *Sawtooth Software Conference Proceedings on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, 253-265.
- Johnson, R. M., & Orme, B. K. (2007). A New Approach to Adaptive CBC. Retrieved from <http://www.sawtoothsoftware.com/download/techpap/acbc10.pdf>

- Kaltenborn, T. (2012). VALCA: Validating Conjoint Analysis - Evidence from the German Telecommunication Market. In D. Vrontis (Ed.), *Building new business models for success through competitiveness and responsibility: Conference readings book proceedings* (pp. 855-867). [S. l.]: EuroMed Press.
- Lohrke, F. T., Holloway, B. B., & Woolley, T. W. (2009). Conjoint Analysis in Entrepreneurship Research: A Review and Research Agenda. *Organizational Research Methods*, 13(1), 16-30. <http://dx.doi.org/10.1177/1094428109341992>
- Louviere, J. J., & Woodworth, G. (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research*, 20(4), 350-367.
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2004). *Stated choice methods: Analysis and applications* (3rd ed.). Cambridge: Cambridge University Press.
- McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.
- Meißner, M., Decker, R., & Adam, N. (2011). Ein empirischer Validitätsvergleich zwischen Adaptive Self-Explicated Approach (ASE), Pairwise Comparison-based Preference Measurement (PCPM) und Adaptive Conjoint Analysis (ACA) (in German). *Zeitschrift für Betriebswirtschaft*, 81(4), 423-446. <http://dx.doi.org/10.1007/s11573-011-0449-8>
- Moore, W. L. (2004). A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing*, 21(3), 299-312. <http://dx.doi.org/10.1016/j.ijresmar.2004.01.002>
- Netzer, O., & Srinivasan, V. (2011). Adaptive Self-Explication of Multiattribute Preferences. *Journal of Marketing Research*, 28, 140-156. <http://dx.doi.org/10.1509/jmkr.48.1.140>
- Olson, E. L. (2013). It's not easy being green: the effects of attribute tradeoffs on green product preference and choice. *Journal of the Academy of Marketing Science*, 41(2), 171-184. <http://dx.doi.org/10.1007/s11747-012-0305-6>
- Orme, B. K., & Johnson, R. M. (2008). Testing Adaptive CBC: Shorter Questionnaires and BYO vs. "Most Likelies". Retrieved from <http://www.sawtoothsoftware.com/download/techpap/acbc3home.pdf>
- Rossi, P. (2013). Bayesian Inference for Marketing/Micro-econometrics. Retrieved from <http://cran.r-project.org/web/packages/bayesm/bayesm.pdf>
- Rust, H. (2011). *Autoträume der jungen Generation: Lehren für das strategische Management* (1st ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Sawtooth Software. (2009a). The Adaptive Choice-Based Conjoint (ACBC) Technical Paper. Retrieved from <http://www.sawtoothsoftware.com/download/techpap/acbctech.pdf>
- Sawtooth Software. (2009b). The CBC/HB System for Hierarchical Bayes Estimation: Version 5.0 Technical Paper. Retrieved from <http://www.sawtoothsoftware.com/download/techpap/hbtech.pdf>
- Toubia, O., Hauser, J. R., & Garcia, R. (2007). Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application. *Marketing Science*, 26(5), 596-610. <http://dx.doi.org/10.1287/mksc.1060.0257>
- Toubia, O., Hauser, J. R., & Simester, D. J. (2004). Polyhedral Methods for Adaptive Choice- Based Conjoint Analysis. *Journal of Marketing Research*, (16), 116-131. <http://dx.doi.org/10.1509/jmkr.41.1.116.25082>
- Wertenbroch, K., & Skiera, B. (2002). Measuring Consumers' Willingness to Pay at the Point of Purchase. *Journal of Marketing Research*, (29), 228-241. <http://dx.doi.org/10.1509/jmkr.39.2.228.19086>
- Wilde, E. de, Cooke, A. D. J., & Janiszewski, C. (2008). Attentional contrast during sequential judgements: A source of the number-of-levels effect. *Journal of Marketing Research*, 45(4), 437-449. <http://dx.doi.org/10.1509/jmkr.45.4.437>
- Wittink, D. R., Huber, J., Zandan, P., & Johnson, R. M. (1992). The Number of Levels Effect in Conjoint: Where does it come from and can it be eliminated? *Sawtooth Software Conference Proceedings on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*.
- Wittink, D. R., Lakshman, K., & Reibstein, D. (1989). The Effect of Differences in the Number of Attribute Levels on Conjoint Analysis. *Marketing Letters*, 2, 113-123.

- Yu, J., Goos, P., & Vandebroek, M. (2011). Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*, 28(4), 378-388. <http://dx.doi.org/10.1016/j.ijresmar.2011.06.002>
- Zenker, S., Eggers, F., & Farsky, M. (2013). Putting a price tag on cities: Insights into the competitive environment of places. *Cities*, 30, 133-139. <http://dx.doi.org/10.1016/j.cities.2012.02.002>
- Ziegler, A. (2012). Individual characteristics and stated preferences for alternative energy sources and propulsion technologies in vehicles: A discrete choice analysis for Germany. *Transportation Research Part A*, 46(8), 1372-1385. <http://dx.doi.org/10.1016/j.tra.2012.05.01>

Appendix A

Utility values in the Monte Carlo simulation

Table A1. Assumed utility values in the Monte Carlo simulation study (150 simulated respondents)

Utility values		Group 1 (50 simulated respondents)	Group 2 (50 simulated respondents)	Group 3 (50 simulated respondents)
Attribute 1	Level 1	-1	-0,2	0,2
	Level 2	-0,5	-0,1	0,6
	Level 3	0	0	-0,2
	Level 4	0,5	0,1	-0,6
	Level 5	1	0,2	0
Attribute 2	Level 1	-0,75	-0,75	-0,75
	Level 2	-0,375	-0,375	-0,375
	Level 3	0	0	0
	Level 4	0,375	0,375	0,375
	Level 5	0,75	0,75	0,75
Attribute 3	Level 1	-0,5	0,5	-1
	Level 2	0	0	0
	Level 3	0,5	-0,5	1
Attribute 4	Level 1	-0,8	-0,8	-0,8
	Level 2	0	0	0
	Level 3	0	0	0
	Level 4	0,8	0,8	0,8
Attribute 5	Level 1	-1	-1	-1
	Level 2	1	1	1
Attribute 6	Level 1	-0,2	-0,2	-0,2
	Level 2	0	0	0
	Level 3	0,2	0,2	0,2
Attribute 7	Level 1	-1	1	0,2
	Level 2	-0,1	0,1	-0,2
	Level 3	0,1	-0,1	-0,4
	Level 4	1	-1	0,4
Attribute 8	Level 1	-0,9	-0,9	-0,9
	Level 2	-0,5	-0,5	-0,5
	Level 3	0	0	0
	Level 4	0,5	0,5	0,5
	Level 5	0,9	0,9	0,9

Appendix B
Main phases of an ACBC questionnaire

ACBC

Click your preferred choice for each feature below.

feature	select feature
Brand	<input type="checkbox"/> BMW
	<input type="checkbox"/> Kia
	<input type="checkbox"/> Renault
	<input type="checkbox"/> Toyota
	<input checked="" type="checkbox"/> VW

For each product, indicate whether it is a possibility or not.

	Product 1	Product 2	Product 3
Brand	VW	Renault	VW
Drive system	Gas	Hybrid	Electric
Engine performance	105 KW	94 KW	105 KW
Design	Classical	Futuristic	Futuristic
Equipment	Standard	Standard	Premium
Range	450 km	900 km	150 km
Energy consumption (€)	7,50 €	14,50 €	4,50 €
Price	26.990 €	16.990 €	16.990 €
Your choice:	<input type="checkbox"/> A possibility <input checked="" type="checkbox"/> Won't work	<input type="checkbox"/> A possibility <input checked="" type="checkbox"/> Won't work	<input checked="" type="checkbox"/> A possibility <input type="checkbox"/> Won't work

We don't want to jump to conclusions, but we have noticed you have chosen products with certain characteristics. Are any of these an absolute must? If so, mark the one **most important**.

<input type="checkbox"/> Brand: - Renault
<input type="checkbox"/> Range: - 900 km
<input checked="" type="checkbox"/> Equipment: - Premium

Among these three, which product would you most want to get.

	Product 1	Product 2	Product 3
Brand	VW	VW	BMW
Drive system	Electric	Electric	Electric
Engine performance	105 KW	105 KW	105 KW
Design	Classical	Futuristic	Futuristic
Equipment	Premium	Premium	Premium
Range	450 km	150 km	900 km
Energy consumption (€)	7,50 €	4,50 €	4,50 €
Price	26.990 €	16.990 €	16.990 €
Your Choice:	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Figure B1. Main phases of an ACBC questionnaire (author's illustration)

Appendix C

Main phases of a HIT-CBC questionnaire

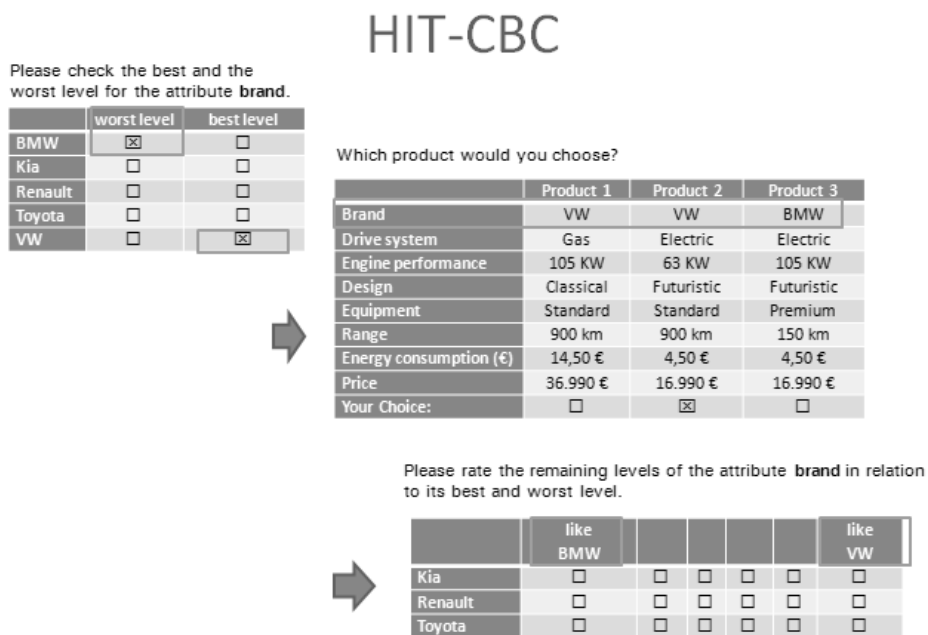


Figure C1. Main phases of a HIT-CBC questionnaire (author's illustration)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).